European Journal of Information Systems

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/tjis20

Will PLS have to become factor-based to survive and thrive?

Ned Kock

EUROPEAN JOURNAL OF INFORMATION SYSTEMS
2024, VOL. 33, NO. 6, 882–902
https://doi.org/10.1080/0960085X.2024.2360564

THE OPERATIONAL RESEARCH SOCIETY
Taylor & Francis
Taylor & Francis Group

RESEARCH ARTICLE

Will PLS have to become factor-based to survive and thrive?

Ned Kock

Texas A&M Regents Professor and Chair, Division of International Business and Technology Studies, Texas A&M International University, Laredo, TX, USA

ABSTRACT
Structural equation modelling (SEM) is a general method that aims at estimating models with latent variables (LVs), where the LVs are measured indirectly and with some imprecision via questionnaires. This is done usually employing question-statements answered on Likert-type scales. In this paper we discuss various forms of SEM, and demonstrate that composite-based models, common in classic partial least squares (PLS) implementations, are poorly aligned with the very idea of SEM. We argue that minimisation of type I and II errors, or false positives and negatives respectively in hypothesis testing, can only happen if LVs are implemented as factors (and not as composites). This requires the use of modern, factor-based PLS methods, which have some advantages not only over classic PLS implementations, but also over covariance-based SEM approaches. Our main goal with this paper is to stimulate debate, whether pro or against our views. If we are generally correct in our thinking, the impact on how quantitative research is conducted in the field of information systems, as well as many other fields, could be quite dramatic. The reason for this is the widespread use of SEM in information systems, business, and the behavioural sciences.

ARTICLE HISTORY
Received 31 August 2022
Accepted 14 May 2024

KEYWORDS
Composite; factor; measurement with error; partial least squares; structural equation modeling

1. Introduction

Structural equation modelling (SEM) is a data analysis method that allows a researcher to test both a structural model and a measurement model, simultaneously. The structural model, which aims at summarising elements of a theoretical model, usually involves a set of variables that cannot be measured directly without error, called latent variables (LVs); and causal relationships among these LVs, represented through arrows. The measurement model involves variables that measure LVs with error, typically as responses to question-statements on Likert-type scales in questionnaires. (For reference, these and other key terms are listed in alphabetical order, with their respective definitions, in Appendix A).

SEM has traditionally been associated with the implementation of LVs as factors; where the LVs cause their indicators, leading to the emergence of measurement residuals when the LVs are regressed on their indicators. (Mathematically, this can be done, even though the indicators are caused by the LVs). Each measurement residual can be seen as accounting for the variance in its corresponding LV that is not accounted for by its indicators. The variance accounted for by the indicators in this context is the true reliability associated with each LV. If the indicators were to account for 100 percent of the variance in their LVs, there would be no measurement residual, and thus the LVs would be composites – exact linear combinations of the indicators, and nothing else.

SEM with composites has been growing in use, particularly due to the widespread use of partial least squares (PLS) techniques (Ghasemy et al., 2020; Hair et al., 2019; Lowry & Gaskin, 2014; Memon et al., 2021; Ringle et al., 2020; Sarstedt et al., 2020; Shmueli et al., 2016). Classic PLS techniques tend to focus on the identification of combinations of weights through which indicators are aggregated into LVs, as composites. The field of information systems (IS) is closely associated with the development, implementation, assessment, and use of classic PLS techniques (Chin, 1998; Hajiheydari & Ashkani, 2018; Kock & Hadaya, 2018; Kock & Lynn, 2012; Mahmud et al., 2017; Petter, 2018). PLS techniques may also be used to implement factor-based SEM (Kock, 2019a,b), as we will discuss in this paper.

In this paper, we discuss various forms of SEM, and demonstrate that composite-based models are poorly aligned with the idea of measurement with error, whether measurement is reflective or formative. From a conceptual standpoint, first LVs are causally linked as part of the structural model to be tested in SEM, and then those LVs are measured with error. SEM allows a researcher to use this measurement with error approach to recover coefficients of association among LVs *without* error. This leads to the minimisation of type I and II errors, or false positives and negatives respectively, in hypothesis testing. But this can only happen, we argue, if LVs are implemented as factors (and not as composites). We put forth a key

CONTACT Ned Kock nedkock@gmail.com

argument in this paper, which is that PLS-based methods will have to become factor-based to survive and thrive in the context of SEM.

We also discuss a proposed solution: a new form of SEM that builds on PLS algorithms to generate correlation-preserving factors, which we see as fundamentally a numeric computing solution, hence its particular relevance to the audience of a prestigious IS journal such as this. It should be stressed to readers that the leading journal in SEM is the appropriately named *Structural Equation Modeling* journal, edited by someone we regard as the world's foremost authority in SEM, George Marcoulides, who has also weighed in on issues we address here (Hardin & Marcoulides, 2011; Marcoulides & Saunders, 2006; Marcoulides et al., 2009). We see IS and SEM as somewhat related fields; some IS inventions can be used in SEM methods, and SEM is widely used by IS researchers.

To allow for a more straightforward discussion, and without any impact on its generality or broad applicability, all variables and parameters presented are standardised, unless stated otherwise. That is, all variables are scaled so that they have a mean of zero and standard deviation of one. This makes the parameters presented throughout our discussion directly comparable with one another.

2. Our main goal with this paper

Our main goal with this paper is to stimulate debate, whether pro or against our views. If we are generally correct in our thinking, the impact on how quantitative research is conducted in the field of IS, as well as many other fields, could be quite dramatic. The reason for this is the widespread use of SEM in IS, business, and the behavioural sciences. Classic PLS techniques have made SEM analyses easier to be conducted, but have come under a great deal of criticism due to being composite-based. And for good reason.

Underscoring the need for this paper is a recently published issue of the journal *Communications of the Association for Information Systems*, centred on a lead article on recent developments in PLS. While that lead article was ostensibly aimed at IS researchers, it completely ignored a new form of factor-based SEM that builds on PLS algorithms, recently developed by Kock (2019a), which is central to the discussion presented in this paper. This is problematic because this new form of factor-based SEM arguably addresses the vast majority of the criticisms in that lead article.

We believe that the lead article ignored the new form of factor-based SEM that builds on PLS algorithms, mentioned above, not because the authors of the lead article simply decided to ignore it. Rather, it is reasonable to believe that the authors needed further clarifications about that new SEM method, after which they could take a position pro or against it. Hence the need for this paper, where we provide such clarifications through a discussion that brings together various different threads related to SEM. Those individual threads have been elaborated on elsewhere, but not in an integrated way. Extensive coverage of each individual thread would require multiple papers, possibly one paper for each of the main threads that are being integrated. That would, obviously, be beyond the scope of this paper. We hope that our discussion, which is fundamentally conceptual, and buttressed by summarised illustrations, will provide the basis for major progress in SEM as a whole.

3. Measurement with error and factors

LVs are representations of mental constructs, or latent constructs, that must exist in the mind of a person who wants to measure those constructs. That person is typically a researcher who aims to test a structural model that is made up of LVs. There are two main ways in which LV measurement can be conducted: reflective and formative. Generally, reflective measurement relies on redundant question-statements, and formative measurement relies on non-redundant question-statements. This is illustrated in Figure 1.

For example, a researcher who wants to measure "job satisfaction" could use question-statements like "I like my job" and "my job is great", to be answered on Likert-type scales. This example refers to reflective LV measurement because respondents would tend to provide highly correlated answers to these question-statements. On the other hand, the researcher may use question-statements like "I like my boss" and "I like my office", which would entail formative LV measurement. Here highly correlated answers are not expected; respondents may like their bosses but not their offices, and vice-versa.

The mental constructs associated with LVs do not always have to be abstract ideas such as "job satisfaction". They can be more concrete, such as the actual use of electronic communication media ("e-comm. media use"). Still, reflective measurement involves the use of redundant question-statements, while formative measurement involves the use of non-redundant question-statements. For example, question-statements of the type "I use e-comm. media" and "using e-comm. media is important to me" would be used in reflective measurement. question-statements of the type "I use email" and "I use video conferencing" would be used in formative measurement.

Non-redundant question-statements add value by covering different facets of an LV. But why would one want to use redundant question-statements at all? One key reason is that each question-statement is expected to measure the LV with some degree of imprecision, in part because the researcher sees the question-

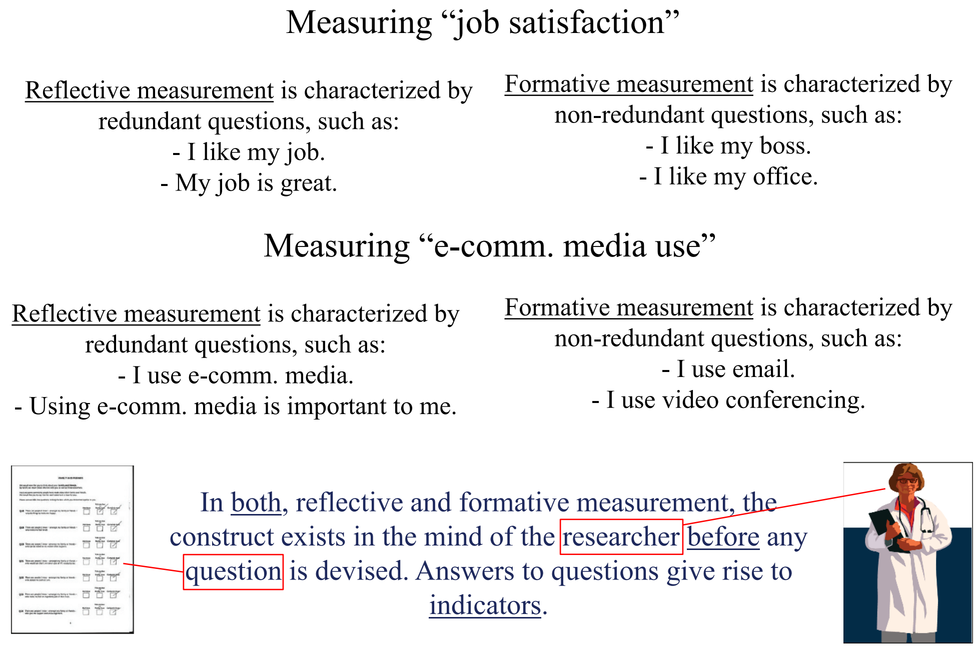


Figure 1. Latent constructs always exist before indicators.

statement as closely aligned with the underlying mental construct. If multiple redundant question-statements are used, one can then check whether: the respondents understood the question-statements as associated with the same LV as the researcher did (convergent validity); and the respondents did not mistake the question-statements as being associated with another LV (discriminant validity). Good convergent validity would imply high correlations among answers to question-statements for the same LV, and good discriminant validity would imply low correlations among answers to question-statements for different LVs.

As we can see from the discussion above, if questionnaires are used to measure LVs, then the mental constructs associated with the LVs must exist before the question-statements. In this scenario, the LVs are quantified as factors and the answers to question-statements as indicators, and the factors always cause the indicators (see Figure 2), whether measurement is reflective or formative. In advocating this, we are admittedly at odds with much of the literature on formative measurement in IS and other fields (see, e.g., Cenfetelli & Bassellier, 2009; Petter et al., 2007; Sarstedt et al., 2016). If we were to believe that a set of indicators caused their corresponding LV, then we would have to conclude that the indicators existed before the construct associated with the LV, which is impossible if the indicators were designed to measure the construct – the latter must have existed before in the mind of the designer of the question-statements associated with the LV.

If we regress a factor on its indicators, we obtain weights. Also, this leads to the emergence of a measurement residual, which accounts for the variance in the factor that is not accounted for by the indicators. Unless one of the indicators is a perfect measure of their LV, which would make that indicator identical to its factor and obviate the need for the other indicators, there is always some residual variance that is not accounted for based on the indicators. If we regress the indicators on their factor, we obtain loadings, which are factor-indicator correlations. Also, this requires the consideration of indicator error terms, which account for the variance in the indicators that are not accounted for by their factor. The measurement residual could be viewed as an extra "indicator", which is uncorrelated with the actual indicators, and for which the weight equals the loading.

At this point the reader may ask: If a factor causes its indicators, how can the factor be regressed on indicators that cannot exist before the factor? After all, the factor explains the indicators, not the other way around. The answer to this question is that, when one analyses empirical data, only the indicators are available. Therefore, it is unavoidable that, to successfully estimate a factor, one needs to use the indicators. This can be done by mathematically expressing the LV as an aggregation of the indicators plus an uncorrelated measurement residual; even though we clearly acknowledge that it is the factor that causes the indicators. In doing so, we are not assuming that the indicators cause their LV. We are simply resorting to a mathematical formulation to express the LV in terms of its indicators. This applies to both reflective and formative measurement.

If the indicators were to be seen as giving rise to their LV, then the LV could be quantified as a composite, which would be an exact linear combination of the indicators. In this case, there would be no

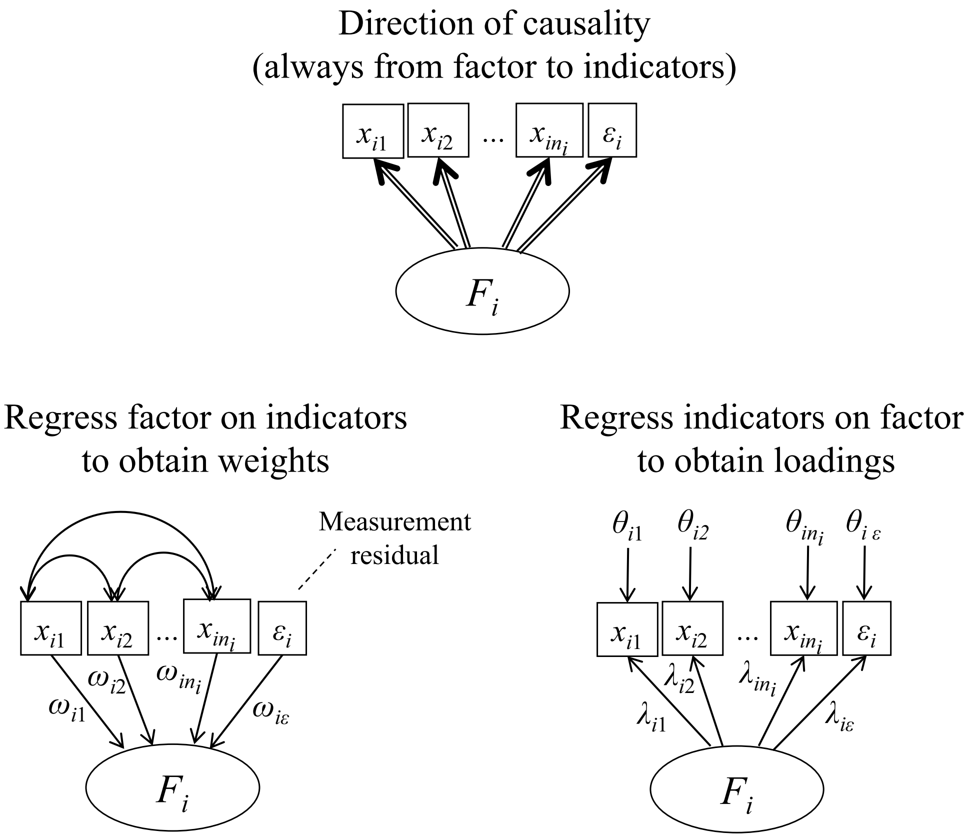


Figure 2. Causality and regression directions. Causality direction indicated via compound arrows; regression direction indicated via single arrows; for the measurement residual the weight equals the loading.

measurement residual. This view is often associated with formative measurement (Hardin et al., 2011), and the idea that the indicators can, under certain circumstances, cause their factor. The problem with this idea is that the question-statements associated with the indicators are devised by a researcher based on a mental idea that must exist before the question-statements themselves. It is therefore impossible that indicators can cause factors.

The position that indicators cannot logically cause the LVs that they (i.e., the indicators) are designed to measure, even in formative measurement, is important in the context of this paper because we argue here that PLS needs to become factor-based to survive and thrive. If indicators were seen as causing LVs when formative measurement is employed, that would open the door for the argument that classic composite-based PLS should be preferred with formative LVs. This would provide the impetus for continued use of composite-based PLS, which would weaken our argument. We realise however, that this is a controversial topic (Hardin et al., 2011).

4. Structural equation modelling

Through SEM a researcher can simultaneously test a structural model, specified as a set of structural relationships among LVs (known as the "structural model" or "inner model"); and a measurement model, specified as a set of relationships among LVs and indicators (also known as the "outer model"). The relationships among LVs are quantified in various ways, notably via path coefficients. The LV-indicator relationships are also quantified in various ways, notably via loadings.

Two main classes of SEM methods find widespread use today: covariance-based (CB) and variance-based SEM (Kline, 1998; Kock & Lynn, 2012; Schumacker & Lomax, 1996). The former is fundamentally a factor-based class of methods. The latter, variance-based SEM, builds on PLS techniques, and can be based on composites and factors. These are discussed in the following sections.

5. CB-SEM

An analysis employing CB-SEM usually starts with a two-stage least squares (TSLS) regression, where initial values of a set of parameters, notably loadings and path coefficients, are estimated and stored in a column vector $\hat{\theta}$. For this the LVs are at first estimated based on dominant indicators; whose initial loadings are set to 1. The TSLS regression is an extension of ordinary least squares (OLS) regression that corrects for biases in path coefficients due to variation from omitted variables or from exogenous LVs being transferred to endogenous LVs indirectly; that is, in cases where the exogenous LVs are not modelled as direct causes of the endogenous LVs in question, but transfer variation to the endogenous LVs via other LVs.

Once the initial parameter estimates are generated and stored in $\hat{\theta}$, new estimates are iteratively obtained based on the equation below, until final convergence is achieved. This process does not entail the estimation of the LVs; their initial estimates used in the TSLS regression are discarded. The iterative process centred on the equation below is a generalisation of a method developed by Isaac Newton. In the equation below, new parameter estimates are obtained by subtracting from the vector of parameters $\hat{\theta}$ the product between the inverse of the Hessian matrix $H\left(\hat{\theta}\right)$ and the vector of first derivatives $u\left(\hat{\theta}\right)$.

$$\hat{\theta} = \hat{\theta} - H\left(\hat{\theta}\right)^{-1} u\left(\hat{\theta}\right).$$

Each element of $u\left(\hat{\theta}\right)$ is calculated as the first derivative of a function F of the difference between the model-implied and actual indicator covariance matrices, with respect to each parameter. Typically, the function F is the maximum likelihood function; though other functions can also be used, such as the generalised least squares function. Each element of the Hessian matrix $H\left(\hat{\theta}\right)$ is calculated as the second derivative of the function F with respect to each parameter.

Iterations are conducted until each element of $u\left(\hat{\theta}\right)$ approaches zero; i.e., becomes smaller than a small fraction. Note that in each iteration the Hessian matrix $H\left(\hat{\theta}\right)$ has to be re-estimated, which makes CB-SEM a rather computationally complex method. While this is often not a problem in terms of wait time for users of CB-SEM software, it does seem to be associated with more variability in the final set of parameters for each sample being analysed, possibly because of propagation of sampling error. This means, essentially, that the standard errors for the parameters estimated via CB-SEM tend to be larger than those generated by computationally simpler methods.

6. PLS-PM

Karl Jöreskog was one of the main contributors to the development of CB-SEM. He was a student of Herman Wold, who developed what is generally known as the PLS path modelling (PLS-PM) method. This method is much more computationally efficient than CB-SEM, and generates approximations of the LVs as composites. Those LV approximations allow for a large number of parameters to be obtained directly; e.g., path coefficients among linked LVs, and LV-indicator weights and loadings. However, the parameters generated by PLS-PM do not converge to the true values as sample sizes increase. That is, the PLS-PM method is not "statistically consistent".

An analysis employing PLS-PM starts with each LV being estimated as a composite $\hat{C}$ that is a standardised sum of its indicators. Then several OLS regressions are conducted among linked LVs and indicators, generating initial values of various parameters; notably path coefficients, weights, and loadings. The PLS-PM method then proceeds by alternating between two steps known as the inside and outside approximations (Lohmöller, 1989).

The inside approximation entails re-estimating each composite $\hat{C}_i$ according to the equation below, where: $Stdz(\cdot)$ is the standardisation function; and A_i is the number of composites $\widehat{C}_j$ $(j = 1 \ldots A_i)$ that are neighbours of the composite $\widehat{C}_i$. Two composites are said to be neighbours if their corresponding LVs are causally linked, whether they are at the beginning or end of the arrows. The values of $\hat{v}_{ij}$ are set according to one of three schemes: (a) the signs of the correlations among neighbours, in the centroid scheme; (b) the correlations among neighbours, in the factorial scheme; or (c) the path coefficients or correlations among neighbours, in the path weighting scheme, depending on whether the arrows go in or out respectively.

$$\hat{C}_i = Stdz\left(\sum_{j=1}^{A_i} \hat{v}_{ij}\hat{C}_j\right).$$

The outside approximation entails calculating LV-indicator weights $\hat{w}_{ij}$ employing one of two modes: Mode A, where the weights $\hat{w}_{ij}$ are calculated by regressing each indicator on its composite; or Mode B, where the weights $\hat{w}_{ij}$ are calculated by regressing each composite on its indicators. Once this is done, each composite $\hat{C}_i$ is re-estimated according to the equation below, where: $Stdz(\cdot)$ is the standardisation function; and n_i is the number of indicators x_{ij} associated with the composite $\hat{C}_i$.

$$\hat{C}_i = Stdz\left(\sum_{j=1}^{n_i} \hat{w}_{ij}x_{ij}\right).$$

Iterations are conducted until each of the weights $\hat{w}_{ij}$ differs from its value in the previous iteration by less than a small fraction. Note that the PLS-PM method is much less computationally complex than CB-SEM. This arguably leads to less variability in PLS-PM in the final set of parameters for each sample being analysed than in CB-SEM, where the variability possibly stems from the propagation of sampling error. In other words, the standard errors of the parameters estimated via PLS-PM tend to be smaller than those estimated via CB-SEM.

The use of composites, as opposed to factors, brings about a problem that becomes more serious as LV reliabilities decrease: the PLS-PM method converges to biased parameter estimates as sample sizes grow to infinity. Notably, loadings tend to be overestimated, path coefficients for non-zero effects (at the population level) tend to be underestimated, and path coefficients for zero effects tend to be overestimated. In part because of these biases, the PLS-PM method is often seen as not being an appropriate method to conduct SEM analyses, in contrast to CB-SEM.

7. PLSF-SEM

As noted earlier in this paper, a new form of SEM that builds on PLS algorithms to generate correlation-preserving factors (PLSF-SEM) has been recently developed (Kock, 2019a, b). PLSF-SEM is designed for SEM analyses where LVs are modelled as factors, and not as composites. It relies on the consistent PLS technique, developed by one of the greatest mathematical statisticians to have ever lived, the late Theo Dijkstra; who, like Karl Jöreskog, was one of Herman Wold's former students (Huang, 2013). This new PLSF-SEM method starts with a PLS-PM analysis employing the centroid scheme. Using the weights generated by this analysis, two equations from the consistent PLS technique (see, e.g., Dijkstra & Schermelleh-Engel, 2014) are used to produce consistent estimates of LV reliabilities and LV-indicator loadings.

The PLSF-SEM method then proceeds in a stochastic fashion to a composite estimation stage. This stage first generates a set of independent and identically distributed random variables that stand in for the measurement residuals associated with each of the LVs. It then iterates across the equations below, where: $\hat{F}_i$ is the factor, and $\hat{C}_i$ is the composite, associated with the LV indexed by i; x_i is the matrix of indicators associated with factor $\hat{F}_i$; $\hat{\lambda}_i$ is the vector of loadings associated with the factor; the $'$ superscript indicates the transpose operation; $\Sigma_{x_i x_i}$ is the covariance matrix of the indicators associated with factor $\hat{F}_i$; $\Sigma_{x_i \hat{\theta}_i}$ is the matrix of covariances among the indicators and their errors; $diag(\cdot)$ is a function that returns the diagonal version of a matrix; and the superscript $+$ denotes the Moore – Penrose pseudoinverse transformation.

$$\hat{F}_i = Stdz\left(\hat{C}_i \hat{\omega}_{iC} + \hat{\varepsilon}_i \hat{\omega}_{i\varepsilon}\right).$$

$$\hat{\theta}_i = x_i - \hat{F}_i \hat{\lambda}_i{}'.$$

$$\hat{\omega}_i = \Sigma_{x_i x_i}{}^{-1}\left(\Sigma_{x_i x_i} - diag\left(\Sigma_{x_i \hat{\theta}_i}\right)\right)\hat{\lambda}_i{}'^{+}.$$

$$\hat{C}_i = \frac{1}{\hat{\omega}_{iC}}\left(x_i \hat{\omega}_i\right).$$

Since the consistent PLS equations yield consistent estimates of the reliabilities $\hat{\rho}_i$, we can calculate the composite and measurement residual weights as $\hat{\omega}_{iC} = \sqrt{\hat{\rho}_i}$ and $\hat{\omega}_{i\varepsilon} = \sqrt{1-\hat{\rho}_i}$, respectively. Iterations are conducted until each of the weights $\hat{\omega}_i$ differs from its value in the previous iteration by less than a small fraction. At the end of this composite estimation stage we have a set of composites $\hat{C}_i$ that can be used as a basis for the generation of the final correlation-preserving factors.

From measurement error theory (Nunnally, 1978; Nunnally & Bernstein, 1994) we know that the correlation between each pair of factors in a model is related to the correlation between the corresponding composites as follows: $\hat{\Sigma}_{F_i F_j} = \Sigma_{\hat{C}_i \hat{C}_j} / \sqrt{\hat{\rho}_i \hat{\rho}_j}$. Based on this, the PLSF-SEM method then proceeds to a factor estimation stage where iterations are conducted employing the equations below, where: $\Sigma_{\hat{F}_i \hat{F}_j}$ is the correlation among each pair of estimated factors; $\Sigma_{\hat{C}_i \hat{\varepsilon}_i}$ is the correlation between an estimated composite and its corresponding measurement residual; and $\Sigma_{\hat{F}_i \hat{\varepsilon}_i}$ is the correlation between an estimated factor and its measurement residual.

$$\hat{\varepsilon}_i = Stdz\left(\hat{\varepsilon}_i + \left(\hat{\Sigma}_{F_i F_j} - \Sigma_{\hat{F}_i \hat{F}_j}\right)\frac{\hat{\Sigma}_{F_i F_j}}{\hat{\omega}_{i\varepsilon}}\left(\hat{C}_j \hat{\omega}_{jC} + \hat{\varepsilon}_j \hat{\omega}_{j\varepsilon}\right)\right).$$

$$\hat{F}_i = Stdz\left(\hat{F}_i + \left(\hat{\omega}_{iC} - \Sigma_{\hat{F}_i \hat{C}_i}\right)\hat{C}_i \hat{\omega}_{iC}\right).$$

$$\hat{\varepsilon}_i = Stdz\left(\hat{\varepsilon}_i - \Sigma_{\hat{C}_i \hat{\varepsilon}_i}\hat{C}_i \hat{\omega}_{iC} + \left(\hat{\omega}_{i\varepsilon} - \Sigma_{\hat{F}_i \hat{\varepsilon}_i}\right)\hat{F}_i \hat{\omega}_{i\varepsilon}\right).$$

$$\hat{F}_i = Stdz\left(\hat{C}_i \hat{\omega}_{iC} + \hat{\varepsilon}_i \hat{\omega}_{i\varepsilon}\right).$$

$$\hat{\varepsilon}_i = Stdz\left(\frac{1}{\hat{\omega}_{i\varepsilon}}\left(\hat{F}_i - \hat{C}_i \hat{\omega}_{iC}\right)\right).$$

The iterations continue until the absolute sum of the differences in $\Sigma_{\hat{F}\hat{F}} - \hat{\Sigma}_{FF}$ falls below a small fraction. When this is achieved, the correlations among the estimated factors $\Sigma_{\hat{F}\hat{F}}$ match the target correlations $\hat{\Sigma}_{FF}$. Finally, the PLSF-SEM method ends with a TSLS regression, whereby consistent estimates of a large number of parameters become available. These include: path coefficients, loadings, and weights. Moreover, correlation-preserving estimates of the factors also become available, and can be used in a variety of tests.

As we can see, the computational complexity of PLSF-SEM is higher than that of PLS-PM, but still much lower than that of CB-SEM. This is so in part because of the latter method's calculation and use of Hessian matrices, although other aspects could be behind this greater computation complexity (e.g., more parameters having to be estimated concurrently). This leads to lower standard errors for PLSF-SEM than CB-SEM as sample sizes grow from small to very large, as both PLSF-SEM and CB-SEM converge to the true values of the parameters for very large samples. Therefore, PLSF-SEM has higher statistical efficiency than CB-SEM, even though both are statistically consistent methods.

At this point an expert reader could counter that Lehmann and Casella (2006) have provided a mathematical proof of the statistical efficiency of CB-SEM via maximum likelihood, which poses a challenge to our argument above. In our view, there are two ways in which we can reconcile these apparently contradictory views. The first is by noting that the proof by Lehmann and Casella (2006) makes a number of assumptions that are not made by PLSF-SEM, and whose violation could lead to exceptions to the proof. For example, they state that: "For small sample sizes, [CB-SEM via maximum likelihood] estimators can be unsatisfactory . . . the estimator can take on negative values although the estimand is known to be non-negative" (Lehmann & Casella, 2006, p. 99). Note that if a method displays higher statistical efficiency than another method, that advantage would be particularly appealing at small sample sizes. Kock (2019a) demonstrated that PLSF-SEM performs quite well with small samples.

The second way in which we can reconcile these apparently contradictory views is that the implementation of different algorithms via software involves constraints that are not covered by mathematical proofs. Notably, an iterative algorithm that is more complex will tend to have more software loops that incorporate stop conditions (for convergence) requiring loss of precision and thus a greater impact of sampling error on standard errors. For instance, one such stop condition could be a change in value for a variable that is lower than 0.0001 from the previous loop's value (implying convergence to a solution). The computational complexity of PLSF-SEM is much lower than that of CB-SEM, in part because of the latter method's calculation and use of Hessian matrices of second order partial derivatives, thus possibly leading CB-SEM to have greater standard errors.

Since in PLSF-SEM all LVs are assumed to be factors, whether formative or reflective measurement is used, there are no distinct formative and reflective modes as in PLS-PM (modes A and B). This applies to second-order LV implementations, which are typically assessed as formative. Because all of the parameters usually estimated via CB-SEM and PLS-PM are available from a PLSF-SEM analysis, empirical analyses tests designed in the context of those methods can be used in PLSF-SEM. For example, loadings, cross-loadings, correlations among LVs, and average variances extracted are used in PLSF-SEM (see, e.g., Amora, 2021) to conduct confirmatory factor analyses as normally done in CB-SEM (Kline, 1998), also incorporating some innovations employed in PLS-PM (see, e.g., Rasoolimanesh, 2022). Fit indices obtained from the comparison of model-implied and empirical indicator covariance matrices are used in PLSF-SEM (see, e.g., Kock, 2020) as normally done in CB-SEM (Kline, 1998).

The two most widely used software tools to conduct PLS-PM, namely SmartPLS and WarpPLS (Memon et al., 2021), have been taking notably different paths with respect to factor-based implementations of SEM algorithms. Since version 4.0, the latest as of this writing, SmartPLS offers CB-SEM as an addition to PLS-PM, but does not offer PLSF-SEM. WarpPLS has been offering PLSF-SEM since version 5.0 (it is in version 8.0 at the time of this writing), as an addition to PLS-PM, but does not offer CB-SEM. One could argue that offering PLSF-SEM is more of a methodological software innovation than offering CB-SEM, since the latter has been available from other SEM software tools for quite some time (e.g., Amos, Mplus, and Lavaan).

8. Illustrating the differences

Figure 3 shows a set of results, primarily related to accuracy of estimates. True values are shown on the top-left area. Estimates obtained with each of the three methods are shown on the right part of the figure. On the bottom-left area we present bar charts with root-mean-square errors (RMSEs) associated with the differences among estimates and the corresponding true values. We used a large sample (N = 10,000) to minimise sampling bias; that is, the estimates shown are close to the asymptotic convergence values. In keeping with the discussion presented so far in this paper, the LVs were modelled as factors that cause their corresponding indicators.

The model is typical of IS research, or more specifically of business research on technological impacts, from a conceptual standpoint. It predicts that the degree to which a technology is used (T) by various individuals is fully mediated via its effect on a mediator (M) with respect to its ultimate effect on the job performance of the individuals in a particular task (P). Consistently with the idea of full mediation, the true path coefficients are strong, at $\beta_{MT} = .650$ and $\beta_{PM} = .600$, but well below the values that could induce full-collinearity variance inflation factors above the 3.3 threshold (Kock & Lynn, 2012). The true loadings assume a measurement model that

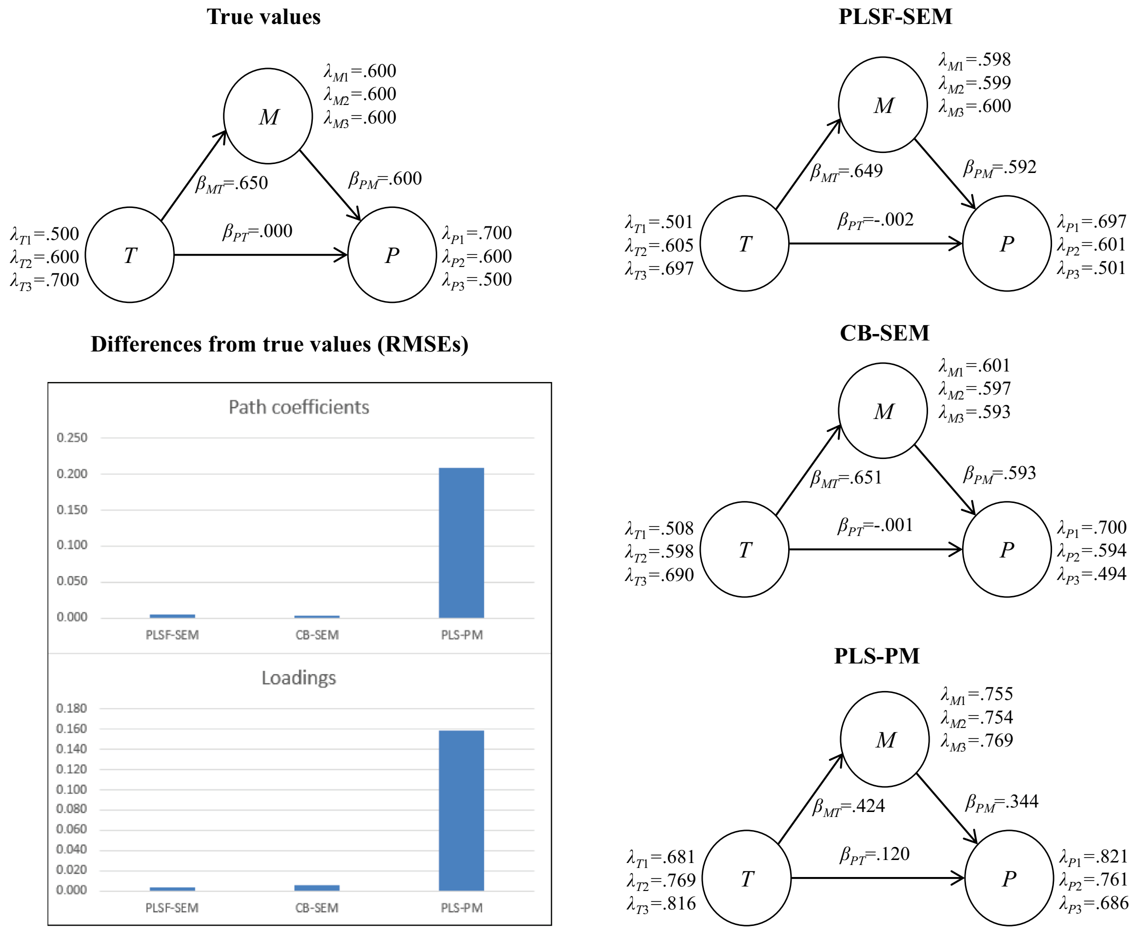


Figure 3. True values and estimates for large sample (N = 10,000). We used an N = 10,000 to minimize sampling bias (i.e., estimates are close to the asymptotic convergence values); true values are shown on the top-left; estimates obtained with each method are shown on the right; the RMSEs aggregate differences among estimates and true values.

would pass widely accepted validity and reliability criteria.

As we can see, the PLSF-SEM and CB-SEM methods present virtually the same performance in terms of similarity with the true values of the estimates of path coefficients and loadings generated by the two methods (for a more elaborate validation, see: Kock, 2019a). The PLS-PM method performs very poorly in this respect, notably: underestimating the path coefficients associated with the two strong effects; overestimating the path coefficient associated with the non-existent (i.e., zero) effect; and overestimating the loadings.

Figure 4 shows the standard errors for the three methods, for the path $\beta_{MT} = .650$ and a sample size of 300; this is a typical sample size used in IS research. The standard error associated with a parameter estimate generated by a specific method is a measure of the variability of that parameter estimate across samples. In our illustrative model, it was generated as the standard deviation for the path in question across

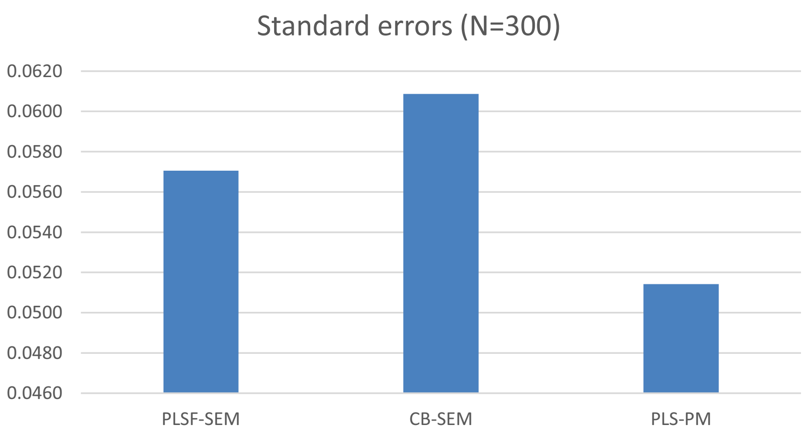


Figure 4. Standard errors (N = 300). For path $\beta_{MT} = .650$.

1,000 samples generated through a Monte Carlo simulation, computed directly from the samples, where the simulation used the true values as the population values. As noted earlier, the standard error for any parameter is likely to go up with the computational complexity of the method, other things being equal. Typically, it is inversely related to the statistical power of the method, or its "sensitivity". The higher the power, then the less likely it is that the method will not recognise non-zero effects, and thus yield false negatives.

It can be seen that the PLSF-SEM method has a lower standard error than CB-SEM, and the PLS-PM method has the lowest standard error among all three methods (see: Kock, 2019a). This means that PLSF-SEM has greater statistical power than CB-SEM. Even though PLS-PM yields the lowest standard errors, the power of this method is about the same as that of PLSF-SEM, because it underestimates path coefficients associated with non-zero effects. This underestimation offsets the benefits from its low computational complexity. Since only PLSF-SEM and CB-SEM are statistically consistent, asymptotically converging to the true values, it follows that PLSF-SEM is the most statistically efficient of the two. It converges to the true values faster.

Note that the zero effect is overestimated by the PLS-PM method as .120, while PLSF-SEM and CB-SEM correctly estimate it at close to .000. This pattern happens consistently across sample sizes (Kock, 2019a, b), and creates a major problem for PLS-PM – false positives. While PLSF-SEM and CB-SEM yield false positives below 5 percent across samples of various sizes, the percentage of false positives yielded by PLS-PM goes up as sample sizes increase. This is illustrated in Figure 5.

As we can see, with a sample size of 253, the probability that PLS-PM will yield a false positive in our model is approximately 60 percent (much higher than the acceptable level of 5 percent). If the sample size is increased to 326, the probability of a false positive goes up to 70 percent. With a sample size a bit over 1,000 a false positive is almost certain to be committed. In other words, the larger the sample, the more likely it is that PLS-PM will lead researchers to find support for hypotheses associated with effects that do not exist in reality. This is exactly the opposite to what one would expect from a trustworthy analysis method.

This remarkably problematic situation has a relatively simple cause. Since PLS-PM does not converge to the true values as sample sizes increase, like CB-SEM and PLSF-SEM do, it will converge to a value that is different from zero for paths that are actually zero at the population level. In our model, the path coefficient for the zero path will be overestimated by PLS-PM as approximately .120. Given that standard errors go down as sample sizes go up, the ratio between .120 and the standard error associated with the zero path in our model will progressively go up as sample sizes increase. As this ratio goes up, the probability that a type I error, or false positive, will be committed also goes up. This occurs whether hypothesis testing relies on p values or confidence intervals.

The illustrative model results presented above are complemented by a set of comparable results, included here with permission, taken from the full-blown Monte Carlo simulation conducted by Kock (2019a). The simulation employs the approach discussed by Kock (2016), where the author provides a detailed step-by-step description of how to create and use simulated data (which can be seen as a pseudocode that can be implemented by other researchers). The simulations compare a slightly broader range of methods: PLSF-SEM, CB-SEM through full-information maximum likelihood, ordinary least squares regression with summed indicators, and PLS-PM. They are included in Appendix B (path coefficients, loadings, and weights), Appendix C (full collinearity variance inflation factors), Appendix D (standard errors), and Appendix E (false positives). The underlying data is the same as that reported by Kock (2019a), but has been re-organised in order to provide insights that are

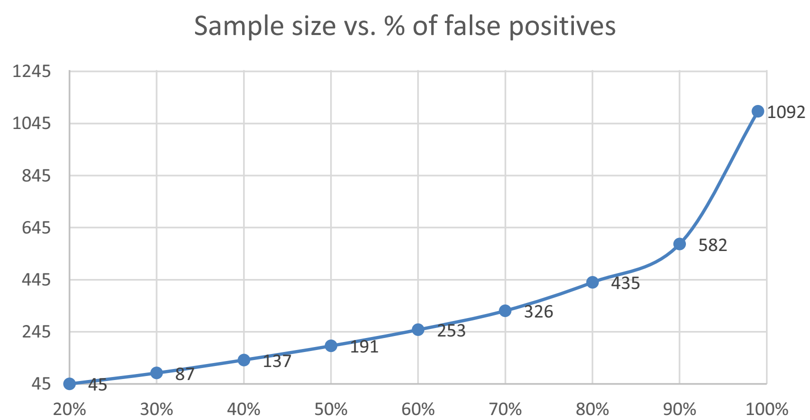


Figure 5. Sample sizes and percentages of false positives for PLS-PM. For path $\beta_{PT} = .000$; vertical axis = sample sizes; horizontal axis = percentages of false positives.

both new and directly related to the discussion above. As the reader will see, the results presented in these appendices both support and enhance the results discussed above.

9. Composite-based models?

If indicators were to cause their LVs, and account for all of their variance, then the resulting models would be composite-based. There are two main possible alternatives for these types of models. The first would have independent indicators, which could be seen as the "ideal" of formative measurement, as there would be no redundancy among the indicators. The second alternative would have indicators depending on a common variable, and thus sharing a certain level of redundancy. These two alternatives are illustrated in Figure 6, which also highlights the paradoxical nature of composite-based models.

If we have a composite-based model with independent indicators, the correlations among any pair of composites would be zero, because the composites would aggregate uncorrelated indicators. Therefore, all path coefficients in such a model would also be zero. Such a model would be of little use in testing any theoretical framework. Any effects at the population level that were hypothesised to be non-zero effects would be associated with zero path coefficients, thus possibly leading to type II errors (false negatives) induced by the misguided use of uncorrelated indicators.

On the other hand, if the indicators were to depend on a common variable in a composite-based model, where they are not caused by their LVs, we would then arguably have a situation resembling common method bias (MacKenzie & Podsakoff, 2012), even though common method bias could be produced by other issues. The non-zero path coefficients in such a model would differ from zero only because of the variation coming from the common variable. The non-zero path coefficients would not be due to the structural relationships. Here, any nonexistent effects at the population level that were included in the model (presumably by mistake) would be associated with non-zero path coefficients, thus leading to type I errors (false positives).

Additionally, no combination of weights at the population level would be recoverable, because an infinite number of combinations would lead to the same results. Let us say, for instance, that we had a composite that aggregated three indicators ($C_i = x_{i1}\omega_{i1} + x_{i2}\omega_{i2} + x_{i3}\omega_{i3}$) where the true weights were $\omega_{i1} = .267$, $\omega_{i2} = .535$, and $\omega_{i3} = .802$. Since the indicators would be either independent or dependent on an unknown common variable, the weights would not be recoverable. That is, no algorithm would converge to the true values of the weights. Since composites are used, composite-based algorithms (e.g., PLS-PM) may do better than factor-based algorithms (e.g., PLSF-SEM), but both would yield incorrect approximations.

10. Analytic composites

Indices, such as the Dow Jones Industrial Average, are widely used in business. The Dow Jones Industrial Average has increasingly been dependent on technology companies, which provide the third highest percentage of its composition, behind financial services and industrial companies. Indices like the Dow Jones

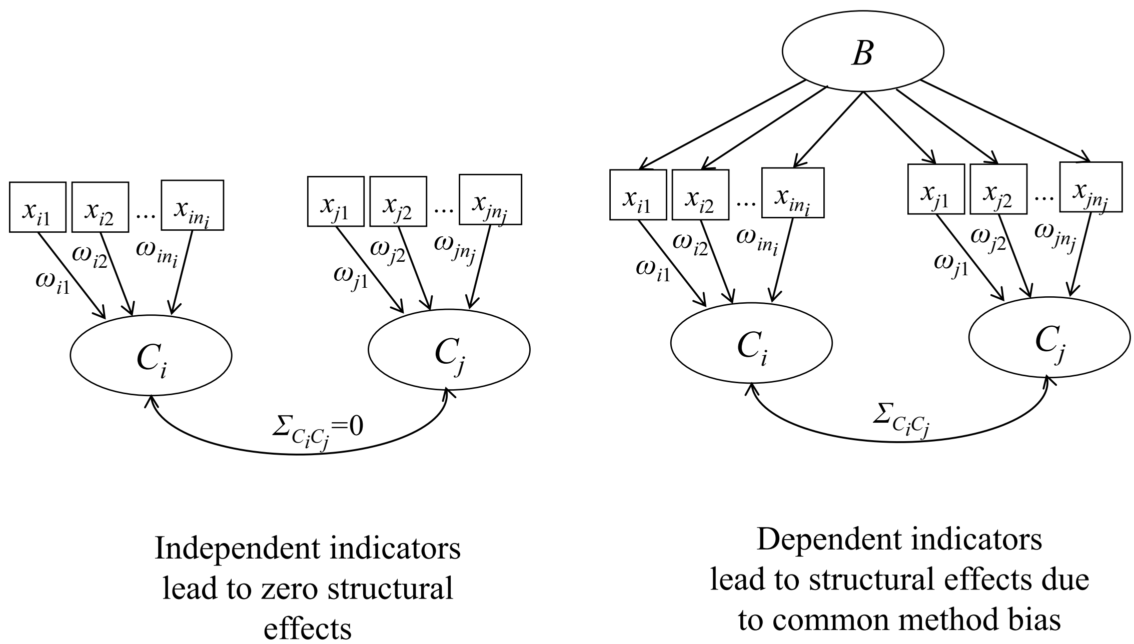


Figure 6. The paradox of composite-based models.

Industrial Average could be seen as composites, given that they are aggregations of indicators.

Let us revisit our example, where we had a composite that aggregated three indicators ($C_i = x_{i1}\omega_{i1} + x_{i2}\omega_{i2} + x_{i3}\omega_{i3}$) where the true weights were $\omega_{i1} = .267$, $\omega_{i2} = .535$, and $\omega_{i3} = .802$. That is, the second weight is approximately twice the first weight, and the third weight approximately three times the first weight. In this example, the composite was designed in a particular way, presumably to serve a purpose. For instance, theory could have suggested this particular configuration of weights for the composite.

We refer to this type of composite as an "analytic" composite, because the weights of the indicators are what they are by design, not because an algorithm estimated them. We avoid the term "index", instead using the term analytic composite, because of the model fit connotation of the former term in SEM. With analytic composites, there is no need to "discover" the weights, because they are known, by definition (the researcher sets them). Whether the configuration of weights of a particularly analytic composite is appropriate or not depends on the purpose of the composite.

The purpose of the Dow Jones Industrial Average, which is based on 30 companies and thus has 30 indicators, has traditionally been to provide a representation of the overall U.S. stock market. However, the perception that it provides an inadequate representation has led to the development of more comprehensive indices, such as the S&P 500 and the Russell 3000 indices; following 500 and 3000 companies, respectively.

That is, analytic composites like the Dow Jones Industrial Average may, or may not, fulfill the needs for which they were originally created. Notably, both, the S&P 500 and the Russell 3000, have technology companies making up the largest percentage of their composition, and thus arguably better reflect the visible dominance of technology companies in the U.S. economy than the Dow Jones Industrial Average.

The above is an example of the use of secondary data to create composites, which is different from the use of questionnaire-based data, and apparently a case in which the indicators both "form" and "cause" the analytic composite. We prefer to view this case as one in which the indicators are used to design the analytic composite. Also, here the indicators have not been created beforehand to measure the analytic composite; they existed irrespective of the composite. In cases like this, there is no right combination of weights that can be discovered by software. In analytic composites weights are defined beforehand by the user of the composite to achieve a purpose. For example, there is an equal-weight version of the Dow Jones Industrial Average that is used as a basis for related exchange-traded funds (e.g., EDOW) that are favoured by some investors.

11. Prediction

One recent trend in the composite-based PLS-PM literature is to present this method as being particularly useful for prediction (Shmueli et al., 2016). Frequently this is interpreted as a justification for recommendations of its use in SEM. The problem is that prediction and SEM are very different types of data analysis applications. Perhaps, the most fundamental difference is that prediction aims at extrapolating parameters from one empirical sample to another, where the samples are often similar to one another. SEM aims at extrapolating parameters from one empirical sample to the entire population from which that sample was taken. This is illustrated in Figure 7.

Any quantitative method that is particularly good at prediction can obviously be put to the test in the stock market, where precise historical data is abundantly available. Even if moderately successful, the prospective gains from using the method in the context of investing would be enormous. For example, an individual investor who could use predictive models to obtain a gain of only 0.5% per day in the stock market would theoretically be able to turn an initial investment of 10 thousand dollars into almost 2 trillion dollars in approximately 15 years. (At the time of this writing, the wealthiest person in the world had a net worth of a little under 250 billion U.S. dollars.) While there is potential, we are unaware of any successful application of PLS-PM in the context of stock market prediction.

Assessing whether PLS-PM is particularly good at prediction is beyond the scope of this paper. Nevertheless, we can safely say that prediction bears little resemblance to SEM. One possible application in the stock market would be to develop a model with PLS-PM that would predict the price of a security (e.g., one share of a bitcoin fund) based on the price of other securities (e.g., shares of gold and silver funds, as indicators of a predictor LV). Here prices in the past year could be used for prediction of prices in the following month, as suggested by the model, leading to buy and sell (or sell short) stock trades.

Typically, as illustrated above, the base sample used for prediction is quite similar to the sample being predicted (although that is not always the case). In SEM, on the other hand, the analysis of an empirical sample is used to test a structural model summarising elements of a theory that explains the behaviour of an entire population – where the population is usually much larger than the sample analysed. The notions of type I and II errors, statistical power, minimum required sample size, and statistical significance,

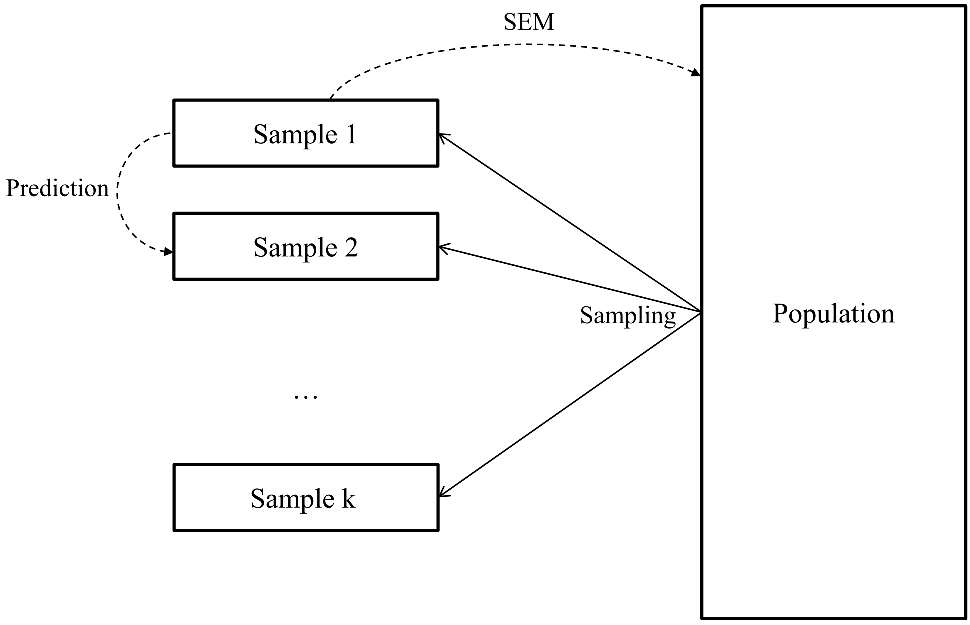


Figure 7. Prediction versus SEM. Sampling indicated via full arrows; extrapolation indicated via dashed arrows.

among others, are well defined in the context of SEM. However, they would have to be re-defined, if at all used, in the context of prediction. In spite of this, composite-based PLS-PM techniques aimed at prediction are increasingly used to justify the use of this method for SEM. This is, in our view, highly problematic.

12. Discussion

The main underlying reason behind the problematic nature of composite-based PLS-PM, when used in the context of SEM, is that question-statements associated with indicators of an LV cannot be devised without the mental idea referring to the LV existing in the mind of the researcher who creates the question-statements. If no mental idea exists, then meaningless question-statements could conceivably be grouped into an LV. But if the question-statements are meaningless, then their indicators would be uncorrelated with other indicators.

In this scenario, the indicators of an LV would be correlated with the LV, which could be seen as "emerging" from the indicators, but would not be correlated with indicators of other LVs. Therefore, if LVs are correlated, the correlation would not be due to the structural model, which usually summarises elements of a theory. That is, any non-zero associations among LVs, which are typically needed for theory-testing, would be due to common sources of variation other than the structural model; which characterises common method bias. The bottom line: no theory could be appropriately tested with such emerging LVs.

The structural model, which specifies the links among LVs, should not be conflated with the measurement model. The measurement model specifies LV-indicator links. Since indicators are supposed to measure LVs with error, they cannot exist before the mental ideas associated with the LVs. In other words, a measure of something cannot exist before that "something" that it is supposed to measure. In SEM question-statements are normally answered on Likert-type scales. Thus, question-statements associated with LVs are guaranteed to measure the LVs with error, because the true LVs at the population level are assumed to exist on ratio scales. What makes SEM unique and extremely useful is that it allows for the estimation of parameters as if there was no measurement error.

12.1. Behavioral and design constructs

How do our views and arguments above, which summarise related discussions presented throughout this paper, fit with the notions that latent constructs (represented by LVs in SEM models) can be modelled as either behavioural or design constructs (see: Benitez et al., 2020; Henseler, 2020; Schuberth et al., 2018), with behavioural or design constructs being modelled in ways that depend on whether they are respectively reflective or formative?

Müller et al. (2018) distinguish between behavioural and design constructs by noting that behavioural constructs are modelled as reflectively measured LVs typically used in the behavioural sciences, and design constructs as formatively measured LVs modelled as composites and typically used in disciplines where "artifacts" (e.g., an index used in finance or economics) are the target of research. Within the framework proposed in this paper, behavioural constructs are best operationalised as factors, and design constructs as analytic composites.

In our view, a behavioural construct should be modelled as factor-based, whether reflective or formative measurement is employed, if the measures employed are devised to quantify the construct (which is usually the case in SEM). In other words, if a designer of a set of question-statements has the construct's idea in mind, then the direction of causality is LV > indicators. That is, the LV should be viewed as a factor-based representation of the construct. The fundamental reason for this is that the construct must exist in the researcher's mind before the question-statements are created to measure it, in both reflective and formative measurement, thus having temporal precedence over the question-statements (a precondition of being a cause). Reflective measurement employs redundant question-statements, and formative measurement employs non-redundant question-statements.

As noted earlier, a researcher who wants to measure "job satisfaction" could use question-statements like "I like my job" and "my job is great", which would implement reflective measurement because respondents would tend to provide highly correlated answers to these redundant question-statements. On the other hand, the researcher may use question-statements like "I like my boss" and "I like my office", which would entail formative measurement. Highly correlated answers to these question-statements would not be expected; respondents may like their bosses but not their offices, and vice-versa.

Now, with respect to design constructs, our view is that they can indeed be employed in SEM models, but not as LVs in the usual SEM analysis sense (where an SEM method is used to "discover" weights or loadings). Composites used in finance, such as the Dow Jones Industrial Average, could be seen as implementations of design constructs, and such measures are widely used in business. Measures like the Dow Jones Industrial Average could be seen as composites, given that they are aggregations of indicators. We refer to this type of composite as an "analytic" composite, because the weights of the indicators are what they are by design. We avoid the term "index", instead using the term analytic composite, because of the model fit connotation of the former term in SEM.

With analytic composites, which are artefacts designed for a purpose, there is no need for an SEM software to "discover" the weights needed for aggregating the indicators into the composites; because those weights are known, as they are set by design. Therefore, we would not expect an SEM software to treat analytic composites as LVs measured indirectly through multiple indicators, even though the analytic composites could be included in SEM models as single-indicator variables (which are not technically LVs).

Whether the configuration of indicators and weights of a particularly analytic composite is appropriate or not depends on the purpose of the composite, and the theoretical framework used as a basis for its design. As we pointed out earlier in this paper, the purpose of the Dow Jones Industrial Average, which is based on 30 companies and thus has 30 indicators, has traditionally been to provide a representation of the overall U.S. stock market. Nevertheless, the view that it provides an inadequate representation has, in part due to the underrepresentation of technology companies, led to the development of more comprehensive analytic composites like the S&P 500 and the Russell 3000.

12.2. Why do we need estimates of LVs?

If CB-SEM has been successfully used for so many years to test models without estimating LVs, why do we need estimates of LVs now? The answer to this question is at the core of the popularity of PLS-PM; this method is popular in no small measure due to the fact that it generates approximations of LVs. This in turn leads to a dramatic growth in the number of tests that can be easily and directly implemented through software tools that automate the method.

For example, without LV scores it is very cumbersome to conduct and interpret the results of moderating effects tests, which become very simple to conduct and visualise in two-dimensional and three-dimensional graphs if those scores are available (see Figure 8). The problem with PLS-PM is that it relies on composites, which are approximations of LVs, thus yielding biased parameters (e.g., path coefficients for moderating links). Hence the need for factor-based methods such as PLSF-SEM, which generate LV estimates (not only approximations).

At this point a reader may point out that LV scores can be generated based on the final results of a regular CB-SEM analysis. This is in fact an area where much research has been conducted, including research by André Beauducel, one of the world's leading experts in LV score estimation methods (Beauducel & Herzberg, 2006; Beauducel & Hilger, 2022). Several unrefined and refined methods exist to do so; among the refined ones that are implemented in software tools are the Thurstone and Bartlett methods (Bartlett, 1937; DiStefano et al., 2009; Hershberger, 2005; Thurstone, 1935). The drawback here is that these methods seem to also yield low quality approximations of LVs. In fact, they seem to be in some cases of lower quality than the LV approximations produced by PLS-PM and ordinary least squares regression with summed indicators (see Appendix C).

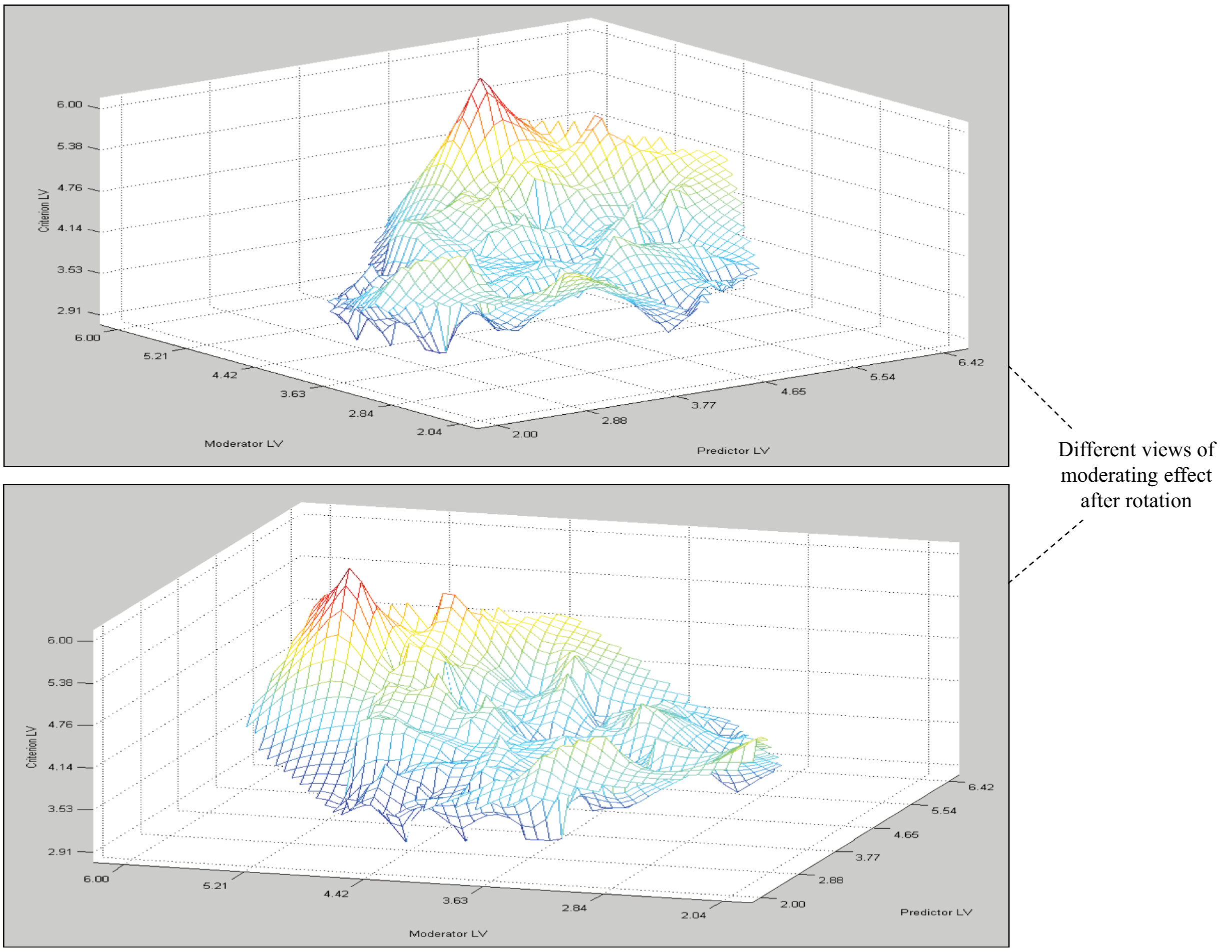


Figure 8. These graphs of moderating effects cannot be generated without LV scores. These graphs are among many that can be generated based on LV scores, helping with the visualization of complex effects.

12.3. Reproducibility of results in PLSF-SEM

As noted earlier, in PLSF-SEM independent and identically distributed random variables that stand in for the measurement residuals are initially created for each of the LVs. The method assigns weights to these uncorrelated measurement residuals based on the consistent PLS equations. With these, and initial estimates of “special” composites that satisfy a small set of equations, the method proceeds to the generation of the final correlation-preserving factors. It could be argued that the creation and use of random variables would lead to irreproducibility of results in PLSF-SEM.

In other words, users of software implementations of PLSF-SEM may obtain slightly different parameter estimates each time they analyse an empirical dataset, which is undesirable. This problem can be easily solved by setting the random seed to a fixed value prior to creating identically distributed random variables that stand in for the measurement residuals. This essentially avoids different results each time an analysis is conducted with the same model and empirical data, making the results reproducible. Stochastic methods that rely on the creation of random variables typically require this type of solution.

12.4. Estimates of LVs and their use outside an SEM analysis

While we have argued in this paper that the LV estimates produced by PLSF-SEM are invaluable in the test of a structural model, which aims at summarising elements of a theoretical model, we do not feel as strongly about the use of those LV estimates outside the SEM analysis in which they are employed (e.g., to produce estimates of various other parameters).

The parameters estimated via PLSF-SEM for an empirical sample are used to make assumptions about the population from which the sample was presumably drawn. However, the LV estimates are not assumed to be exactly the ones in the population. Possibly this is a limitation that also applies, in general terms, to the various parameters estimated in the context of an SEM analysis.

Using parameters outside the context of the SEM analysis that yielded them, based on a given sample, in the context of a second sample, would be better aligned with what we have discussed earlier in this paper as a prediction application. As we argued earlier, it is our belief that prediction bears little resemblance to SEM. In SEM a researcher would normally test hypotheses based on an empirical sample, and extrapolate the results of those tests to a population that is assumed to include the empirical sample. The results of these tests are typically in the form of support, or not, for each of the hypotheses being considered.

12.5. Comparing PLSF-SEM with other SEM methods

Table 1 provides a side-by-side comparison of PLSF-SEM, CB-SEM, PLS-PM, and PLSc. The latter, also known as the consistent PLS technique, is not actually a full parameter estimation technique. It is a technique that relies on results from a PLS-PM analysis employing the centroid scheme, which are then used to correct certain parameters. To the best of our knowledge, two corrected classes of parameters are produced by PLSc, namely reliabilities and loadings. These are used by PLSF-SEM early in its formulation, as a basis for LV estimation. Once the LV estimates are obtained by PLSF-SEM, all of the parameters are re-estimated, including reliabilities and loadings. In practice, the estimates of reliabilities and loadings produced by PLSF-SEM are very close in value to those yielded by PLSc.

It appears, from this comparative analysis, that PLSF-SEM should always be preferred over CB-SEM, PLS-PM and PLSc. The exception to this general rule of thumb is when common method bias has been identified as existing in an empirical dataset, and must be controlled for, in which case CB-SEM maybe be advisable in some cases. That is, if other more straightforward techniques cannot be effectively used, such as the technique of common structural variation reduction (Kock, 2021). We discuss this further below, at the end of this subsection.

Only PLSF-SEM, CB-SEM, and PLSc conduct consistent parameter estimation. Only PLSF-SEM and PLS-PM conduct LV estimation, although PLS-PM actually yields approximations of LVs as composites. Some authors differentiate between estimates and approximations, using the latter term for estimates that are known to be biased (as is the case with PLS-PM). PLSc does not yield LV estimates, other than the ones produced by the PLS-PM analysis employing the centroid scheme. Those are actually de-coupled from the parameters corrected by PLSc.

Only PLSF-SEM and PLS-PM allow for moderating effects estimation, because the LV estimates are needed to build the interaction variables employed in moderating effects estimation. Only PLSF-SEM and PLS-PM allow for nonlinear (e.g., quadratic) relationship estimation, because the LV estimates are needed for that. PLSF-SEM, CB-SEM, and PLS-PM allow for endogeneity testing and control. That is done via instrumental variables by PLSF-SEM and PLS-PM, with the latter yielding biased results. In CB-SEM, covariances among structural errors and predictors of LVs can be included as parameters to be estimated, thus allowing for endogeneity testing and control to be conducted more directly.

PLSF-SEM, CB-SEM, and PLS-PM allow for common method bias testing. In PLSF-SEM and PLS-PM that can be accomplished through tests such as the full-collinearity variance inflation factors test (Kock & Lynn, 2012), and Harman's single factor test Kock (2021a). In CB-SEM common method bias testing can be accomplished by including covariances among indicator error terms as parameters to be estimated.

Only CB-SEM allows for common method bias control, which is accomplished by the researcher including covariances among indicator error terms as parameters to be estimated. This is the key advantage of CB-SEM over PLSF-SEM. However, this is a theoretical advantage, because common method bias tends to affect multiple indicators. In practice, this tends to cause the number of parameters to be estimated to become too large – resulting in model identification problems. That is, the number of parameters to be estimated, which include covariances among indicator error terms, exceeds the information provided by the observed data. This is typically characterised by the Hessian matrix of second order partial derivatives not being invertible.

12.6. Can indicators be the ingredients that form a composite?

If indicators are question-statements designed to measure a construct, our view is that the construct

Table 1. Comparing PLSF-SEM with other SEM methods.

Feature	PLSF-SEM	CB-SEM	PLS-PM	PLSc
Consistent parameter estimation	Yes	Yes	No	Yes
LV estimation	Yes	No	Yes	No
Moderating effects estimation	Yes	No	Yes	No
Nonlinear relationship estimation	Yes	No	Yes	No
Endogeneity testing	Yes	Yes	Yes	No
Endogeneity control	Yes	Yes	Yes	No
Common method bias testing	Yes	Yes	Yes	No
Common method bias control	No	Yes	No	No

PLSc = consistent PLS technique (included for completeness); PLSc is a parameter correction technique that relies on PLS-PM employing the centroid scheme; PLS-PM actually yields approximations of LVs, as composites, not estimates.

(operationalised as an LV) always causes the indicators. If those indicators are seen as weakly correlated "ingredients" that form an LV, then formative LV measurement may be employed. But the operationalisation of the LV is as a factor, not as a composite. This will lead to the emergence of a measurement residual, which accounts for the variance in the LV that is not accounted for by the indicators. That is, the reliability of the LV will be lower than 1.

Note that, for a question-statement associated with an indicator to be created without any construct in mind, it would have to be something like "I like whatever" or "&fa*y j%3$2"; to be answered on a Likert-type scale (e.g., going from "strongly disagree" to "strongly agree"). That is, the question-statement would have to be rather meaningless. If a question-statement has a clear meaning, then the creator of the question-statement has a construct in mind, even if that person does not make a conscious effort to think of a construct. Such question-statements could be aggregated into a formative LV, with the LV modelled as a factor (not a composite). It is our position that question-statements that have a clear meaning always have a mental construct to which they belong, even if the creator (or creators) of the question-statements does not consciously think of them as belonging to any construct.

13. Conclusion

In this paper, we discussed various forms of SEM, and demonstrated that composite-based models are poorly aligned with the idea of measurement with error. We discussed the recently developed factor-based PLSF-SEM method, a new form of SEM that builds on PLS algorithms to generate correlation-preserving factors. We showed that PLSF-SEM is, like CB-SEM, statistically consistent, asymptotically converging to the true values. We also showed that PLSF-SEM is the most statistically efficient of the two, as it converges to the true values "faster"; i.e., with smaller sample sizes.

The above statistical efficiency property provides a strong basis for the continued use and refinement of the PLSF-SEM method. Our key argument put forth in this paper was that PLS-based methods will have to become factor-based, like the PLSF-SEM method is, to survive and thrive in the context of SEM. We have also expressed concern about the recent trend in the composite-based PLS-PM literature of presenting this method as being particularly useful for prediction, as a justification for recommendations of its use in SEM; primarily because prediction and SEM are very different from one another.

Could PLSF-SEM be a viable replacement for CB-SEM as well? We believe that the answer to this question is "yes". Not only is PLSF-SEM more statistically efficient than CB-SEM, but it also presents another advantage. The advantage is near universal convergence to solutions, a property that it shares with PLS-PM. The inherent complexity of CB-SEM (e.g., concurrent estimation of multiple parameters, generation of Hessian matrices and their inversion) not only leads to standard errors that are higher than those generated by PLSF-SEM; but, also, makes convergence in CB-SEM impossible in many cases, even when theoretical identification criteria are met.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

Amora, J. T. (2021). Convergent validity assessment in PLS-SEM: A loadings-driven approach. *Data Analysis Perspectives Journal*, *2*(3), 1–6.

Bartlett, M. S. (1937). The statistical conception of mental factors. *British Journal of Psychology*, *28*(1), 97–104. https://doi.org/10.1111/j.2044-8295.1937.tb00863.x

Beauducel, A., & Herzberg, P. Y. (2006). On the performance of maximum likelihood versus means and variance adjusted weighted least squares estimation in CFA. *Structural Equation Modeling*, *13*(2), 186–203. https://doi.org/10.1207/s15328007sem1302_2

Beauducel, A., & Hilger, N. (2022). Coefficients of factor score determinacy for mean plausible values of Bayesian factor analysis. *Educational and Psychological Measurement*, *82*(6), 1069–1086. https://doi.org/10.1177/00131644221078960

Benitez, J., Henseler, J., Castillo, A., & Schuberth, F. (2020). How to perform and report an impactful analysis using partial least squares: Guidelines for confirmatory and explanatory IS research. *Information & Management*, *57*(2), 1–16. https://doi.org/10.1016/j.im.2019.05.003

Cenfetelli, R. T., & Bassellier, G. (2009). Interpretation of formative measurement in information systems research. *MIS Quarterly*, *33*(4), 689–707. https://doi.org/10.2307/20650323

Chin, W. W. (1998). Issues and opinion on structural equation modeling. *MIS Quarterly*, *22*(1), vii–xvi.

Dijkstra, T. K., & Schermelleh-Engel, K. (2014). Consistent partial least squares for nonlinear structural equation models. *Psychometrika*, *79*(4), 585–604. https://doi.org/10.1007/s11336-013-9370-0

DiStefano, C., Zhu, M., & Mindrila, D. (2009). Understanding and using factor scores: Considerations for the applied researcher. *Practical Assessment, Research & Evaluation*, *14*(20), 1–11.

Ghasemy, M., Teeroovengadum, V., Becker, J. M., & Ringle, C. M. (2020). This fast car can move faster: A review of PLS-SEM application in higher education research. *Higher Education*, *80*(6), 1121–1152. https://doi.org/10.1007/s10734-020-00534-1

Hair, J. F., Risher, J. J., Sarstedt, M., & Ringle, C. M. (2019). When to use and how to report the results of PLS-SEM. *European Business Review*, *31*(1), 2–24. https://doi.org/10.1108/EBR-11-2018-0203

Hajiheydari, N., & Ashkani, M. (2018). Mobile application user behavior in the developing countries: A survey in Iran. *Information Systems*, *77*(3), 22–33. https://doi.org/10.1016/j.is.2018.05.004

Hardin, A. M., Chang, J. C. J., Fuller, M. A., & Torkzadeh, G. (2011). Formative measurement and academic research: In search of measurement theory. *Educational and Psychological Measurement*, *71*(2), 281–305. https://doi.org/10.1177/0013164410370208
Hardin, A., & Marcoulides, G. A. (2011). A commentary on the use of formative measurement. *Educational and Psychological Measurement*, *71*(5), 753–764. https://doi.org/10.1177/0013164411414270
Henseler, J. (2020). *Composite-based structural equation modeling: Analyzing latent and emergent variables*. Guilford Publications.
Hershberger, S. L. (2005). Factor scores. InB. S. Everitt & D. C. Howell (Eds.), *Encyclopedia of statistics in behavioral science* (pp. 636–644). John Wiley.
Huang, W. (2013). *PLSe: Efficient estimators and tests for partial least squares* [Doctoral dissertation]. University of California.
Kline, R. B. (1998). *Principles and practice of structural equation modeling*. The Guilford Press.
Kock, N. (2016). Non-normality propagation among latent variables and indicators in PLS-SEM simulations. *Journal of Modern Applied Statistical Methods*, *15*(1), 299–315. https://doi.org/10.22237/jmasm/1462076100
Kock, N. (2019a). From composites to factors: Bridging the gap between PLS and covariance-based structural equation modelling. *Information Systems Journal*, *29*(3), 674–706. https://doi.org/10.1111/isj.12228
Kock, N. (2019b). Factor-based structural equation modeling with WarpPLS. *Australasian Marketing Journal*, *27*(1), 57–63. https://doi.org/10.1016/j.ausmj.2019.02.002
Kock, N. (2020). Using indicator correlation fit indices in PLS-SEM: Selecting the algorithm with the best fit. *Data Analysis Perspectives Journal*, *1*(4), 1–4.
Kock, N. (2021a). Common structural variation reduction in PLS-SEM: Replacement analytic composites and the one fourth rule. *Data Analysis Perspectives Journal*, *2*(5), 1–6.
Kock, N. (2021b). Harman's single factor test in PLS-SEM: Checking for common method bias. *Data Analysis Perspectives Journal*, *2*(2), 1–6.
Kock, N., & Hadaya, P. (2018). Minimum sample size estimation in PLS-SEM: The inverse square root and gamma-exponential methods. *Information Systems Journal*, *28*(1), 227–261. https://doi.org/10.1111/isj.12131
Kock, N., & Lynn, G. S. (2012). Lateral collinearity and misleading results in variance-based SEM: An illustration and recommendations. *Journal of the Association for Information Systems*, *13*(7), 546–580. https://doi.org/10.17705/1jais.00302
Lehmann, E. L., & Casella, G. (2006). *Theory of point estimation*. Springer.
Lohmöller, J.-B. (1989). *Latent variable path modeling with partial least squares*. Physica-Verlag.
Lowry, P. B., & Gaskin, J. (2014). Partial least squares (PLS) structural equation modeling (SEM) for building and testing behavioral causal theory: When to choose it and how to use it. *IEEE Transactions on Professional Communication*, *57*(2), 123–146. https://doi.org/10.1109/TPC.2014.2312452
MacKenzie, S. B., & Podsakoff, P. M. (2012). Common method bias in marketing: Causes, mechanisms, and procedural remedies. *Journal of Retailing*, *88*(4), 542–555. https://doi.org/10.1016/j.jretai.2012.08.001
Mahmud, I., Ramayah, T., & Kurnia, S. (2017). To use or not to use: Modelling end user grumbling as user resistance in pre-implementation stage of enterprise resource planning system. *Information Systems*, *69*(2), 164–179. https://doi.org/10.1016/j.is.2017.05.005
Marcoulides, G. A., Chin, W. W., & Saunders, C. (2009). A critical look at partial least squares modeling. *MIS Quarterly*, *33*(1), 171–175. https://doi.org/10.2307/20650283
Marcoulides, G. A., & Saunders, C. (2006). Editor's comments – PLS: A silver bullet? *MIS Quarterly*, *30*(2), iii–ix. https://doi.org/10.2307/25148727
Memon, M. A., Ramayah, T., Cheah, J. H., Ting, H., Chuah, F., & Cham, T. H. (2021). PLS-SEM statistical programs: A review. *Journal of Applied Structural Equation Modeling*, *5*(1), 1–14. https://doi.org/10.47263/JASEM.5(1)06
Müller, T., Schuberth, F., & Henseler, J. (2018). PLS path modeling–a confirmatory approach to study tourism technology and tourist behavior. *Journal of Hospitality & Tourism Technology*, *9*(3), 249–266. https://doi.org/10.1108/JHTT-09-2017-0106
Nunnally, J. C. (1978). *Psychometric theory*. McGraw Hill.
Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory*. McGraw-Hill.
Petter, S. (2018). "Haters gonna hate": PLS and information systems research. *ACM SIGMIS Database: The DATABASE for Advances in Information Systems*, *49*(2), 10–13. https://doi.org/10.1145/3229335.3229337
Petter, S., Straub, D., & Rai, A. (2007). Specifying formative constructs in information systems research. *MIS Quarterly*, *31*(4), 623–656. https://doi.org/10.2307/25148814
Rasoolimanesh, S. M. (2022). Discriminant validity assessment in PLS-SEM: A comprehensive composite-based approach. *Data Analysis Perspectives Journal*, *3*(2), 1–8.
Ringle, C. M., Sarstedt, M., Mitchell, R., & Gudergan, S. P. (2020). Partial least squares structural equation modeling in HRM research. *The International Journal of Human Resource Management*, *31*(12), 1617–1643. https://doi.org/10.1080/09585192.2017.1416655
Sarstedt, M., Hair, J. F., Jr., Nitzl, C., Ringle, C. M., & Howard, M. C. (2020). Beyond a tandem analysis of SEM and PROCESS: Use of PLS-SEM for mediation analyses! *International Journal of Market Research*, *62*(3), 288–299. https://doi.org/10.1177/1470785320915686
Sarstedt, M., Hair, J. F., Ringle, C. M., Thiele, K. O., & Gudergan, S. P. (2016). Estimation issues with PLS and CBSEM: Where the bias lies! *Journal of Business Research*, *69*(10), 3998–4010. https://doi.org/10.1016/j.jbusres.2016.06.007
Schuberth, F., Henseler, J., & Dijkstra, T. K. (2018). Confirmatory composite analysis. *Frontiers in Psychology*, *9*(1), 1–14. https://doi.org/10.3389/fpsyg.2018.02541
Schumacker, R. E., & Lomax, R. G. (1996). *A beginner's guide to structural equation modeling*. Lawrence Erlbaum.
Shmueli, G., Ray, S., Estrada, J. M. V., & Chatla, S. B. (2016). The elephant in the room: Predictive performance of PLS models. *Journal of Business Research*, *69*(10), 4552–4564. https://doi.org/10.1016/j.jbusres.2016.03.049
Thurstone, L. L. (1935). *The vectors of mind*. University of Chicago Press.

Appendix A: Glossary

Composite. A quantification of an LV, calculated as an exact linear combination of the LV's indicators. Composites are often used as estimates of LVs, assuming (incorrectly, in our view) that the LVs are caused by their indicators and that the indicators explain 100 percent of the variance in their LVs. If the LVs cause their indicators, composites are often used as approximations of the LVs. In this case, the LVs are more properly estimated as factors.

Covariance-based SEM (CB-SEM). A factor-based class of SEM methods that provides estimates of population parameters (e.g., path coefficients and loadings) based on empirical samples, without estimating LVs as part of the iterative convergence process.

Factor. A quantification of an LV, where the LV causes its indicators; leading to the emergence of a measurement residual when the LV is regressed on its indicators. (Mathematically, this regression can be done, even though the indicators are caused by the LVs – not the other way around.)

Factor-based SEM employing PLS techniques (PLSF-SEM). A factor-based class of SEM methods that provide estimates of population parameters (e.g., path coefficients and loadings) based on empirical samples, estimating LVs as part of the iterative convergence process. One characteristic of this class of methods is that they allow for a much wider range of parameter estimates to be generated than methods where LVs are not estimated.

Formative measurement. Form of LV measurement that relies on non-redundant and weakly correlated indicators, which typically store responses to non-redundant question-statements on Likert-type scales in questionnaires. For example, a researcher who wants to measure "job satisfaction" could use question-statements like "I like my boss" and "I like my office". This example refers to reflective LV measurement because respondents would tend to provide weakly correlated answers to these question-statements. Even though formative measurement is different from reflective measurement, both forms of measurement give rise to factors (not composites).

Indicator error term. A variable that accounts for the variance in an indicator that is not accounted for by the indicator's LV, when the indicator is regressed on its LV. Under common method bias conditions, indicator error terms are often correlated with other indicator error terms associated with the same LV and other LVs.

Indicator. Variable that measures an LV with error, typically as a response to a question-statement formulated to be answered on a Likert-type scale in a questionnaire.

Latent variables (LVs). Variables that cannot be measured directly without error, and that are represented as mental constructs in structural models tested via SEM. LVs are measured indirectly through other variables, often referred to as manifest variables or indicators, typically as responses to question-statements on Likert-type scales in questionnaires.

Measurement residual. A variable that accounts for the variance in an LV that is not accounted for by the LV's indicators, when the LV is regressed on its indicators. (Mathematically, this regression can be done, even though the indicators are caused by the LVs – not the other way around.) Measurement residuals occur when LVs are quantified as factors, and do not occur when LVs are quantified as composites.

Partial least squares path modeling (PLS-PM). A composite-based class of path modeling methods that provides approximations of population parameters (e.g., path coefficients and loadings) based on empirical samples, while approximating LVs as composites as part of the iterative convergence process.

Reflective measurement. Form of LV measurement that relies on redundant and strongly correlated indicators, which typically store responses to redundant (but different) question-statements on Likert-type scales in questionnaires. For example, a researcher who wants to measure "job satisfaction" could use question-statements like "I like my job" and "my job is great". This example refers to reflective LV measurement because respondents would tend to provide highly correlated answers to these question-statements. Even though reflective measurement is different from formative measurement, both forms of measurement give rise to factors (not composites).

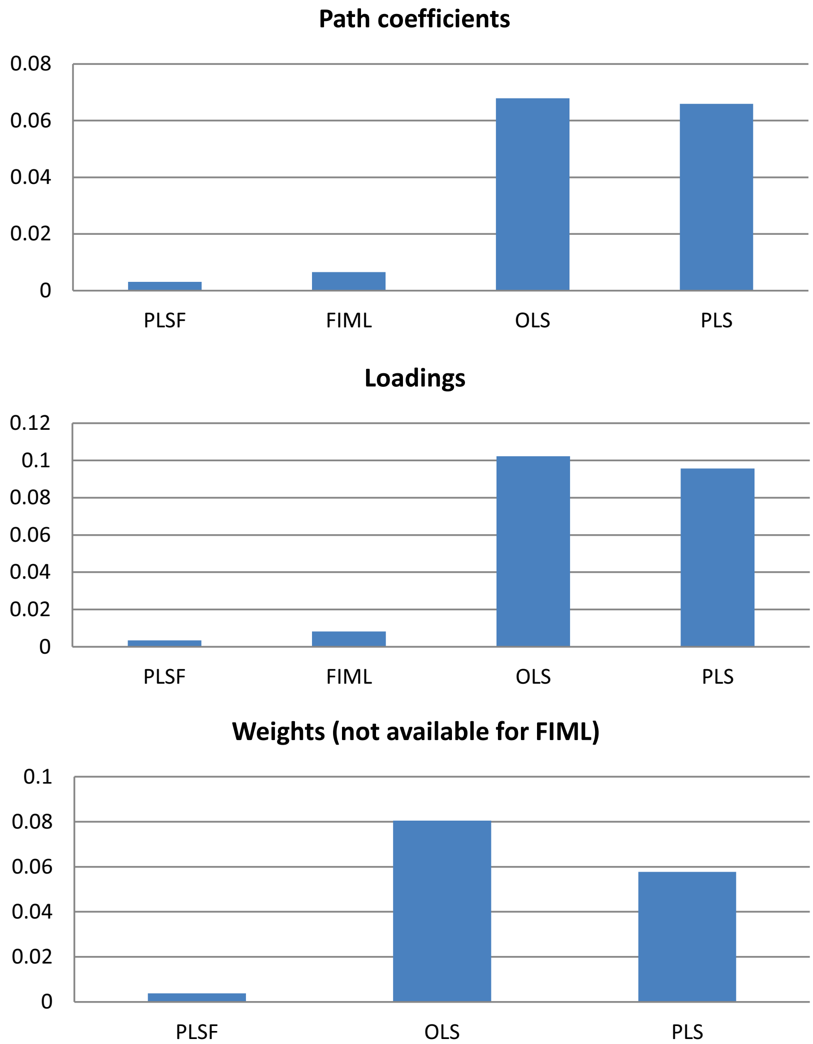
Structural equation modeling (SEM). A data analysis method that allows a researcher to simultaneously test a structural model and a measurement model. The structural model, which aims at summarizing elements of a theoretical model, usually involves a set of LVs; and causal relationships among these LVs, represented through arrows. The measurement model involves links among LVs and indicators.

Structural error term. A variable that accounts for the variance in an LV that is not accounted for by the LV's predictors in a structural model. Those structural model predictors are other LVs, and not indicators.

Appendix B: Path coefficients, loadings, and weights

Figure B1 shows results taken from the full-blown Monte Carlo simulation conducted by Kock (2019a) based on a relatively complex model with 5 LVs and 7 paths with multiple mediating effects. These are based on a large sample (N = 10,000) created based on the true model, to both minimize sampling bias and give us access to a broader range of true parameters (e.g., weights) than are available from the true model. The bar sizes reflect aggregate differences from the true values, measured as RMSEs. The smaller the bars, the better. The analysis methods compared are PLSF=PLSF-SEM; FIML=CB-SEM through full-information maximum likelihood; OLS=ordinary least squares regression with summed indicators; and PLS=PLS-PM.

As we can see, the PLSF and FIML methods had similar performance in terms of the accuracy of estimates of path coefficients and loadings, and performed significantly better than OLS and PLS with regards to those parameters. PLSF also performed quite well in terms of weights, and much better than OLS and PLS. Neither FIML nor other CB-SEM variations typically yield estimates of weights, which can be seen as a limitation in tests that would need weights; such as formative measurement quality assessment tests.



Path coefficients, loadings, and weights. Bar sizes reflect differences from true values, measured as RMSEs. The smaller the bars, the better. PLSF = PLSF-SEM; FIML = CB-SEM through full-information maximum likelihood; OLS = ordinary least squares regression with summed indicators; and PLS = PLS-PM.

Appendix C: Full collinearity variance inflation factors

Similarly to the previous appendix, Figure C1 shows results taken from the full-blown Monte Carlo simulation conducted by Kock (2019a) based on a relatively complex model with 5 LVs and 7 paths with multiple mediating effects. These are based on a large sample (N = 10,000) created based on the true model, to both minimize sampling bias and give us access to a broader range of true parameters (e.g., full collinearity variance inflation factors) than are available from the true model. The bar sizes reflect aggregate differences from the true values, measured as RMSEs. The smaller the bars, the better. The analysis methods compared are PLSF = PLSF-SEM; FIML = CB-SEM through full-information maximum likelihood; OLS = ordinary least

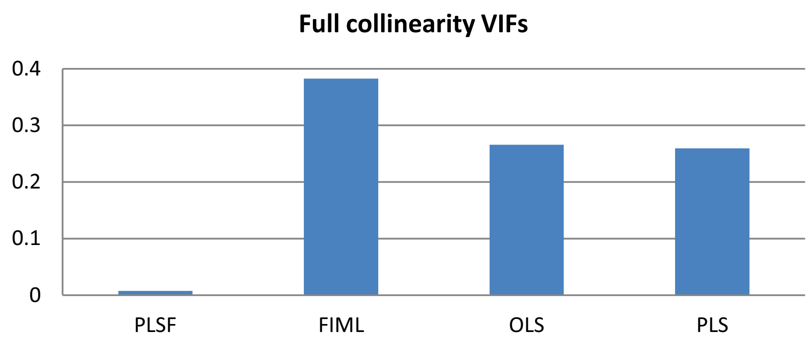


Figure C1. Full collinearity variance inflation factors. Bar sizes reflect differences from true values, measured as RMSEs. The smaller the bars, the better. PLSF = PLSF-SEM; FIML = CB-SEM through full-information maximum likelihood; OLS= ordinary least squares regression with summed indicators; and PLS = PLS-PM. VIFs = variance inflation factors.

squares regression with summed indicators; and PLS = PLS-PM.

Past research has suggested that full collinearity variance inflation factors are rather sensitive parameters, which Kock (2019a) used to assess the quality of correlation-preserving approximations of factor scores. Several unrefined and refined methods exist to generate correlation-preserving approximations of factor scores based on FIML outputs (DiStefano et al., 2009). Kock (2019a) employed two refined methods, the Thurstone and Bartlett methods (DiStefano et al., 2009; Bartlett, 1937; Hershberger, 2005; Thurstone, 1935). Only the Thurstone method yielded viable solutions, which were used to calculate the full collinearity variance inflation factors for FIML. As we can see, PLSF produced the highest quality estimates of correlation-preserving factors.

Appendix D: Standard errors

Similarly to previous appendices, Figure D1 shows results taken from the full-blown Monte Carlo simulation conducted by Kock (2019a) based on a relatively complex model with 5 LVs and 7 paths with multiple mediating effects. The bar sizes are the standard errors; i.e., the standard deviations for the paths whose true values are indicated across 1,000 samples generated through a Monte Carlo simulation, computed directly from the samples, for each of three sample sizes (100, 300, and 500). The analysis methods compared are PLSF = PLSF-SEM; FIML = CB-SEM through full-information maximum likelihood; OLS = ordinary least squares regression with summed indicators; and PLS = PLS-PM.

As we can see, the bars for FIML are consistently larger than those for PLSF, for the three sample sizes employed. This means that there is more dispersion of path coefficient estimates around their average values for FIML than for PLSF. The OLS and PLS methods have even less dispersion, but the problem is that these are less accurate. As shown earlier with respect to accuracy of estimates, the values generated by OLS and PLS are farther from the true values than those yielded by PLSF and FIML.

Appendix E: False positives

Similarly to previous appendices, Figure E1 shows results taken from the full-blown Monte Carlo simulation conducted by Kock (2019a) based on a relatively complex model with 5 LVs and 7 paths with multiple mediating effects. The bar sizes are the percentages of false positives; i.e., instances where a "zero" path coefficient is deemed statistically significant. The percentages are based on 1,000 samples generated through a Monte Carlo simulation, and were computed directly from the samples, for each of three sample sizes (100, 300, and 500). The analysis methods compared are PLSF = PLSF-SEM; FIML = CB-SEM through full-information maximum likelihood; OLS = ordinary least squares regression with summed indicators; and PLS = PLS-PM.

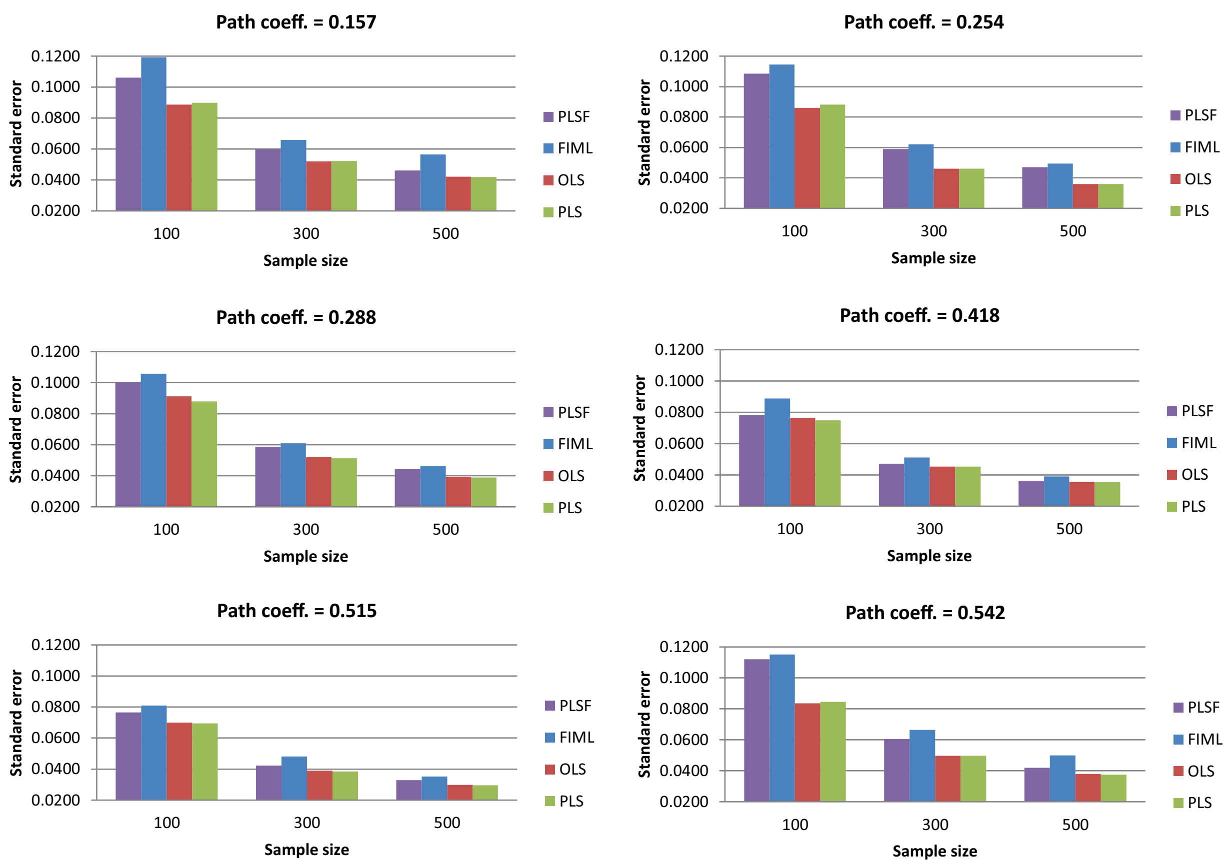
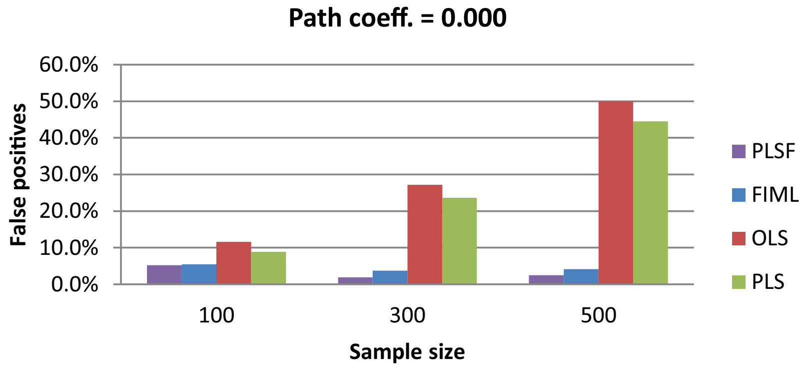


Figure D1 Standard errors. PLSF = PLSF-SEM; FIML = CB-SEM through full-information maximum likelihood; OLS = ordinary least squares regression with summed indicators; and PLS=PLS-PM.

As we can see, the percentages of false positives for PLSF and FIML are about 5% (by convention, the highest acceptable level) for the sample size of 100, and fall below 5% for sample sizes 300 and 500. The OLS and PLS methods, on the other hand, yield progressively higher percentages of false positives as sample sizes go up; exactly what one would like to avoid in SEM analyses. With the sample size of 500, false positives are between 40% and 50% for OLS and PLS.



False positives. Bar sizes reflect the percentages of false positives. The higher the bars, the more cases in which a "zero" path coefficient was identified as statistically significant. PLSF = PLSF-SEM; FIML = CB-SEM through full-information maximum likelihood; OLS = ordinary least squares regression with summed indicators; and PLS = PLS-PM.