Contributing to the Success of PLS in SEM: An Action Research Perspective

Ned Kock

Abstract:

I share with Evermann and Rönkkö (2022) the belief that classic composite-based partial least squares path modeling (PLS-PM) presents shortcomings when used to conduct structural equation modeling (SEM) analyses. The shortcomings can be traced back to one fundamental problem, which is that latent variables (LVs) are approximated in PLS-PM as exact linear combinations of their corresponding indicators. In SEM, each LV is in fact a factor; i.e., a linear combination of the indicators and a measurement residual. My approach to addressing the shortcomings of PLS-PM is rather unique among researchers concerned with quantitative methods. I have employed an action research approach, helping investigators employ SEM in their empirical studies. This has led to my development of a widely used software tool for SEM analyses. I illustrate my action research orientation by discussing three recent methodological developments with which I have been closely involved.

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

1 Introduction

I am thankful for the opportunity to write this rejoinder to the article by Evermann and Rönkkö (2022), which is aimed primarily at information systems (IS) researchers. It should not be surprising to those who are familiar with my work that I share with the authors the belief that classic composite-based partial least squares path modeling (PLS-PM) presents shortcomings when used to conduct structural equation modeling (SEM) analyses. The shortcomings of PLS-PM can be traced back to one fundamental problem, which is that latent variables (LVs) are approximated in PLS-PM as exact linear combinations of their corresponding indicators. In SEM, each LV is in fact a factor; i.e., a linear combination of the indicators and a measurement residual, where the residual accounts for the variance in the LV that is not accounted for by the indicators. Using composites instead of factors to estimate LVs not only leads to biases, but places PLS-PM somewhat at odds with what most methodological researchers define as SEM. Fundamentally, SEM is a factor-based family of related methods.

My main disagreement with Evermann and Rönkkö (2022) is on how to address the problems caused by the use of composites instead of factors to estimate LVs, which are at the source of virtually all of their recommendations. I believe that we should address the source of the problems; not the symptoms, which are their main focal points. Addressing the symptoms leads to a proliferation of tests and recommendations that quickly become overwhelming. Since using composites instead of factors leads to distortions in so many classes of parameters (e.g., loadings, weights, reliability coefficients etc.), these distortions being the symptoms, a separate correction usually has to be devised for each class of parameters that has been distorted. Addressing the source of the problems, by estimating LVs as factors, leads to the removal of the distortions all at once. It should be noted that covariance-based SEM (CB-SEM) does not estimate LVs as part of its iterative convergence process. That is, in my view, a major shortcoming of CB-SEM, even though it is clearly a factor-based method to conduct SEM.

My approach to addressing the shortcomings of PLS-PM is admittedly rather unique among researchers concerned with quantitative methods. As I have done before (Kock, 2004, 2007), I employ an action research approach to address those shortcomings. The main focus of my action research efforts has been to meet the needs of SEM users, which led to my development of WarpPLS (Kock, 2020a), a widely used non-linear SEM software tool first released in 2009. This software implements both composite-based and factor-based algorithms. Its free trial version is a full implementation of the software, not a demo version, and is available for approximately three months. The factor-based algorithms provide the foundation for a set of related methods that are generally referred to here as PLSF-SEM. (Here the "F" refers to the factor-based nature of this SEM method, which builds on PLS-related algorithms.) These methods try to bring together the advantages of composite-based PLS-PM with those of CB-SEM, and without the disadvantages inherent in those techniques.

Action researchers tend to derive many of their research findings from helping their research "clients". WarpPLS users are my main clients, and among those are very sophisticated SEM experts. WarpPLS users in general, and particularly sophisticated expert users, have repeatedly presented me with challenges that pushed me into finding solutions to difficult problems. Among these solutions are new tests to identify lateral collinearity issues and common method bias. The solutions had to be novel, in part because those expert users have used WarpPLS in conjunction with several other SEM software tools, and challenged me to find solutions that were not available from those other tools. I have so far directly helped more than 1,000 users, over a period of a little over 11 years.

In the following sections, I address three recent developments with which I have been closely involved. I believe these developments complement those discussed by the authors in a way that is particularly meaningful to IS researchers, since they have been published recently in IS outlets, including three elite journals in that field: *Information Systems Journal*, *Journal of the Association for Information Systems*, and *Journal of Management Information Systems*. The developments in question refer to lateral collinearity and common method bias assessment, minimum sample size estimation, and factor-based analyses as a complement to composite-based studies. The main publications outlining these developments have had a total of over 3,000 citations so far.

2   Full Collinearity Variance Inflation Factors and Common Method Bias

One mistake that researchers often make is to add two or more redundant LVs to a model, and then connect those LVs with arrows. In this scenario the LVs are redundant in the sense that they measure the same underlying construct. This creates a form of collinearity that we have called "lateral" collinearity (Kock & Lynn, 2012), which is not usually picked up by typical tests of collinearity employing variance inflation factors (VIFs). Traditional collinearity tests address only redundancy among predictor variables, not redundancy among predictor and criteria – such as linked LVs. I have addressed this by developing a new type of measure: full collinearity VIFs (FCVIFs), which are VIFs that are calculated considering all LVs in a model (Kock & Lynn, 2012).

At the time I proposed FCVIFs, it became apparent that they could be used in an effective yet simple test for common method bias. This simple test for common method bias (Kock, 2015b; Kock & Lynn, 2012) relies on the comparison of all of the FCVIFs in a model, one for each LV, against a threshold. There is no need to create a new model, as is the case in other similar tests; e.g., Harman's single factor test (Kock, 2021). If any FCVIF is greater than the threshold, that would be an indication of the existence of common method bias in the model. Based on past research and Monte Carlo simulations I proposed the threshold of 3.3 for models analyzed with composite-based PLS-PM algorithms. Table 1 illustrates this test for a model with three LVs. The row noted as "No CMB" shows the FCVIFs obtained with no common method bias. The row noted as "CMB" shows the FCVIFs obtained with the same model, but with common method bias (i.e., with common method variation contamination added to it).

Table 1. Using FCVIFs to Detect Common Method Bias

	LV1	LV2	LV3
No CMB	1.541	1.472	1.739
CMB	2.619	2.347	3.720
CMB = common method bias.			

As we can see, LV3 has an FCVIF of 3.720, which is greater than 3.3 and thus is an indication of common method bias in the model. This is a good example of my action research orientation, and the desirable outcomes that it generates. The development of FCVIFs was initially aimed at solving a problem faced by researchers employing PLS-PM, namely the need to identify lateral collinearity instances. As their use evolved, FCVIFs proved to be sensitive to a different but related problem – common method bias.

The reason why FCVIFs are useful in the identification of common method bias is that the common method variation that contaminates indicators (and that leads to common method bias) tends to end up in LVs in PLS-PM, because the LVs aggregate indicators. As it turns out, FCVIFs are particularly sensitive to this type of contamination. I discussed this further in an article (Kock, 2015b), which, together with the previous article on lateral collinearity (Kock & Lynn, 2012), has since become widely cited in the context of common method bias assessment.

FCVIFs appear to be useful in the context of factor-based methods, such as PLSF-SEM, although the FCVIF threshold used should arguably be a bit higher (e.g., 5) in this context (Kock, 2015b). This is due to the fact that LVs in factor-based models incorporate more variation from other LVs than in composite-based models; the extra variation being primarily in the measurement residuals contained in the LVs. This extra variation is combined with the common method variation, leading to generally higher FCVIFs. It should be stressed though that only the latter type of variation, common method variation, is pathological.

3   Minimum Sample Size Estimation: The Inverse Square Root and Gamma-Exponential Methods

One of the most controversial issues in PLS-PM has been that of minimum sample size estimation, for which the "10-times rule" has been a favorite due to its ease of use, even though it tends to yield imprecise estimates. According to this rule, the sample size should be greater than 10 times the maximum number of inner or outer model links pointing at any LV in the model. Given my action research orientation, finding an alternative to this rule has long been a target, and I have always wanted to provide a straightforward solution to the problem – as opposed to advising researchers to do very complex tasks, such as conducting sequences of related Monte Carlo simulations, to estimate minimum sample size.

As a result, we have proposed two related methods, based on mathematical equations (an easy-to-use general form is discussed below), as alternatives for minimum sample size estimation in PLS-PM: the inverse square root method, and the gamma-exponential method (Kock & Hadaya, 2018). Based on three Monte Carlo experiments with a relatively complex model containing six LVs and the same number of causal links, whose main results are summarized in Table 2, we demonstrated that both methods are fairly accurate. The methods rely on the minimum absolute significant path coefficient in a model, which assumed the following values in the three Monte Carlo experiments: .397, .147, and .100. Both methods are fully implemented in WarpPLS (Kock, 2020a), and are extensively used by researchers.

Table 2. Performance of Minimum Sample Size Estimation Techniques

Method	Estimate (paths ≥ .397)	Estimate (paths ≥ .147)	Estimate (paths ≥ .100)
Monte Carlo simulation	28	265	599
10-times rule	20	20	20
Inverse square root	40	287	619
Gamma-exponential	26	273	605

The row labeled "Monte Carlo simulation" shows the estimates obtained via the Monte Carlo method, which is a rather laborious method for minimum sample size estimation, and arguably less precise than the gamma-exponential method (for a discussion on this, see Kock & Hadaya, 2018). Note that the 10-times rule underestimates required sample sizes, and grossly so for models with path coefficients as weak as .147 and .100. The inverse square root method slightly overestimates minimum sample sizes, which makes it a conservative method. As shown below, this method is particularly attractive in terms of its simplicity of application. The gamma-exponential method relies on a much more complex equation that can only be solved numerically.

We have proposed a general form of the equation underlying the inverse square root method (see Equation 1) in an article on positivist IS action research (Kock et al., 2017). In this equation $\widehat{N}$ is the sample size estimate; $z_{.95}$ and $z_{.8}$ are the z-scores associated with the values .95 and .8, which assume the use of 95 percent confidence levels (or P values significant at the .05 level) for hypothesis testing and statistical power of 80 percent; and $|\beta|_{min}$ is the minimum significant absolute path coefficient expected (e.g., based on past research) or observed in the model. The equation is general in that it can be customized if one wants to explore results under different confidence levels and statistical power targets.

$$\widehat{N} > \left(\frac{z_{.95} + z_{.8}}{|\beta|_{min}}\right)^2 \tag{1}$$

The Excel function NORMSINV(*x*) can be used to obtain the values for $z_{.95}$ and $z_{.8}$, NORMSINV(.95) and NORMSINV(.8), which are respectively 1.645 and 0.842. A sample size required for a specific model should satisfy Equation 1. For example, assume that the minimum absolute significant path coefficient observed in a model is .2. In this case, a minimum sample size of approximately 155 is needed to achieve a statistical power of 80 percent; calculated as the smallest integer that satisfies Equation 1. Based on citations to the article that proposed the method (Kock & Hadaya, 2018), it appears to have been well received by researchers, which I owe to my action research orientation: to provide a straightforward solution to the problem of minimum sample size estimation in PLS-PM. My further work on this topic suggests that this solution also applies to PLSF-SEM.

Interestingly, my work has led to a conclusion that often surprises researchers: the minimum required sample size does not depend on a model's complexity, but on the minimum absolute significant path coefficient in the model. While typically complex models do indeed have weaker path coefficients, a very simple model with only two connected LVs and a true path coefficient of .100 will require a sample size of around 600 for a statistical power level of 80 percent to be achieved. Note that my focus has been on path coefficients because they are usually associated with hypothesis testing, and thus type I and II errors. Statistical power is the probability that a type II error will be avoided, for each path coefficient, which is why the weakest path coefficient is the one that drives the power of a method used to analyze an entire model with multiple path coefficients.

4   Factor-Based Model Estimation: Going Beyond Composites with PLSF-SEM

In SEM, question-statements are typically devised to be answered on Likert-type scales and included in questionnaires to measure LVs. Consider an LV associated with the concept of "job satisfaction". If reflective measurement is employed, the question-statements would be redundant and somewhat like these: "I like my job" and "My job is great". If formative measurement is used, the question-statements would be non-redundant (i.e., they would refer to different facets of the LV) and be somewhat like these: "I like my boss" and "I like my office".

Indicators are variables that store numeric answers to question-statements associated with an LV. And LVs are assumed to exist, in a latent way, on ratio scales. For this and other reasons, the question-statements associated with an LV clearly measure the LV with error. That is, each indicator is an imprecise measure of its LV, which exists always before the indicators and thus always causes them – whether reflective or formative measurement is used. So, if we regress an indicator on its LV, we will get a loading for that indicator that is lower than 1. For the same reason (i.e., measurement with error), if we regress an LV on its indicators, we will find that the indicators will account for less than 100 percent of the variance in the LV. This percentage (e.g., .85 or 85 percent) is the true reliability of the LV.

So, if the indicators of an LV do not explain 100 percent of its variance, something else has to explain the remaining variance. I have referred to that "something else" as the measurement residual associated with the LV, and developed a generic method later called PLSF-SEM to estimate LVs taking their measurement residuals into consideration (Kock, 2015a, 2019). Initially I have used various approaches to estimate an LV's true reliability, which is a critical element of the method, including Cronbach's alpha (Kock, 2015a). I have later used the consistent PLS technique developed by Theo Dijkstra, which generates a consistent estimate of the true reliability (Kock, 2019).

I have conducted a number of Monte Carlo experiments to assess the performance of the PLSF-SEM method, including its performance in terms of accuracy of estimates, statistical consistency, and statistical efficiency. Figure 1 contrasts the performance of the PLSF-SEM method (noted as "PLSF") against three other methods: CB-SEM via full-information maximum likelihood (noted as "FIML"), ordinary least squares with summed indicators (noted as "OLS"), and PLS-PM (noted as "PLS"). It is based on Monte Carlo experiments building on a reasonably complex model with five LVs and seven causal links. The bars show the root-mean-square errors associated with the deviations from the true values. The smaller the better.
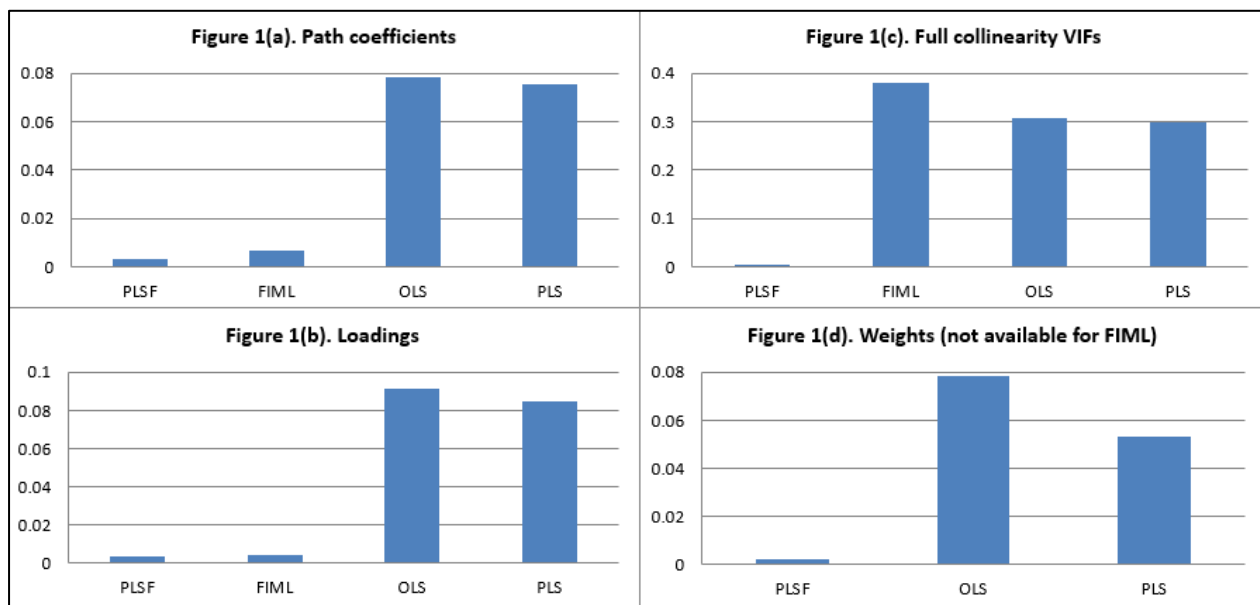

Figure 1. Deviations from true values

As we can see, PLSF-SEM performed as well as CB-SEM in terms of accuracy of path coefficients [Figure 1(a)] and loadings [Figure 1(b)], and much better than the other two methods. Also, PLSF-SEM provided

the most accurate LV estimates among all methods, as indicated by the FCVIFs [Figure 1(c)], which are very sensitive coefficients, as discussed earlier in this article. Based on the FCVIFs, I can see that the LV estimates obtained via CB-SEM (after convergence, using standard techniques) were significantly less accurate than those produced by the other methods. Finally, PLSF-SEM provided the most accurate estimates of weights [Figure 1(d)]; CB-SEM does not produce such estimates.

My results also suggested that while both PLSF-SEM and CB-SEM are statistically consistent methods, in the sense that they both converge to the true values as sample sizes increase, PLSF-SEM yields lower standard errors and thus has greater statistical power. Stated differently, PLSF-SEM is the most statistically efficient among all of the four methods whose performance I compared. One possible reason for this is that PLSF-SEM is computationally simpler than CB-SEM – not requiring, e.g., the calculation and inversion of Hessian matrices of second partial derivatives – and thus PLSF-SEM leads to less propagation of sampling error than CB-SEM.

5   Conclusion

Tackling problems as an action researcher has a number of advantages, one of which is to place one in the "thick of the action" when it comes to addressing SEM user needs. For example, I have had to conduct thousands (literally) of Monte Carlo simulations over the years, to ensure that my solutions are correct, which has given me a much deeper understanding of SEM and related methods than I would have had otherwise. Also, I have developed an SEM software tool, WarpPLS, and helped users to conduct many hundreds of empirical analyses, which allowed me to observe patterns that further improved my understanding of SEM in general. Finally, I have benefitted greatly from close interactions with very sophisticated SEM experts, who have used WarpPLS in conjunction with several other SEM software tools, which helped me develop and implement arguably novel SEM analysis features.

We can illustrate the latter point through a couple of examples. Several years ago, a few SEM experts have asked me to add indices that would allow WarpPLS users to compare model-implied indicator correlation matrices with those obtained from the empirical data, much like CB-SEM software tools do. A key goal was to be able to check whether factor-based PLSF-SEM yielded a better fit with empirical data than composite-based PLS-PM. This would help those experts justify the use of PLSF-SEM, which they wanted to do because PLSF-SEM had been shown to yield lower type I and II errors with simulated data. In response, I developed five new indices (Kock, 2020b), namely the: standardized root mean squared residual (SRMR), standardized mean absolute residual (SMAR), standardized chi-squared (SChS), standardized threshold difference count ratio (STDCR), and standardized threshold difference sum ratio (STDSR).

Soon after incorporating these indices into WarpPLS, I began noticing that PLSF-SEM indeed tended to yield better fit indices than PLS-PM. Now, several years since the new indices were introduced, it is becoming clear that this pattern applies to virtually all empirical datasets. Many users have shared their datasets with me over the years, asking for help, and although I do not keep those datasets, I often inspect various coefficients yielded by different methods and algorithms to better advise users. The fact that virtually all empirical datasets I have seen (well over 100) appear to be factor-based, as indicated by the fit indices, has been reassuring. One could argue that empirical data is often composite-based. Although this claim is illogical, as I have discussed earlier, it is hard to counter a claim based on investigations relying only on simulated data. Besides, those who argue that empirical data is often composite-based can always find issues with how simulated data is created, claiming that if simulated data is created in the "proper" way it will support their claim. Comparative analyses based on real-world empirical data, of which I conducted many, can help settle the issue.

Another example of the contribution of SEM experts to my development of novel solutions refers to a phenomenon known as "Simpson's paradox", which had not been discussed in the context of SEM or path models before (Kock, 2015c; Kock & Gaskins, 2016). Some SEM experts have asked me to add features to WarpPLS that would allow them to identify instances of Simpson's paradox, where path coefficients associated with links and the corresponding correlations have different signs (a very odd and counterintuitive phenomenon). In response, I have developed a table that indicates when Simpson's paradox occurs in connection with links among LVs. I have also developed two new indices of model-wide incidence of the problem: the Simpson's paradox ratio (SPR), and the R-squared contribution ratio (RSCR). I also view this as an important development, because a Simpson's paradox instance is a possible indication of a causality problem, suggesting that a hypothesized causal link is either implausible

or reversed (Kock & Gaskins, 2016). With these coefficients, WarpPLS users can assess the soundness of their models with respect to causality. This is critical in SEM because causally incorrect models may find empirical support, leading to faulty theories.

Acknowledgments

References

Evermann, J., & Rönkkö, M. (2022). Recent developments in PLS. *Communications of the Association for Information Systems.*

Kock, N. (2004). The three threats of action research: A discussion of methodological antidotes in the context of an information systems study. *Decision Support Systems*, 37(2), 265-286.

Kock, N. (Ed.). (2007). *Information systems action research: An applied view of emerging concepts and methods.* Springer.

Kock, N. (2015a). A note on how to conduct a factor-based PLS-SEM analysis. *International Journal of e-Collaboration*, *11*(3), 1-9.

Kock, N. (2015b). Common method bias in PLS-SEM: A full collinearity assessment approach. *International Journal of e-Collaboration*, *11*(4), 1-10.

Kock, N. (2015c). How likely is Simpson's paradox in path models? *International Journal of e-Collaboration*, *11*(1), 1-7.

Kock, N. (2019). From composites to factors: Bridging the gap between PLS and covariance-based structural equation modeling. *Information Systems Journal*, *29*(3), 674-706.

Kock, N. (2020a). *WarpPLS user manual: Version 7.0.* ScriptWarp Systems.

Kock, N. (2020b). 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. (2021). Harman's single factor test in PLS-SEM: Checking for common method bias. *Data Analysis Perspectives Journal*, *2*(2), 1-6.

Kock, N., & Gaskins, L. (2016). Simpson's paradox, moderation, and the emergence of quadratic relationships in path models: An information systems illustration. *International Journal of Applied Nonlinear Science*, *2*(3), 200-234.

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.

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.

Kock, N., Avison, D., & Malaurent, J. (2017). Positivist information systems action research: Methodological issues. *Journal of Management Information Systems*, *34*(3), 754-767.

About the Authors

Ned Kock is Texas A&M Regents Professor, and Chair of the Division of International Business and Technology Studies, in the Sanchez School of Business, at Texas A&M International University. He holds degrees in electronics engineering (B.E.E.), computer science (M.S.), and management (Ph.D.). He has published his research in a number of high-impact journals including *Communications of the ACM*, *Decision Support Systems*, *European Journal of Information Systems*, *European Journal of Operational Research*, *IEEE Transactions (various)*, *Information & Management*, *Information Systems Journal*, *Journal of the Association for Information Systems*, *Journal of Management Information Systems*, *MIS Quarterly*, and *Organization Science*. He is also the developer of WarpPLS, a widely used structural equation modeling software.