Combining composites and factors in PLS-SEM models: A multi-algorithm technique

Ned Kock
Texas A&M International University, USA

Abstract

A multi-algorithm technique is presented for combining latent variables estimated as composites or factors into a single model, in the context of structural equation modeling via partial least squares. The multi-algorithm technique consists of three key steps: selecting composite-based or factor-based outer model analysis algorithms to be used for latent variable estimation; adding the latent variables estimated with the chosen composite-based or factor-based algorithms as new standardized variables; and creating and estimating a final model with the new variables added as single indicators of latent variables.

Keywords: Composites; Factors; Structural Equation Modeling; Partial Least Squares; WarpPLS.

Introduction

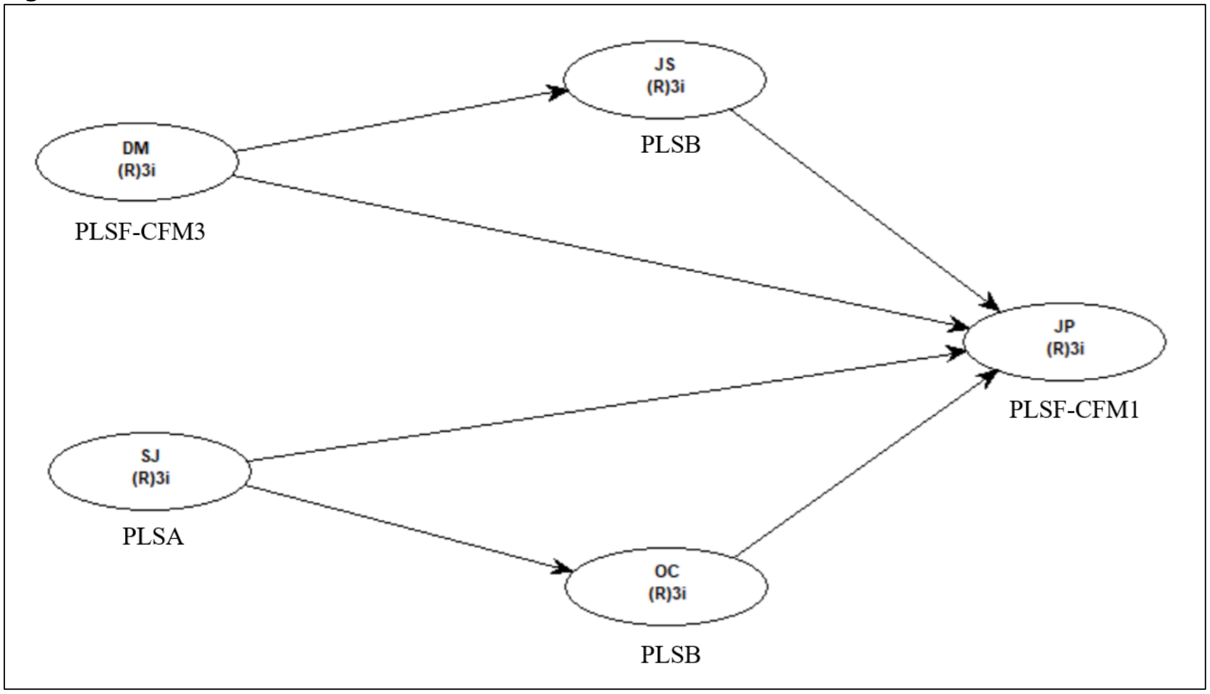
We discuss a multi-algorithm technique for combining latent variables (LVs) estimated as composites and factors in one single model, in the context of structural equation modeling via partial least squares (PLS-SEM). The multi-algorithm technique involves a few primary steps: choosing composite-based or factor-based outer model analysis algorithms to be used for LV estimation; adding the LVs estimated with the chosen composite-based or factor-based outer model analysis algorithms as new standardized variables; and creating and analyzing a model with the new algorithm-specific variables added as single indicators of LVs.

Our discussion is based on an illustrative model analyzed with the software WarpPLS, Version 8.0 (Kock, 2022a). WarpPLS is a widely used SEM software that implements both classic composite-based as well as factor-based PLS-SEM algorithms (Kock, 2019a; 2019b), where LVs can be estimated through various algorithms, among other features that can be useful in advanced SEM analyses (Amora, 2021; 2023; Canatay et al., 2022; Cox, 2024; Hubona & Belkhamza, 2021; Kock, 2015a; 2015b; 2015c; 2016; 2019a; 2019b; 2020a; 2020b; 2020c; 2021a; 2021b; 2021c; 2022a; 2022b; 2022c; 2023a; 2023b; Kock & Gaskins, 2016; Kock & Lynn, 2012; Ma & Zhang, 2023; Moqbel et al., 2020; Morrow & Conger, 2021; Rasoolimanesh, 2022; Samak et al., 2024; Tarkom & Gopal, 2024).

Illustrative model and data

The illustrative model shown in Figure 1 contains two exogenous LVs, namely DM and SJ; and three endogenous LVs, indicated as JS, OC and JP. This illustrative model served as the basis for the creation of a simulated dataset, through the Monte Carlo method (Kock, 2016), with a sample size of 500. Under each LV, we list the composite-based or factor-based outer model algorithm used for LV estimation.

Figure 1: Illustrative model



Notes: DM = democratic management; SJ = scarcity of comparable jobs; JS = job satisfaction; OC = organizational commitment; JP = job performance; notation under LV acronym describes measurement approach and number of indicators, e.g., (R)3i = reflective measurement with 3 indicators.

The outer model analysis algorithms used are Factor-Based PLS Type CFM3 (PLSF-CFM3), Factor-Based PLS Type CFM1 (PLSF-CFM1), PLS Mode A (PLSA), and PLS Mode B (PLSB). The PLSF-CFM3 and PLSF-CFM1 algorithms generate estimates of factors, in two stages, explicitly accounting for measurement error. Like covariance-based SEM algorithms, these algorithms are fully compatible with common factor model assumptions. The PLSA and PLSB algorithms, on the other hand, are classic composite-based PLS algorithms, which do not explicitly account for measurement error.

In their first stages, the PLSF-CFM3 and PLSF-CFM1 algorithms employ a true composite estimation sub-algorithm, which estimates composites based on mathematical equations that follow directly from the common factor model. The second stage employs a variation sharing sub-algorithm, which can be seen as a soft version of the classic expectation-maximization algorithm used in maximum likelihood estimation, with apparently faster convergence and nonparametric properties (Kock, 2019a; 2019b).

The PLSF-CFM3 algorithm employs both loadings and reliabilities from Dijkstra's consistent PLS (a.k.a. PLSc) technique; the former (i.e., loadings) to improve computation efficiency, and the latter (i.e., reliabilities) to estimate measurement error and true composite weights. The PLSF-CFM1 algorithm does not employ Dijkstra's consistent PLS technique at all, instead using Cronbach’s alpha coefficients to estimate measurement error and true composite weights.

The PLSA algorithm is often referred to as the reflective mode of classic composite-based PLS, which is arguably incorrect because both reflective and formative LVs can be used with this algorithm. In it, the inner model influences the outer model through path coefficients and correlations, depending on whether the links go into or out from each LV, respectively. In PLSA the outer model weights are calculated through a least squares regression where the LV is the predictor and the indicators are the criteria.

The PLSB algorithm is often referred to as the formative mode of classic composite-based PLS. Again, this is most likely a misrepresentation based on incorrect assumptions, for the same reason discussed above, namely that both reflective and formative LVs can be used with this algorithm. PLSB is often less stable than PLSA, and usually yields LVs that present higher inter-correlations. This latter characteristic makes some researchers favor the use of PLSB.

In PLSB, similarly to PLSA, the inner model influences the outer model through path coefficients and correlations, depending on whether the links go into or out from each LV, respectively. In the PLSB algorithm, unlike its PLSA counterpart, the outer model weights are calculated through a least squares regression where the indicators are the predictors and the LV is the criterion.

Combining composites and factors in one model

As previously noted, the process of combining composites and factors in one single model involves the main following steps: (1) choosing composite-based or factor-based outer model analysis algorithms to be used for LV estimation; (2) adding the LVs estimated with the chosen composite-based or factor-based outer model analysis algorithms as new standardized variables; and (3) creating and analyzing a model with the new variables added as single-indicator LVs.

Choosing a composite-based or factor-based outer model analysis algorithm to be used for LV estimation can be done through the “View or change general settings” menu option, available under the “Setting” option on the main menu; which allows users to set the outer model analysis algorithm, among other settings (see Figure 2). Once this choice is made, the researcher should perform the SEM analysis, and add the LV or LVs estimated with the specific composite-based or factor-based outer model analysis algorithm as new standardized indicators (see Figure 3). Finally, the researcher should then proceed to create a model with the new variables added as single-indicator LVs (see Figure 4), and perform the SEM analysis one final time.

Note that the final model does not only contain a combination of composites and factors, but of composites and factors estimated via different algorithms. In our case, the algorithms used were PLSF-CFM3 (for DM), PLSF-CFM1 (for JP), PLSA (for SJ), and PLSB (for JS and OC). The final model contains LVs with single indicators, which allow for the estimation of the structural model parameters. However, the final model does not contain important measurement model parameters such as loadings and reliability coefficients.

The measurement model parameters of interest can be obtained from the project files storing the intermediate analyses using the selected algorithms, which should be saved separately for this purpose. Since we used four algorithms in our analysis – namely PLSF-CFM3, PLSF-CFM1, PLSA, and PLSB – there should be four project files storing the results of the intermediate analyses. In addition to these four project files, one extra file should be created to store the results of the analysis of the final model containing the LVs with single indicators.

Figure 2: Choosing an outer model analysis algorithm

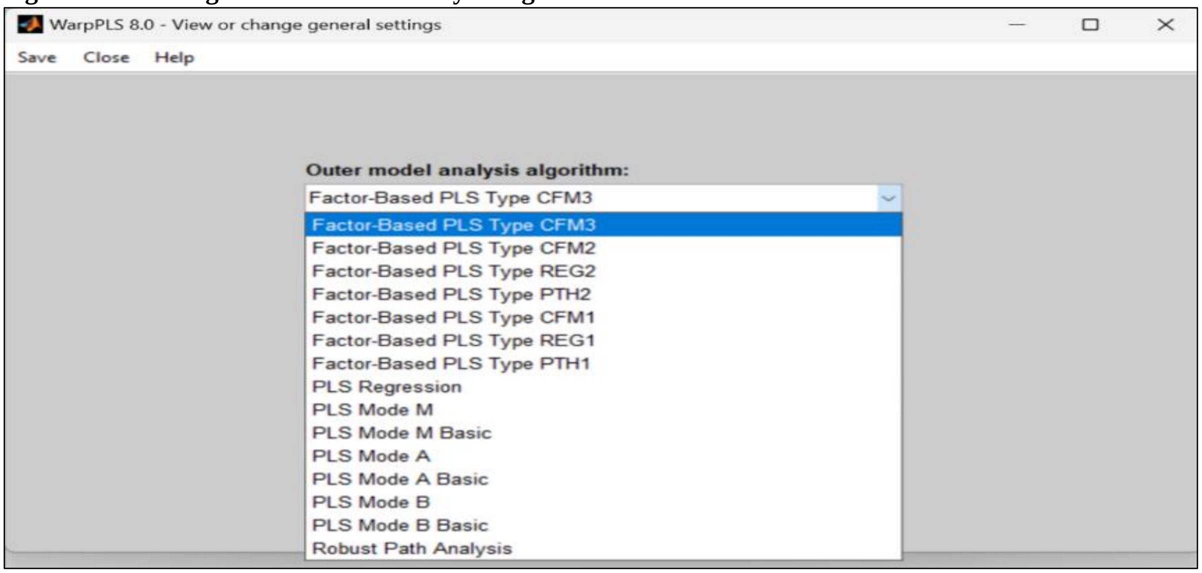


Figure 3: Adding LVs as new standardized indicators

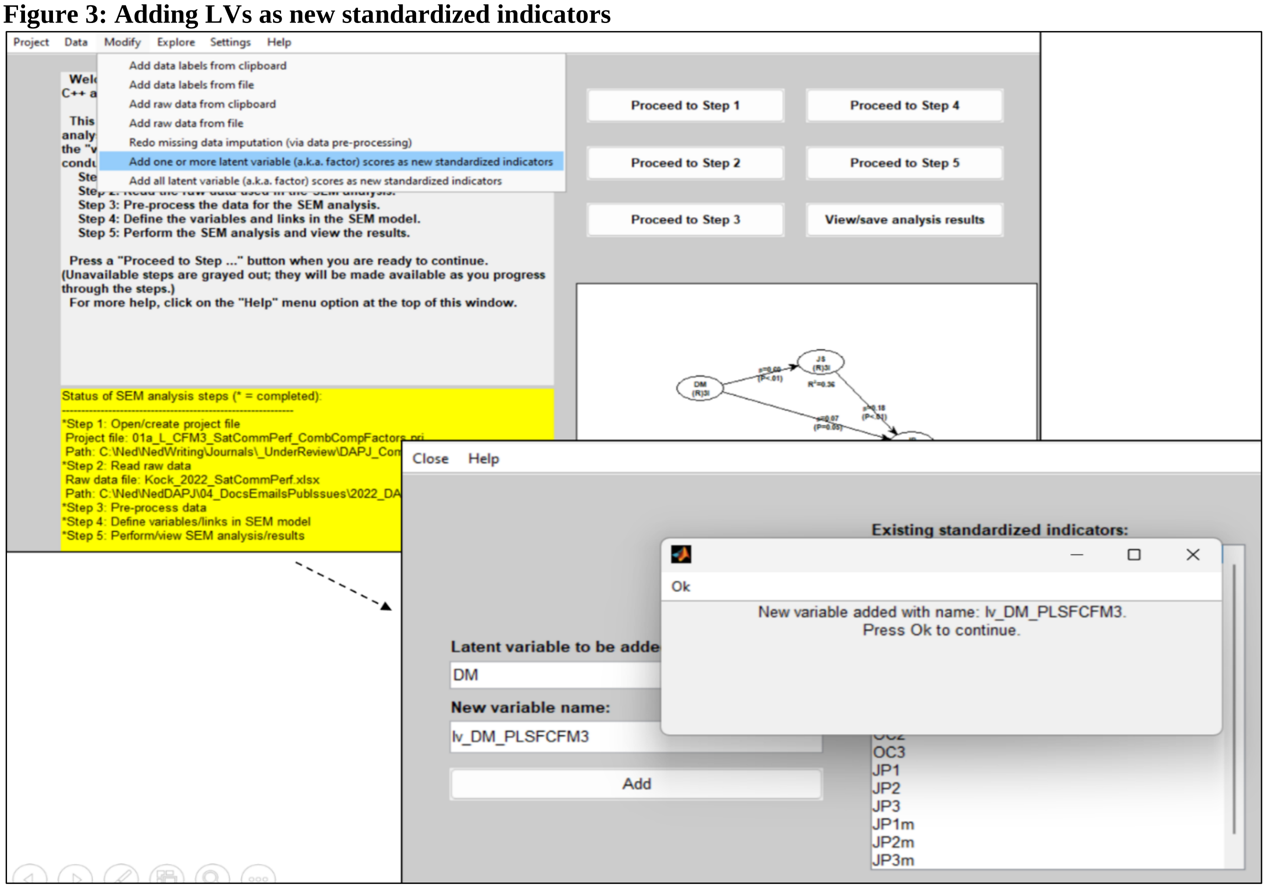


Figure 4: Creating a model with new LVs as indicators

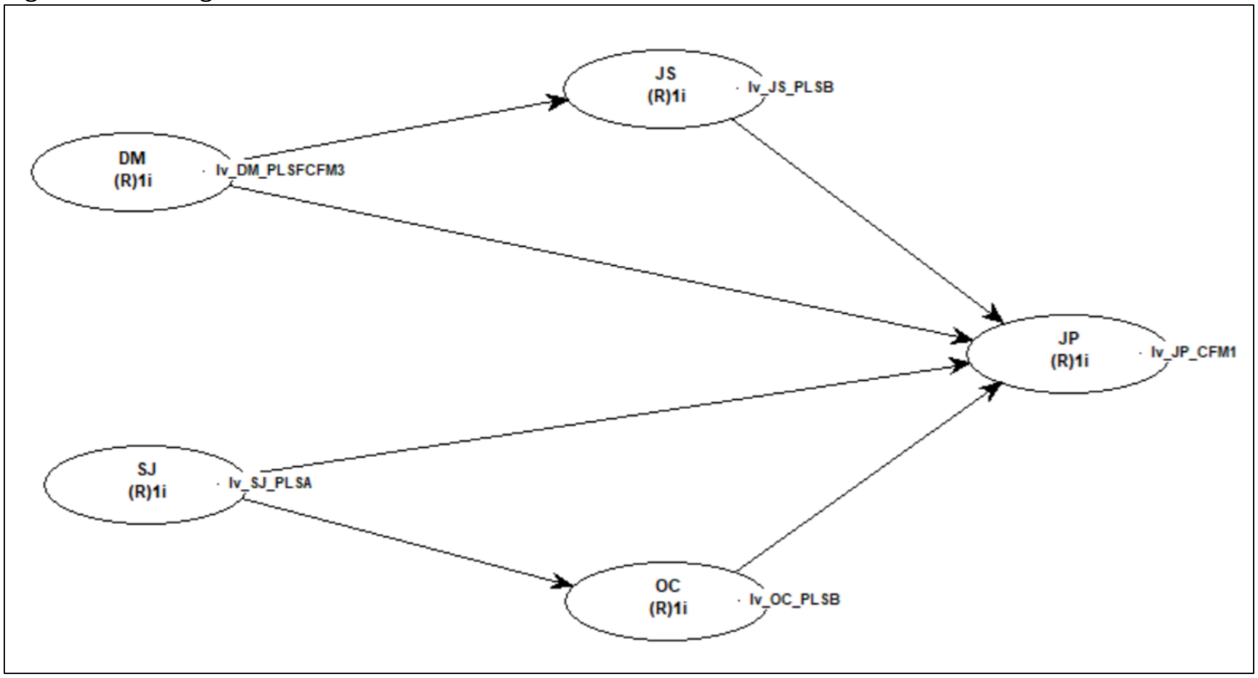
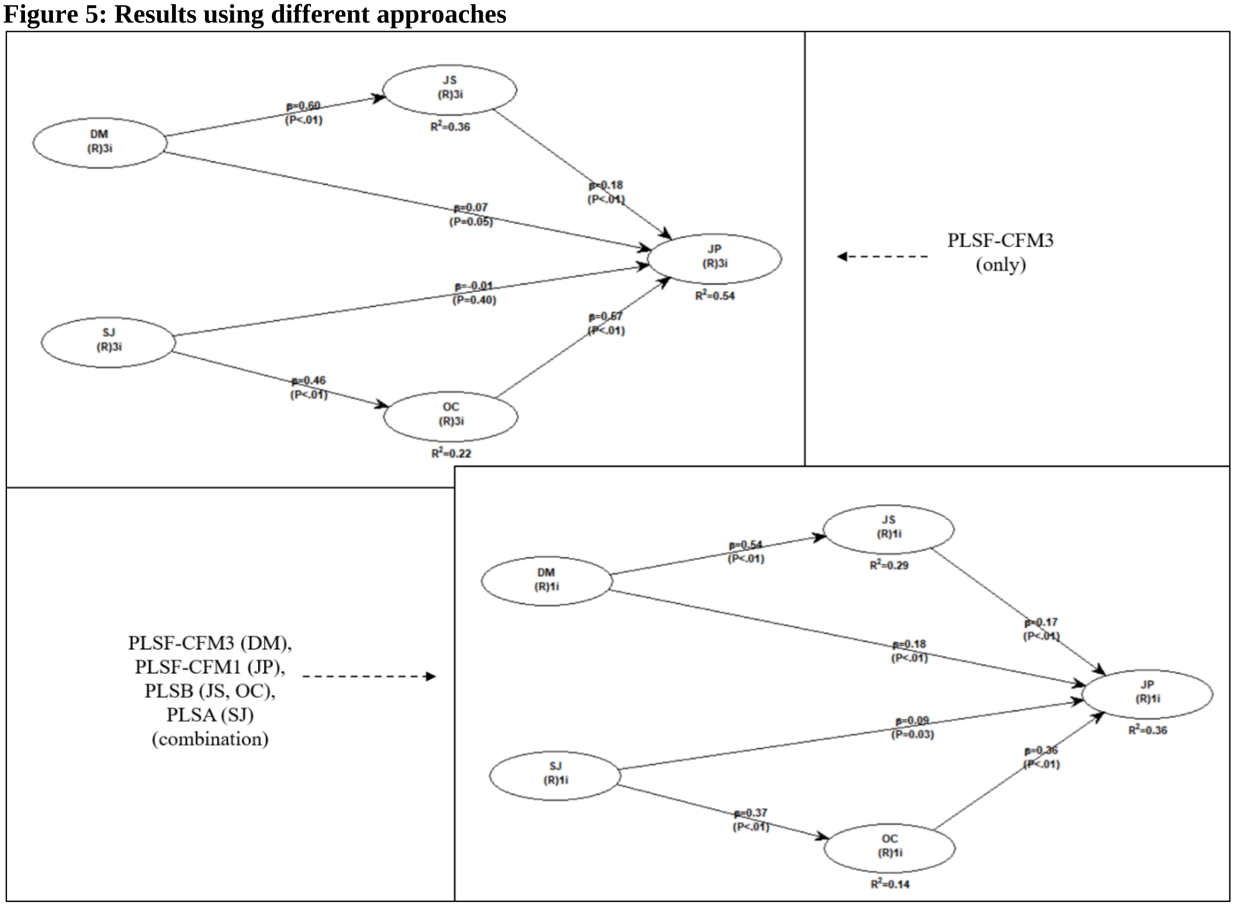


Figure 5: Results using different approaches



Analysis of model with new LVs as indicators

Figure 5 shows the results of the analysis of the final model containing the LVs with single indicators. Also shown are the results of one of the intermediate analyses employing only the PLSF-CFM3 algorithm. The former is at the bottom right of the figure, and the latter at the top left. As we can see, the structural model analysis results, which include path coefficients and R-squared values, varied significantly across models.

In the final model employing LVs estimated as composites and factors, the R-squared coefficients for the endogenous LVs are considerably lower than for the intermediate model employing LVs estimated only as factors. Generally speaking, this is not a desirable outcome, suggesting that the model using LVs estimated only as factors presents a better fit with the data from a structural perspective.

On the other hand, the DM > JP and SJ > JP links, representing hypothesized direct causal relationships, turned out to be significant only in the final model employing LVs estimated as composites and factors. While this may be seen as a desirable outcome from a hypothesis testing perspective, the lower R-squared coefficients generally suggest poorer model fit, and also that the total effects DM >> JP and SJ >> JP are stronger in the model employing LVs estimated only as factors.

Conclusion

In this paper, we presented a multi-algorithm strategy for combining LVs calculated as composites and factors into a single PLS-SEM model, which can be useful to researchers who want to combine LVs estimated via different composite-based or factor-based algorithms in a single model. Researchers may want to do so owing to multiple reasons, such as theoretical considerations and review panel requests.

The multi-algorithm technique is arguably simple to understand and easy to implement. To recap, it consists of three key steps: selecting composite-based or factor-based outer model analysis algorithms to be used for LV estimation; adding the LVs estimated with specific composite-based or factor-based algorithms as new standardized variables; and developing and estimating a model with the new variables added as single-indicator LVs.

Acknowledgments

The author is the developer of the software WarpPLS. He is grateful to WarpPLS users for questions, comments, discussions, and continued use. This article contains revised text, originally written by the author, from a recent edition of the WarpPLS User Manual.

References

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

Amora, J. T. (2023). On the validity assessment of formative measurement models in PLS-SEM. *Data Analysis Perspectives Journal*, 4(2), 1-7.

Canatay, A., Emegwa, T., Lybolt, L. M. & Loch, K. D. (2022). Reliability assessment in SEM models with composites and factors: A modern perspective. *Data Analysis Perspectives Journal*, 3(1), 1-6.

Cox, J. (2024). Combining sub-samples for improved statistical power in PLS-SEM: A constrained latent growth approach. *Data Analysis Perspectives Journal*, 5(1), 1-5.

Hubona, G., & Belkhamza, Z. (2021). Testing a moderated mediation in PLS-SEM: A full latent growth approach. *Data Analysis Perspectives Journal*, 2(4), 1-5.

Kock, N. (2015a). One-tailed or two-tailed P values in PLS-SEM? *International Journal of e-Collaboration*, 11(2), 1-7.

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. (2016). Non-normality propagation among latent variables and indicators in PLS-SEM simulations. *Journal of Modern Applied Statistical Methods*, 15(1), 299-315.

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

Kock, N. (2019b). Factor-based structural equation modeling with WarpPLS. *Australasian Marketing Journal*, 27(1), 57-63.

Kock, N. (2020a). Full latent growth and its use in PLS-SEM: Testing moderating relationships. *Data Analysis Perspectives Journal*, 1(1), 1-5.

Kock, N. (2020b). Multilevel analyses in PLS-SEM: An anchor-factorial with variation diffusion approach. *Data Analysis Perspectives Journal*, 1(2), 1-6.

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

Kock, N. (2021b). 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. (2021c). Moderated mediation and J-curve emergence in path models: An information systems research perspective. *Journal of Systems and Information Technology*, 23(3), 303-321.

Kock, N. (2022a). *WarpPLS User Manual: Version 8.0*. Laredo, TX: ScriptWarp Systems.

Kock, N. (2022b). Testing and controlling for endogeneity in PLS-SEM with stochastic instrumental variables. *Data Analysis Perspectives Journal*, 3(3), 1-6.

Kock, N. (2022c). Using causality assessment indices in PLS-SEM. *Data Analysis Perspectives Journal*, 3(5), 1-6.

Kock, N. (2023a). Assessing multiple reciprocal relationships in PLS-SEM. *Data Analysis Perspectives Journal*, 4(3), 1-8.

Kock, N. (2023b). Using logistic regression in PLS-SEM: Dichotomous endogenous variables. *Data Analysis Perspectives Journal*, 4(4), 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., & 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.

Ma, K. Q., & Zhang, W. (2023). Assessing univariate and multivariate normality in PLS-SEM. *Data Analysis Perspectives Journal*, 4(1), 1-7.

Moqbel, M., Guduru, R., & Harun, A. (2020). Testing mediation via indirect effects in PLS-SEM: A social networking site illustration. *Data Analysis Perspectives Journal*, 1(3), 1-6.

Morrow, D. L., & Conger, S. (2021). Assessing reciprocal relationships in PLS-SEM: An illustration based on a job crafting study. *Data Analysis Perspectives Journal*, 2(1), 1-5.

Rasoolimanesh, S. M. (2022). Discriminant validity assessment in PLS-SEM: A comprehensive composite-based approach. *Data Analysis Perspectives Journal*, 3(2), 1-8.

Samak, A., Islam, M. R., & Hanke, D. (2024). A comparison of data analyses with WarpPLS and Stata: A study of trust and its role regarding internet use and subjective well-being. *Data Analysis Perspectives Journal*, 5(3), 1-6.

Tarkom, A., & Gopal, P. (2024). A comparison of multiple regression analyses in Stata and WarpPLS. *Data Analysis Perspectives Journal*, 5(2), 1-8.