Using indicator correlation fit indices in PLS-SEM: Selecting the algorithm with the best fit

Ned Kock
Texas A&M International University, USA

Abstract

Upon completion of a PLS-SEM analysis, one can obtain the model-implied indicator correlation matrix and compare it with the actual indicator correlation matrix. The latter is obtained directly from the data being analyzed. Indicator correlation fit indices are quantifications of the differences among these two matrices. Our focus in this paper is on the use of indicator correlation fit indices in PLS-SEM for selecting the analysis algorithm with the best fit.

Keywords: Indicator Correlations; Fit Indices; Structural Equation Modeling; Partial Least Squares; WarpPLS.

Introduction

Structural equation modeling (SEM) is a method that allows researchers to build and analyze causal models with latent variables. Upon completion of an SEM analysis, one can obtain the model-implied indicator correlation matrix and compare it with the actual indicator correlation matrix. The latter is obtained directly from the data being analyzed. Indicator correlation fit indices are quantifications of the differences among these two matrices.

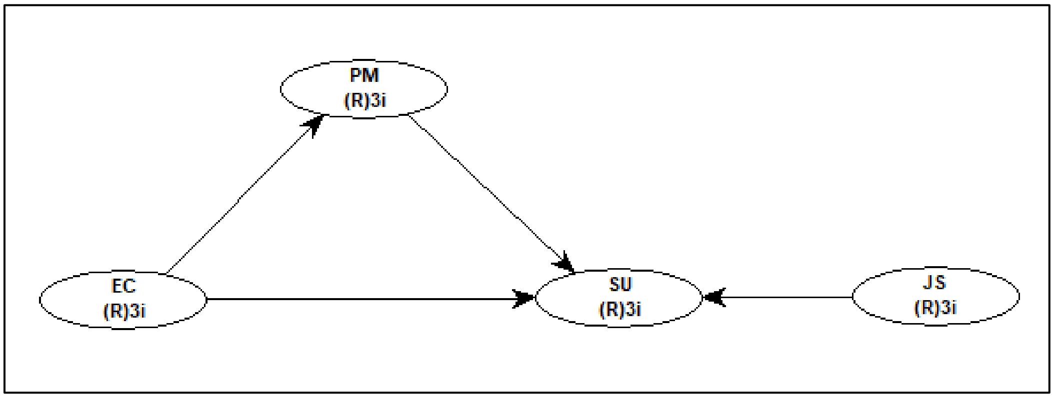
The use of indicator correlation fit indices, sometimes referred to as “global” fit indices, has a long history in SEM. They can be used for composite-based analyses building on classic partial least squares (PLS) algorithms, as well as more modern PLS-SEM algorithms that estimate latent variables as factors, leading to more precise estimates (Kock, 2019a; 2019b). WarpPLS is a leading PLS-SEM software tool that implements both types of algorithms (Kock, 2020a). We use WarpPLS version 7.0 in this paper.

A key assumption underlying the use of indicator correlation matrices is that they provide a rather detailed “signature” of a SEM model, when compared with other sets of parameters. In other words, the assumption is that an indicator correlation matrix is a set of parameters that carries a lot of information about any given model. Our focus in this paper is on the use of indicator correlation fit indices in PLS-SEM for selecting the analysis algorithm with the best fit.

Illustrative model and data

Figure 1 shows the model used as a basis for our discussion. This model, also used in a previous study (Kock, 2020b), includes four latent variables: the degree to which members of project teams use an e-collaboration technology (EC), the degree to which members of project teams use state-of-the-art project management techniques (PM); the business success of the projects conducted by the teams (SU); and the degree to which members of project teams are satisfied with their regular jobs (JS).

Figure 1: Illustrative model used



Notes: EC = e-collaboration technology use; PM = project management techniques use; SU = project success; JS = job satisfaction; notation under latent variable acronym describes measurement approach and number of indicators, e.g., (R)3i = reflective measurement with 3 indicators.

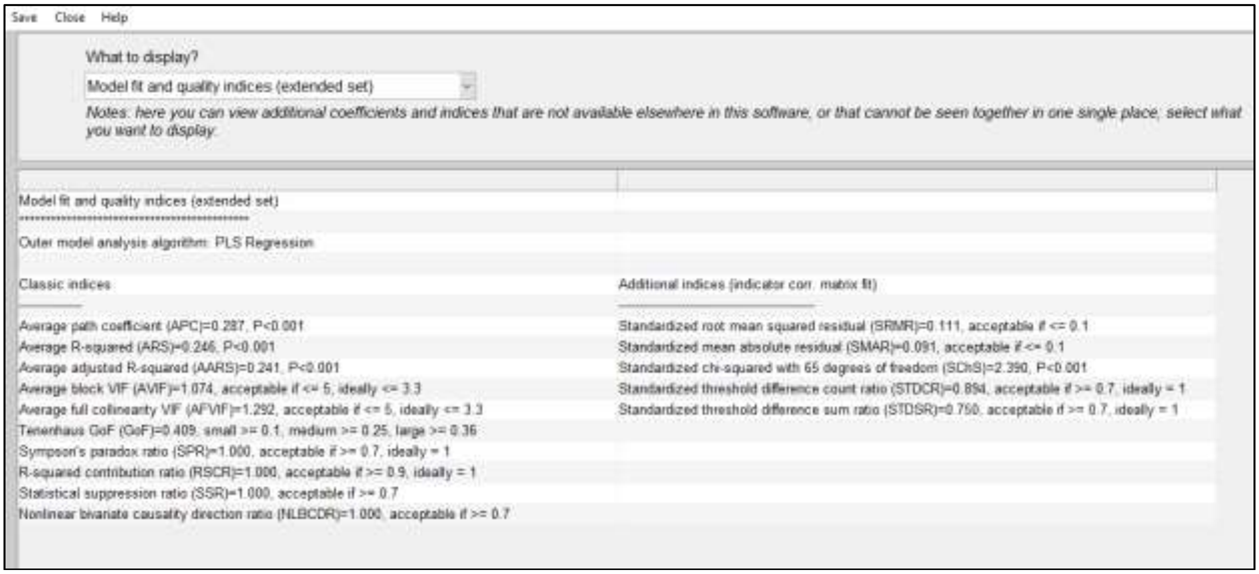
For the purposes of the present discussion, we created data employing the Monte Carlo simulation method (Kock, 2016). We used the model above as a basis. We also used prior research on project teams in various organizations; the teams were involved in the development of new products, such as a new toothpaste, a new car part, or a new pill to treat a disease. The illustrative dataset contained 300 cases, where each case refers to one project team.

Indicator correlation fit indices used in PLS-SEM

After an analysis is conducted, the WarpPLS menu option “Explore additional coefficients and indices” allows a user to obtain an extended set of model fit and quality indices (see Figure 2). This extended set of model fit and quality indices displays the classic indices normally included in reports of SEM analyses employing this software, as well as new indices that allow investigators to assess the fit between the model-implied and empirical indicator correlation matrices. These new indices are 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).

SRMR and **SMAR**. The SRMR index is calculated as the square root of the mean of the sum of the squared differences between the contents of non-redundant cells of the model-implied and empirical indicator correlation matrices. The SMAR index is calculated as the mean of the sum of the absolute differences between those matrices. The model-implied indicator correlation matrix is obtained based on the model parameters (e.g., weights and loadings) estimated by the software. The empirical indicator correlation matrix is simply the matrix containing the correlations among the indicators used in the model. The non-redundant cells of these matrices are the upper or lower triangular cells, excluding the diagonal cells. Generally, SRMR and SMAR values lower than 0.1 indicate acceptable fit.

Figure 2: Extended set of model fit and quality indices



SChS. The SChS index is calculated as the chi-squared coefficient obtained from a test of independence comparing the contents of non-redundant cells of the model-implied and empirical indicator correlation matrices. Here the contents of non-redundant cells of the model-implied indicator correlation matrix are treated as the observed values in a chi-squared test of independence, whereas the corresponding values in the empirical indicator correlation matrix are treated as the expected values. The number of degrees of freedom is calculated as the number of non-redundant cells minus 1, in line with what is usually done in traditional chi-squared tests of independence. For simplicity and consistency of application with respect to other model fit and quality indices, the P value associated with each SChS is calculated as the complement of the P value generated by the chi-squared test of independence (i.e., 1 minus that P value). Normally acceptable fit is indicated by a P value associated with a SChS that is equal to or lower than 0.05; that is, significant at the 0.05 level. This refers to the modified P value; the smaller it is, the better the fit.

STDCR and **STDSR**. The STDCR and STDSR indices are measures of the extent to which a model is free from instances in which the contents of non-redundant cells of the model-implied indicator correlation matrix differ significantly from the corresponding empirical indicator correlation matrix values. Here a heuristic threshold is used to establish whether two values differ significantly; this threshold is 0.2, twice the model-wide acceptable fit threshold for the SRMR and SMAR indices. The STDCR is calculated by dividing the number of non-redundant cells where significant differences do not exist by the total number of non-redundant cells. The STDSR index is calculated as the complement of the ratio obtained by dividing the sum of the absolute values of the differences between non-redundant cells where a significant difference exists by the total sum of the absolute values of the differences between non-redundant cells. These new STDCR and STDSR indices are calculated so that they can be used in ways analogous to other classic fit indices generated by this software. Generally, values of the STDCR and STDSR equal to or greater than 0.7 indicate acceptable fit.

Selecting the algorithm with the best fit

Let us say that we want to select the algorithm with the best fit among two algorithms: "Factor-Based PLS Type CFM3" and "PLS Regression". The former, factor-based algorithm, yielded the following results for the indices, when used to analyze our illustrative model and data: SRMR=0.073, SMAR=0.061, SChS=2.390 (P<0.001), STDCR=0.985, and STDSR=0.946. The latter, composite-based algorithm, yielded the following results: SRMR=0.111, SMAR=0.091, SChS=0.423 (P<0.001), STDCR=0.894, and STDSR=0.750.

These results suggest that, of the two algorithms considered, the one with the best fit is the "Factor-Based PLS Type CFM3" algorithm. This is because this algorithm has the lowest SRMR and SMAR values; and the highest SChS, STDCR, and STDSR values. Moreover, the "Factor-Based PLS Type CFM3" algorithm is the only one between the two with an SRMR value below the recommended threshold of 0.1.

Researchers may want to focus on a subset of the indicator correlation fit indices above to select among various algorithms, for the sake of simplicity. In this case it is recommended that they use the indices in the sequence above, one-by-one. That is, they should use the SRMR first, the SMAR second, and so on, with subsequent indices used as needed as "tiebreakers". For this type of application, often only the SRMR and SMAR indices will be needed.

Acknowledgments

The author is the developer of the software WarpPLS, which has over 10,000 users in more than 33 different countries at the time of this writing. This article contains revised text, originally written by the author, from the WarpPLS User Manual.

References

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). *WarpPLS User Manual: Version 7.0*. Laredo, TX: ScriptWarp Systems.

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