Data Analysis Perspectives Journal, 3(4), 1-7, March 2022

Minimum sample size estimation in SEM: Contrasting results for models using composites and factors

Blessing Ezeugwa
Texas A&M International University, USA

Md Farid Talukder
Texas A&M International University, USA

Muhammad Ruhul Amin
Texas A&M International University, USA

Saha Iqbal Hossain
Texas A&M International University, USA

Faruk Arslan
New Mexico State University, USA

Abstract

Estimating the minimum required sample size is an essential issue for studies that use structural equation modeling employing partial least squares (PLS-SEM). Several PLS-SEM-based studies ignore this critical step or use simple techniques, which lead to inaccurate sample size estimations. This paper illustrates two effective heuristic methods to estimate the minimum required sample size using WarpPLS, a leading PLS-SEM software tool.

Keywords: Minimum Required Sample Size; PLS-SEM; Inverse Square Root Method; Gamma-Exponential Method; WarpPLS.

Introduction

This study focuses on the minimum sample size estimation in the context of partial least squares-based structural equation modeling (PLS-SEM). Estimating the minimum sample size requirement is essential to preserve the statistical power of a PLS-SEM test (Kock & Hadaya, 2018). Using the correct sample size helps to improve the statistical generalizability of the results (Lee & Baskerville, 2003). Therefore, Kock & Hadaya (2018) define minimum sample size estimation as one of the fundamental issues of PLS-SEM. Despite its importance, few PLS-SEM-based studies pay attention to the adequacy of the sample size required to achieve the desired statistical power of the PLS-SEM test. In this paper, we present two heuristic methods to estimate the minimum sample size requirement in both composite and factor-based PLS-SEM study: i) the inverse square root method and ii) the gamma-exponential method. These two

methods are available as features in WarpPLS - a popular PLS-SEM software (Kock, 2020a, 2020b) that implements a number of advanced data analysis features (Amora, 2021; Canatay et al., 2022; Hubona & Belkhamza, 2021; Kock, 2020c; 2021a; 2021b; 2022; Moqbel et al., 2020; Morrow & Conger, 2021; Rasoolimanesh, 2022).

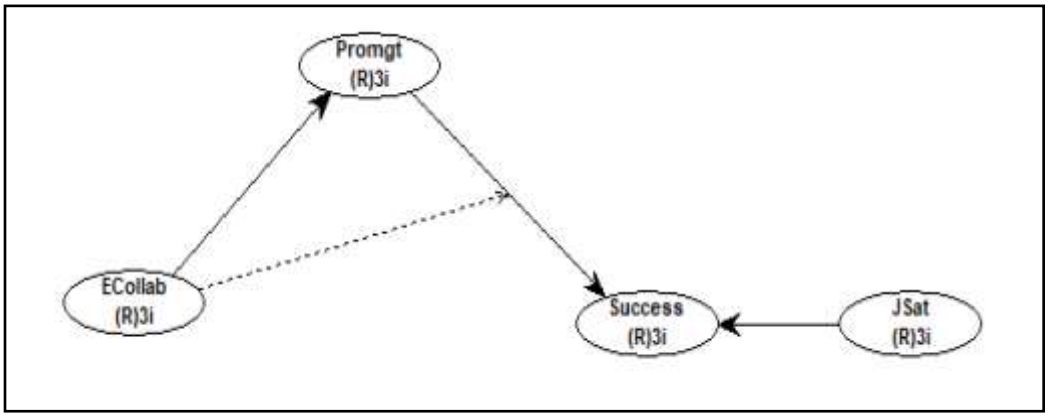
Why is minimum sample size estimation important?

Determining the correct sample size is essential for the reliability of the sampling procedure since an increase in sample size leads to "greater generalizability of the sample points to a sample estimate because of the greater convergence expected from, the larger sample size" (Lee & Baskerville, 2003). Although PLS-SEM is perceived to be an effective method to analyze complex models using smaller sizes, the results may suffer due to inadequate sample size like any other statistical method. Hence, researchers must exert the proper effort to achieve acceptable levels of statistical power in their research settings. While the "10-times rule method" and the "minimum R-squared method" are easy to use methods, they are shown to be not accurate in estimating the minimum required sample size (Kock & Hadaya, 2018). Alternatively, two heuristic methods, i) the inverse square root method and ii) the gamma-exponential method, produce fairly accurate estimations (Kock & Hadaya, 2018). These two methods are presented in the following sections of this paper. Both methods can be executed fairly easily using the WarpPLS software.

Illustrative model and data

The model in Figure 1 is used as a basis in this study. Data for this study was 300 cases obtained from the WarpPLS database (see resources at warppls.com). This model has four latent variables–the degree to which members of project teams use an e-collaboration technology (ECollab), the degree to which they use state-of-the-art project management techniques (Promgt), the business success of the project operated by the team members (Success), and the degree to which the team members are satisfied with their regular jobs (JSat).

Figure 1: Illustrative model used

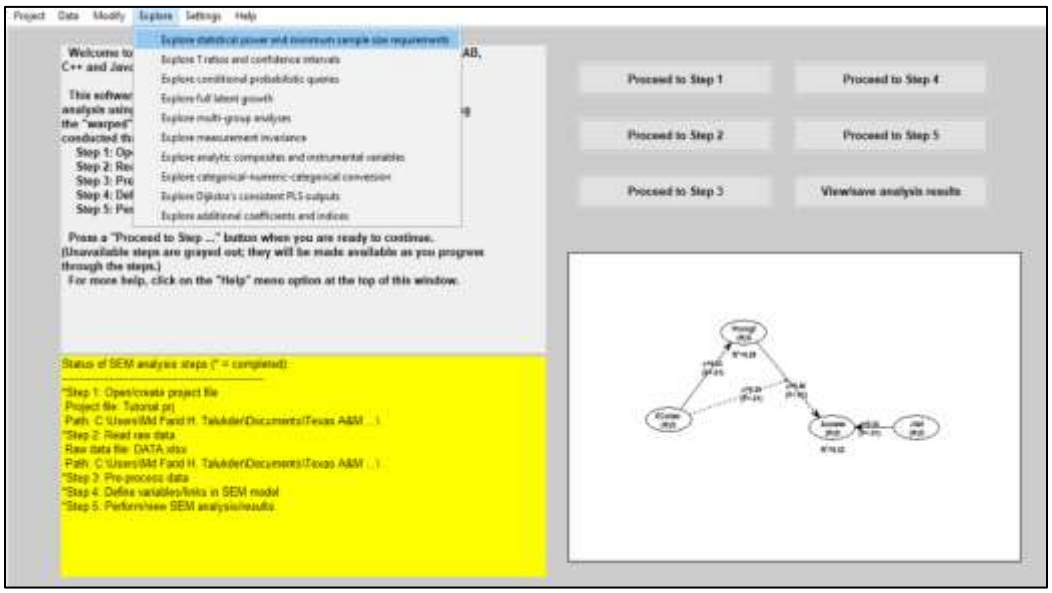


Note: (R)3i = reflective measurement with 3 indicators.

Minimum sample size estimation for composite-based models

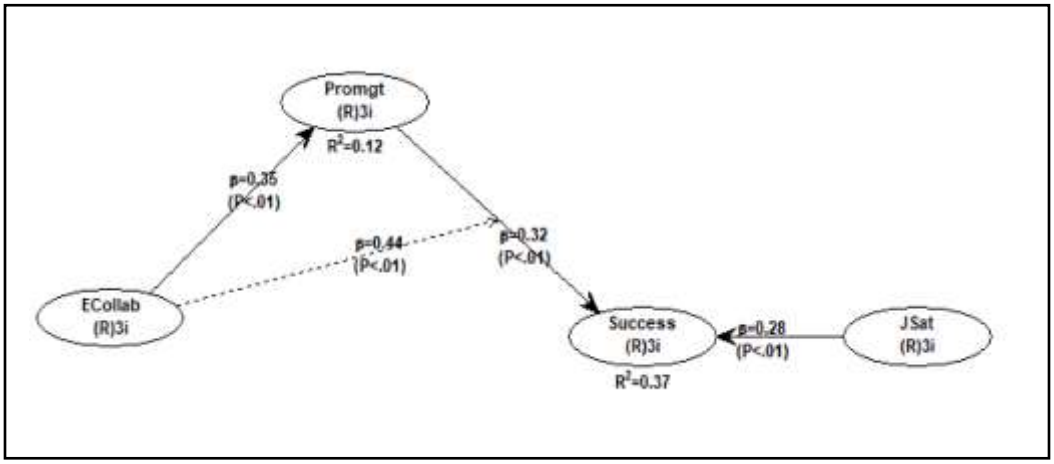
Composite-based SEM algorithms generate a higher minimum required sample size because they tend to underestimate real path coefficients and overestimate path coefficient effects that do not exist (Kock, 2019). Composite-based SEM aggregates indicators but does not fully incorporate measurement error, which is why the classic PLS method yields biased estimates of different parameters, "even as sample sizes grow to infinity" (Kock, 2019).

Figure 2: Minimum sample size estimation



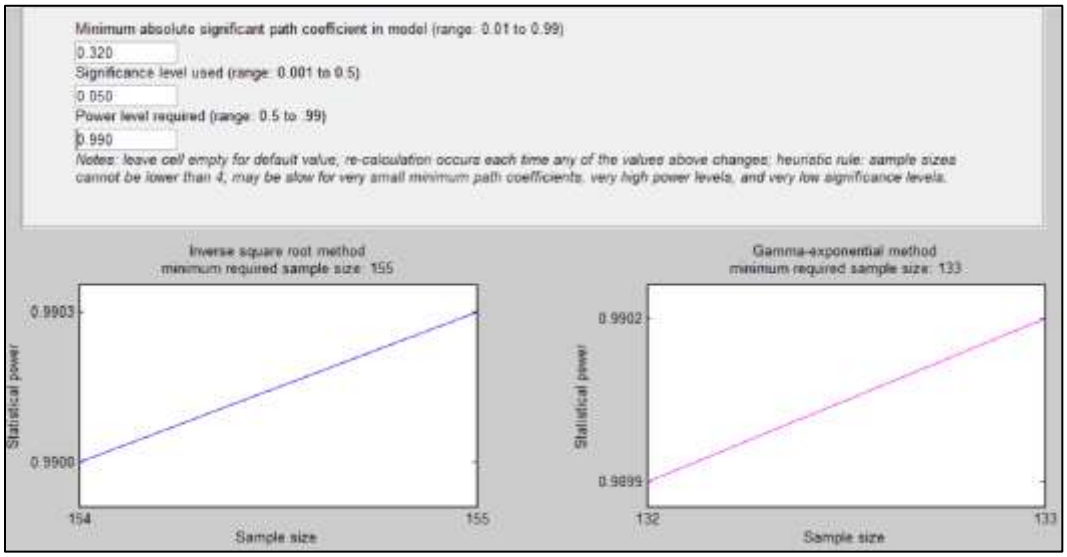
Upon executing the composite-based model (Step 5), go to the main software window and choose the option "Explore" (Figure 2). This option allows users to estimate the minimum sample size using minimum absolute path coefficient, power level, and significance level.

Figure 3: Illustrative model of composite-based method



As shown in Figure 3, the minimum absolute significant path coefficient is 0.32. Based on this, we used significance level (0.05) and power level (0.99) to estimate the minimum required sample size using both the inverse square root method and the gamma-exponential method. Figure 4 shows the inverse square root method generated a larger minimum required sample size (155) than the gamma-exponential method (133). In this case, the preferred minimum sample size is 155. Note that this power level is quite high; usually the value of 0.80 is acceptable.

Figure 4: Sample size estimation when power is 0.99



Minimum sample size estimation for factor-based models

The factor-based method, broadly defined, possesses unique characteristics extensively used in structural equation modeling (SEM) to estimate factors and accounts fully for the measurement error (Kock, 2019). Therefore, this method generates unbiased sample size parameters, a significant advantage over the composite-based method (Kock, 2019). This method has equal statistical consistency with covariance-based SEM but greater statistical efficiency (Kock, 2019).

Figure 5 shows settings to estimate models using factor-based SEM. The WarpPLS "view or change general settings" enables the outer model analysis algorithm to be changed from the composite-based method "PLS Regression" to "Factor-Based PLS Type REG2" for factor-based SEM estimation. This is one of the several factor-based algorithms available from the software.

Figure 6 shows the illustrative model that is used as a basis for our factor-based analysis. Based on the minimum path coefficient (0.46), significance level (0.05), and power level (0.99), we estimated the minimum required sample size based on the inverse square root and gamma-exponential methods. The inverse square root method yields a larger minimum required sample size (75) than the gamma-exponential method (53); see Figure 7. In this case, 75 is the preferred minimum sample size for data collection and analysis. Our sample size (300) is larger than the estimated required minimum sample size for both composite and factor-based analysis.

Note that Figure 6 shows that ECollab has a moderating effect on the link of Promgt > Success. This indicates that the interaction variable ECollab*Promgt is a predictor of Success, in addition to the direct predictors ECollab and Promgt.

Figure 5: Settings for factor-based SEM

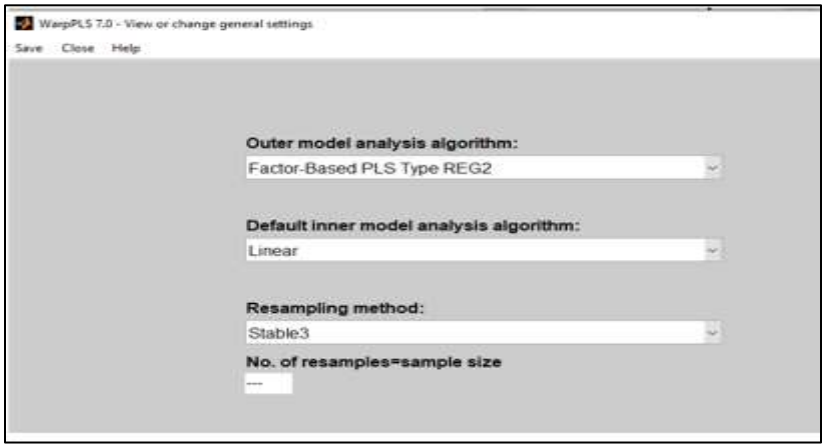


Figure 6: Illustrative model of factor-based method

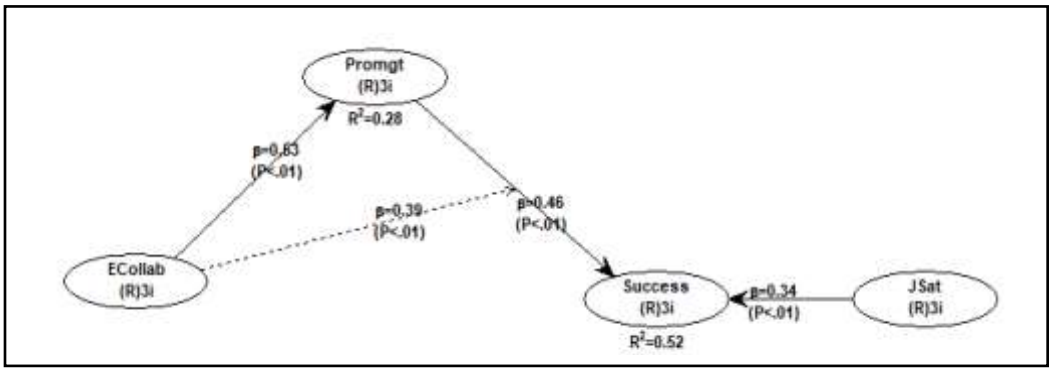
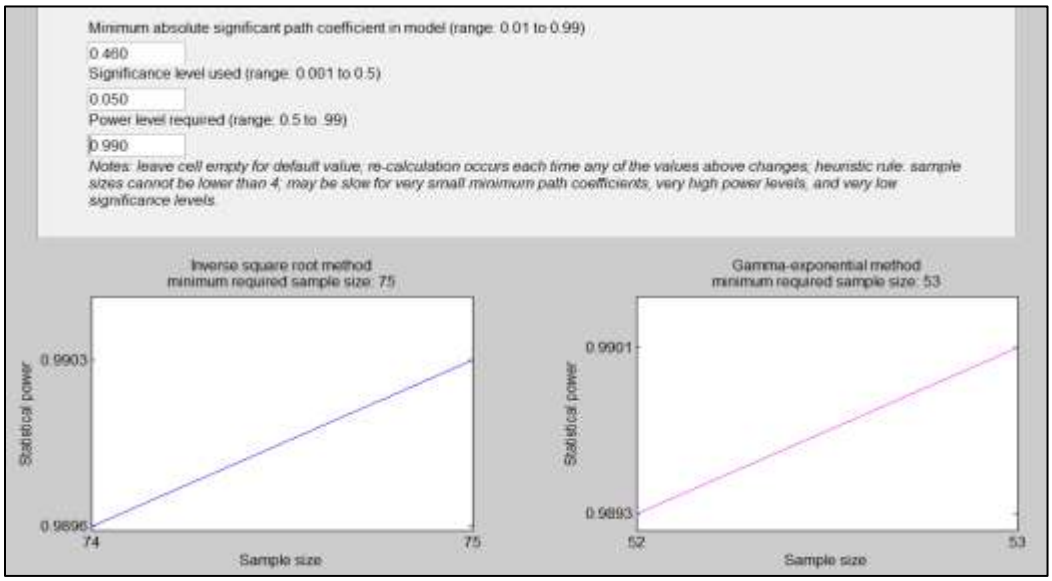


Figure 7: Sample size estimation when power is 0.990



The results shows that the coefficient (0.39) for the moderating effect is lower than the coefficients (0.53 and 0.46) for other direct effects. We did not use 0.39 as the minimum absolute path coefficient to estimate the minimum sample size in this study because the argument of using a moderating effect as the minimum absolute path coefficient to estimate the minimum sample size is still being debated, since moderating effects frequently give rise to nonlinear effects that are significantly stronger than the underlying moderating effects themselves (Kock, 2021c). Further, the model's control variable is not used for minimum sample estimation since it is not hypothesized for our study.

Conclusion

In this study, we demonstrated a minimum sample size estimation analysis in the context of PLS-SEM. Our presentation shows that the minimum required sample size is estimated to be greater with a composite-based SEM algorithm when compared with a factor-based SEM algorithm. The inverse square root method is preferred in both composite and factor-based methods since it produces a more conservative estimate (i.e., a larger minimum required sample size) than the gamma-exponential method.

Acknowledgments

We are grateful to the inventor of WarpPLS for providing a platform to contribute to topics related to its users and other researchers and practitioners in diverse fields interested in structural equation modeling and the use of WarpPLS.

References

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

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.

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. (2014). Advanced mediating effects tests, multi-group analyses, and measurement model assessments in PLS-based SEM. *International Journal of e-Collaboration*, 10(1), 1-13.

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

Lee, A. S., & Baskerville, R. L. (2003). Generalizing generalizability in information systems research. *Information systems research*, 14(3), 221-243.

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.