Assessing univariate and multivariate normality in PLS-SEM

Kathy Qing Ma
Texas A&M International University, USA

Weiyong Zhang
Old Dominion University, USA

Abstract

Partial least squares structural equation modeling (PLS-SEM) has gained popularity among researchers in part due to its relaxed requirement for multivariate normality. One important step in performing structural equation modeling (SEM) is to test the normality assumption. In this paper, we illustrate how to assess univariate and multivariate normality in PLS-SEM using WarpPLS.

Keywords: Univariate Normality; Multivariate Normality; Structural Equation Modeling; Partial Least Squares; WarpPLS.

Introduction

Partial least squares structural equation modeling (PLS-SEM) is a modeling technique that enables researchers to analyze causal-predictive relationships between latent variables. It has gained increasing popularity due to its ability to estimate complex path models with relaxed restrictions on sample sizes and the assumption of normality (Kock, 2016; Kock & Hadaya, 2018).

Classic PLS-SEM algorithms compute composites as exact linear combinations of indicators, whereas factor-based PLS-SEM algorithms generate estimates of both composites and factors to fully account for measurement error. WarpPLS is a leading software that supports both classic and factor-based PLS-SEM algorithms, providing users with accurate estimates while taking into account measurement error (Amora, 2021; Canatay et al., 2022; Ezeugwa, et al., 2022; Hubona & Belkhamza, 2021; Kock, 2019a; 2019b; 2019c; 2020a; 2020b; 2020c; 2021a; 2021b; 2022a; 2022b; 2022c; Moqbel et al., 2020; Morrow & Conger, 2021; Rasoolimanesh, 2022). In this paper, we use WarpPLS version 8.0 to illustrate how to assess univariate and multivariant normality in PLS-SEM.

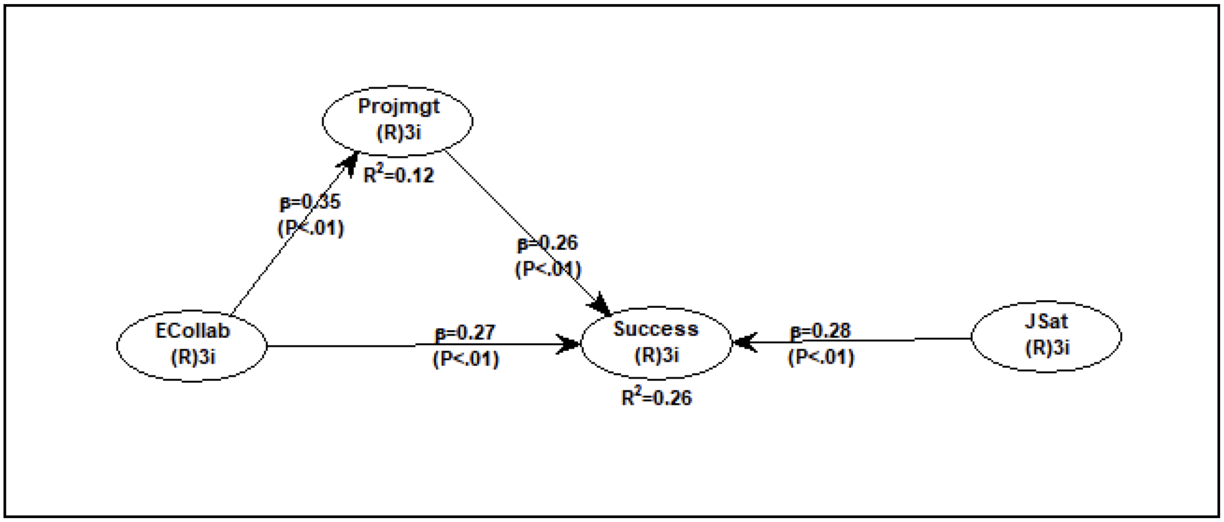
The normality assumption and PLS-SEM

Although the assumption of normality is required by a number of statistical methods, such as linear regression, analysis of variance (ANOVA), and multivariate analysis of variance (MANOVA), PLS-SEM has been shown to perform uniquely well with non-normal data (Kock, 2016; 2022a; Kock & Hadaya, 2018). When normality is in doubt, steps can be taken to check the data for univariate and multivariate normality (Yazici & Yolacan, 2007). WarpPLS allows users to assess both univariate and multivariate normality via a variety of measures and tests, including skewness, excess kurtosis coefficients, normality tests, as well as histograms.

Illustrative model and data

Figure 1 presents the model used for the upcoming discussion. It consists of 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. Figure notations are ECollab for EC, Projmgt for PM, Success for SU, and JSat for JS.

Data was generated through Monte Carlo simulation (see Kock, 2016 for details). The data set contains 300 cases, each case referring to one project team involved in the development of new products in various organizations. The unit of analysis is a project team, not an individual. Examples of new products include a new toothpaste, a new car part, a new pill to treat a disease, to name a few.

It is hypothesized that e-collaboration technology use (EC) facilitates the use of project management techniques (PM), which in turn increases project success (SU). E-collaboration technology use (EC) also increases project success (SU) directly, without PM mediation. Job satisfaction (JS) of project team members has been included as a control variable.

Assessing univariate normality in PLS-SEM

Under the “Data” tab, users can access the menu option “View or save correlation and descriptive statistics for indicators” to check for univariate normality. This option allows users to view the descriptive statistics for indicators of all latent variables included in the model, including skewness, excess kurtosis coefficient, results of two normality tests (i.e., Jarque-Bera and robust Jarque-Bera tests of normality), and histograms.

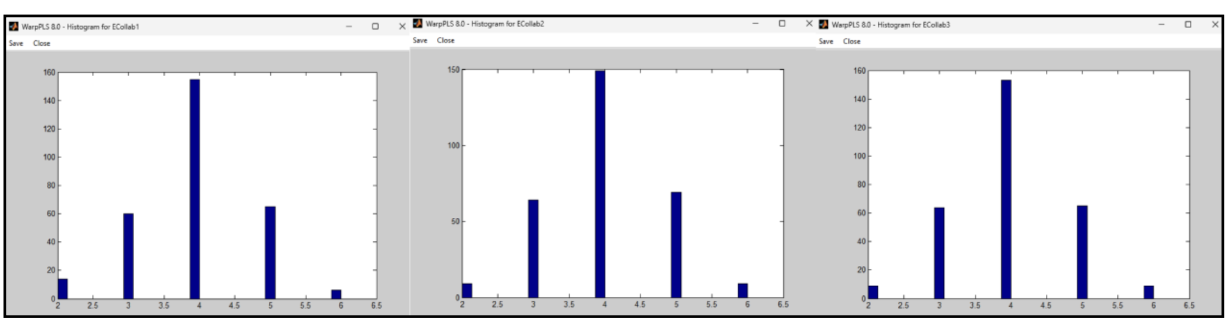
Figure 2 shows the descriptive statistics for the three indicators of EC. As illustrated, the skewness and excess kurtosis values for all three indicators of EC are close to zero and range from -0.216 to -0.006 and 0.008 to 0.139, respectively; which suggests a normal univariate distribution for each of the EC indicators (Kock, 2016). All three indicators pass the original Jarque-Bera (JB) test of normality yet fail the robust Jarque-Bera (RJB) test of normality, which has been shown to have identical or slightly higher power than the original JB test for detecting alternatives to normality (Gel & Gastwirth, 2008).

Figure 2: Descriptive statistics for EC indicators

	ECollab1	ECollab2	ECollab3	Projmgt1	Projmgt2	Projmgt3	Success1
ECollab3	0.370	0.389	1.000	0.269	0.201	0.242	0.196
Projmgt1	0.174	0.212	0.269	1.000	0.431	0.374	0.181
Projmgt2	0.220	0.189	0.201	0.431	1.000	0.408	0.183
Projmgt3	0.200	0.185	0.242	0.374	0.408	1.000	0.199
Success1	0.095	0.074	0.196	0.181	0.183	0.199	1.000
Success2	0.102	0.005	0.113	0.143	0.194	0.180	0.392
Success3	0.029	-0.029	0.150	0.131	0.238	0.224	0.375
(No. diff. vals.)	5.000	5.000	5.000	5.000	6.000	5.000	6.000
(No. diff. vals./N)	0.017	0.017	0.017	0.017	0.020	0.017	0.020
(Mean)	3.963	4.017	4.003	4.017	4.023	3.997	4.010
(SD)	0.827	0.828	0.820	0.820	0.832	0.808	0.812
(Min)	2.000	2.000	2.000	2.000	2.000	2.000	2.000
(Max)	6.000	6.000	6.000	6.000	7.000	6.000	7.000
(Median)	4.000	4.000	4.000	4.000	4.000	4.000	4.000
(Mode)	4.000	4.000	4.000	4.000	4.000	4.000	4.000
(Skewness)	-0.216	-0.031	-0.006	-0.104	0.096	0.082	0.019
(Exc. kurtosis)	0.139	0.008	0.097	-0.430	0.119	-0.324	0.284
(Unimodal-RS)	Yes	Yes	Yes	Yes	Yes	Yes	Yes
(Unimodal-KMV)	Yes	Yes	Yes	Yes	Yes	Yes	Yes
(Normal-JB)	Yes	Yes	Yes	Yes	Yes	Yes	Yes
(Normal-RJB)	No	No	No	Yes	No	No	No
(Histogram)	View	View	View	View	View	View	View

The corresponding histograms for the three EC indicators are presented in Figure 3. As one can see, the histograms for all three indicators are approximately bell-shaped and symmetric about the mean; which again suggests univariate normality for each of the EC indicators. Taken together, one could likely conclude that the data follows a normal distribution for all three indicators of EC.

Figure 3: Histograms for EC indicators



Assessing multivariate normality in PLS-SEM

By selecting "View/save analysis results", users can access the menu option "View latent variable coefficients". Figure 4 shows the descriptive statistics for all latent variables in the model, allowing users to check for multivariate normality. As exhibited, the skewness and excess kurtosis statistics for PM and JS are close to zero (0.043 and -0.333 for PM, and 0.062 and -0.500 for JS), which implies that the data is normally distributed. These two latent variables also pass the JB and RJB tests of normality.

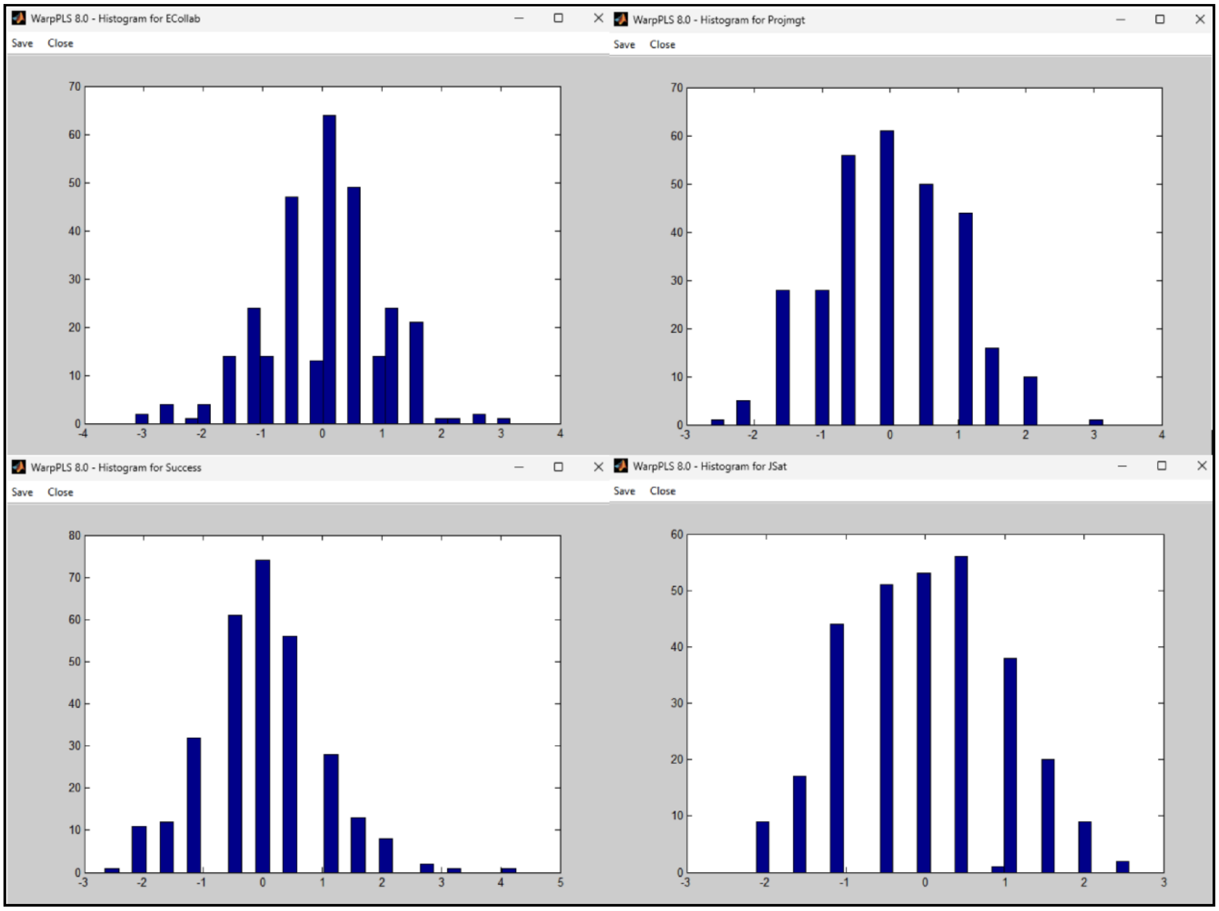
EC and SU score relatively higher on skewness and excess kurtoses, with values of -0.200 and 0.454 for EC, and 0.342 and 1.073 for SU; which are still considered acceptable as they fall within the -1.5 and +1.5 range (Tabachnick & Fidell, 2013). Out of the two latent variables, EC fails the RJB test of normality while SU fails both the JB and RJB tests of normality.

Figure 4: Descriptive statistics for all latent variables

	ECollab	Projmgt	Success	JSat
R-squared		0.124	0.258	
Adj. R-squared		0.121	0.251	
Composite reliab.	0.813	0.820	0.811	0.821
Cronbach's alpha	0.656	0.671	0.650	0.673
Avg. var. extrac.	0.592	0.603	0.589	0.605
Full collin. VIF	1.146	1.243	1.213	1.098
Q-squared		0.127	0.258	
(No. diff. vals.)	55.000	45.000	49.000	48.000
(No. diff. vals./N)	0.183	0.150	0.163	0.160
Min	-3.141	-2.638	-2.653	-2.126
Max	3.158	3.128	4.256	2.560
Median	0.009	-0.019	0.005	-0.024
Mode	0.009	-0.019	0.005	-0.024
Skewness	-0.200	0.043	0.342	0.062
Exc. kurtosis	0.454	-0.333	1.073	-0.500
Unimodal-RS	Yes	Yes	Yes	Yes
Unimodal-KMV	Yes	Yes	Yes	Yes
Normal-JB	Yes	Yes	No	Yes
Normal-RJB	No	Yes	No	Yes
Histogram	View	View	View	View

Figure 5 presents the histograms for the four latent variables in the model. As displayed, the histograms for EC, PM, and JS are roughly bell-shaped and symmetric about the mean, whereas the histogram for SU is somewhat right-skewed. Based on the statistics and histograms, one could possibly conclude that the data approximates a normal distribution for EC, PM, and JS yet is slightly positive-skewed for SU.

Figure 5: Histograms for all latent variables



Conclusion

It is recommended that researchers test for normality when the assumption of normality is in doubt. In this paper, we illustrate how to assess univariate and multivariate normality in the context of structural equation modeling via partial least squired (PLS-SEM). Our findings indicate that the assumption of multivariate normality is violated, which justifies the use of composite-based or factor-based PLS-SEM algorithms due to their robustness to deviations from normality (Kock, 2016; 2019a; 2019b; 2019c; Kock & Hadaya, 2018). WarpPLS proves to be a powerful tool for analyzing the data inasmuch as it provides composite-based PLS-SEM algorithms together with factor-based PLS-SEM algorithms.

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.

Ezeugwa, B., Talukder, M. F., Amin, M. R., Hossain, S. I., & Arslan, F. (2022). Minimum sample size estimation in SEM: Contrasting results for models using composites and factors. *Data Analysis Perspectives Journal*, 3(4), 1-7.

Gel, Y. R., & Gastwirth, J. L. (2008). A robust modification of the Jarque–Bera test of normality. *Economics Letters*, 99(1), 30-32.

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. (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. (2019c). Factor-based structural equation modeling: Going beyond PLS and composites. *International Journal of Data Analysis Techniques and Strategies*, 11(1), 1-28.

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

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

Kock, N. (2022c). 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.

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.

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics (6th ed.)*. Boston, MA: Pearson.

Yazici, B., & Yolacan, S. (2007). A comparison of various tests of normality. *Journal of Statistical Computation and Simulation*, 77(2), 175-183.