Journal of Cases on Information Technology
Volume 27 • Issue 1 • January-December 2025

Decoding Structural Equation Modeling:
Insights on Data Assumptions, Normality, and Model Fit in Advancing Digital Marketing Strategies

Jack Ng Kok Wah
https://orcid.org/0000-0002-3055-953X
Multimedia University, Cyberjaya, Malaysia

ABSTRACT

The study utilizes structural equation modeling to examine issues related to normality, missing data, and sampling errors in digital marketing engagement research. The primary focus is on exploring relationships between self-esteem, social comparison, social interactions, perceived social support, and psychological well-being, with perceived social support as a mediating factor. Confirmatory factor analysis is applied to evaluate model fit using data from 400 social media users. Skewness and Kurtosis values are assessed to ensure normality, with scores kept within the acceptable range of -2 to +2. Questionnaires with over 30% missing values are excluded to maintain data quality, and the "10-times rule" is used to ensure adequate sample size and reduce sampling errors. Results confirm a normal distribution and indicate that the model aligns with SEM assumptions, meeting all fit indices. The research offers insights into SEM's application in digital marketing and suggests future studies should investigate advanced modeling techniques for further exploration.

KEYWORDS

Structure Equation Modelling, Normality Assumptions, Model Fit Measures, Digital Marketing Technologies, Missing Data Check

INTRODUCTION

Modern research tools and techniques are essential for informed decision-making across various sectors (Kumar & Upadhaya, 2017). A notable advancement is the introduction of semPower by Moshagen and Bader (2024), a tool that facilitates power analysis in structural equation modeling (SEM). This tool enhances researchers' ability to determine appropriate sample sizes, thereby improving the robustness and reliability of findings in behavioral research methodologies. The study underscores the significance of power analysis for deriving valid study conclusions. SEM, a pivotal research technique in social media engagement, allows researchers to explore complex variable relationships. Vargör and Öğretmen (2024) compared covariance-based SEM (CB-SEM) with partial least squares SEM (PLS-SEM), emphasizing that while CB-SEM is more robust for theory testing, PLS-SEM excels in predictive accuracy and handling complex models with small sample sizes, making it ideal for exploratory research. Before analyzing data, the study outlines SEM's foundational aspects, including modeling criteria, assumptions, and key concepts. SEM is crucial for assessing normality, handling missing data, and measuring sampling errors, ensuring robust research findings. Confirmatory factor analysis (CFA) plays a central role in evaluating model fit, starting

DOI: 10.4018/JCIT.369092

with a predetermined structure that anticipates latent factors and their associated indicator variables. For example, Du et al. (2024) assess SEM fit indices and cutoff values in network psychometrics, highlighting the shift from exploratory to confirmatory network analysis. Their findings emphasize the importance of appropriate fit measures and cutoff thresholds to ensure model validity and enhance psychological network models' robustness.

Issues and Gaps

SEM has long been a foundational tool in advanced statistical analysis across various fields. Despite its extensive application, the correct understanding and application of SEM remain elusive for many researchers, with several issues and gaps hindering its optimal use, especially in relation to data assumptions, normality, and model fit. One of the primary issues is the inadequate understanding and application of key assumptions in SEM, particularly the assumption of normality. Many SEM studies fail to address or justify the data assumptions they rely on, which undermines the robustness and reliability of their findings. The normality assumption is critical for several estimation techniques, yet it is often not properly tested, leading to the use of data that is not normally distributed and resulting in biased estimates and invalid conclusions (Hair et al., 2014).

While some researchers attempt to address non-normality by using methods such as bootstrapping, there remains significant debate over which techniques are most appropriate for different types of data (Nevitt & Hancock, 2001). Another significant challenge in SEM is the issue of missing data, which is a common occurrence in large-scale research. The improper treatment of missing values can have a substantial impact on research outcomes. Although methods such as multiple imputation (MI) have been proposed to handle missing data, Amusa and Hossana (2024) emphasize that there is no consensus on the best approach for PLS-SEM, leading to estimation inaccuracies and flawed model fit assessments. Inadequate handling of missing data is a gap that continues to affect the reliability of SEM results, as it can introduce bias and distort conclusions drawn from the analysis.

Furthermore, the evaluation of model fit presents another critical gap in SEM research. SEM model fit indices, including chi-square, root mean square error of approximation (RMSEA), and comparative fit index (CFI), are commonly used to assess the adequacy of a model, yet these indices are not without their limitations. Sole reliance on these indices can lead to misleading conclusions, as they may not fully capture the complexity of a model's fit. Scholars like Goretzko et al. (2024) and Schermelleh-Engel et al. (2003) argue that researchers must integrate multiple fit indices and conduct sensitivity analyses to ensure the robustness of their model fit assessments. Over-reliance on a single fit index without a comprehensive understanding of their limitations is a recurring issue in SEM practice.

Researchers often fail to consider alternative model evaluation methods that could provide more accurate assessments of model fit, such as Bayesian methods or cross-validation techniques (Boomsma, 2000). These limitations in evaluating model fit, along with the broader issues of data assumptions and missing data, create gaps in the proper application of SEM, which hinder its potential to provide accurate and reliable results in research. Addressing these issues through better understanding and application of SEM techniques is crucial for advancing the field and ensuring the validity of findings.

Another significant gap in SEM research is the lack of exploration into how the choice between CB-SEM and PLS-SEM impacts the results. These two SEM approaches can yield different outcomes depending on the type of data and research questions being addressed. However, as noted by Vaithilingam et al. (2024), there is a lack of research that explicitly compares these methods and offers clear guidelines on when to use each approach. This gap causes confusion among researchers, making it difficult for them to determine the most suitable modeling technique for their specific research context. The absence of definitive guidance on the strengths and limitations of CB-SEM and PLS-SEM leaves many researchers unsure about which method will provide the most accurate and reliable results. Furthermore, while existing studies offer comprehensive theoretical guidelines on