WEP Warwick Evans Publishing

International Journal of Social Sciences and Public Administration

ISSN: 3005-9836 (Print), ISSN: 3005-9585 (Online) | Volume 6, Number 3, Year 2025
DOI: https://doi.org/10.62051/ijsspa.v6n3.17
Journal homepage: https://ijsspa.org

Applications of Structural Equation Modelling in Social Sciences Research

Wang Jing[1,2], Ali Salman[3]

[1]Infrastructure University Kuala Lumpur, Selangor, Malaysia
[2]Zhengzhou University of Technology, Zhengzhou, China
[3]Universiti Malaysia Kelantan, Kelantan, Malaysia

ABSTRACT

Structural Equation Modeling (SEM) has emerged as a powerful analytical tool in social sciences research, offering a robust approach to examining complex relationships among observed and latent variables. Unlike traditional regression models, SEM enables researchers to simultaneously analyze multiple dependent and independent variables, account for measurement errors, and evaluate theoretical models comprehensively. This paper provides an overview of SEM applications in social sciences, emphasizing its advantages, key methodological steps, and limitations.

KEYWORDS

Structural Equation Modeling; Social Sciences; Multivariate Analysis; Research Methodology; Model Validation.

1. INTRODUCTION

Structural Equation Modeling (SEM) is a multivariate statistical technique that integrates factor analysis and path analysis to examine complex relationships among variables Bollen [1]. It accomplishes this with two main components, which are the measurement model and the structural model. The measurement models focus on validating the proposed measurements by examining how latent variables relate to their indicators. Meanwhile, the structural models assess the hypothesized relationships between latent variables, allowing researchers to test the statistical hypotheses that guide the study. This approach is particularly valuable when analyzing latent constructs, such as attitudes, motivations, and behavioral intentions, which cannot be directly observed [2]. SEM has gained prominence in social sciences due to its ability to simultaneously estimate multiple relationships and account for measurement errors [3]. Therefore, SEM has been widely applied in psychology, sociology, business, and communication research [4]. This paper explores the methodological framework of SEM, its applications, and emerging challenges.

Since SEM involves a lot of specialized terminology, especially related to the different types of variables proposed in the model, providing an initial definition of key terms helps make the findings easier to understand. Some of the frequently used terms in SEM are:

Structural Equation Model: This comprehensive model combines both the structural model and the measurement model, encompassing all measured and observed variables.

Measurement Model: This refers to the portion of the structural equation model that includes all observed variables linked to the latent variables, along with their relationships, variances, and errors.

Structural Model: This part of the overall structural equation model includes both latent and observed variables.

Exogenous Variables: These are variables that are not affected by other variables in the model.

Endogenous Variables: These are variables that are influenced by other variables in the model.

Indicator Variables: These are variables that can be directly observed and quantified, also referred to as manifest variables in some contexts.

Latent Variables: These are variables that cannot be directly measured.

2. ADVANTAGES OF SEM

SEM distinguishes itself from traditional statistical methods through its unique characteristics. First, according to Hu and Bentler [5], SEM relies on covariance as its primary statistic, offering more detailed insights than correlation. While conventional regression analysis focuses on minimizing the differences between observed and predicted individual data points, SEM seeks to minimize the discrepancies between observed and predicted covariance matrices. In essence, SEM aims to uncover the patterns of relationships among variables and explain as much of their variance as possible [6]. Second, SEM allows for the inclusion of latent variables in the analysis, extending beyond the relationships between observed variables and constructs. It enables the simultaneous examination of multiple dependent relationships, while also accounting for potential measurement errors in all variables [7]. Third, SEM employs confirmatory factor analysis to reduce measurement error by evaluating multiple indicators for each latent variable, while also enhancing model visualization through its graphical interface [4]. Four, SEM provides the benefit of evaluating complete models instead of just individual coefficients. It can assess models with several dependent variables, include mediating variables, and model error terms for all indicator variables [5]. Last, SEM takes into consideration measurement errors across all variables. When the fit indices of a proposed structural model are inadequate, it enables specification searches to discover models that better match the sample variance-covariance matrix [8].

SEM process involves two main steps: validating the measurement model and fitting the structural model [2]. In the first step, the measurement model is validated using Confirmatory Factor Analysis (CFA), which tests how well observed variables represent latent constructs and assesses their reliability and construct validity. The second step focuses on fitting the structural model, which examines the relationships between latent variables using path analysis to evaluate the significance and strength of these relationships, including direct, indirect, and total effects. This two-step approach allows SEM to test complex hypotheses by validating measurement models and assessing the relationships between latent constructs simultaneously. For estimating variable relationships with SEM, various software programs like CALIS, EQS, AMOS, and LISREL are available. AMOS is often preferred for its versatility across all stages of data analysis [9, 10].

3. METHODOLOGICAL STEPS IN SEM

The primary objective of creating a hypothesized diagram in SEM is to identify a model that fits the data effectively from a statistical standpoint and ensures that each parameter in the model holds a meaningful, substantive interpretation [11]. The development of a SEM generally involves the following five steps:

1)Model Specification: Develop a hypothesized model grounded in conceptual principles, outlining the relationships between both observed and latent variables.

2)Model Identification: Verify that the model is identifiable, ensuring there is enough data to estimate all required parameters.

3)Model Estimation: Apply estimation methods, such as maximum likelihood estimation, to calculate the model parameters.

4)Model Fit Evaluation: Assess the fit of the model to the data using various fit indices, including absolute fit indices, incremental fit indices, and parsimony fit indices, to determine the model's adequacy.

5)Model Modification: If the model fit is inadequate, adjust it by adding paths or changing the structure, ensuring the changes are supported by theory.

3.1. Model Specification

In SEM, model specification refers to the process of defining the theoretical relationships between variables, both observed and latent, in a way that reflects the researcher's hypotheses or theoretical framework [4]. The accuracy and clarity of model specification are essential, as any errors in specification can result in biased parameter estimates and incorrect conclusions [1]. This step involves determining which variables are endogenous and exogenous, as well as specifying the direct, indirect, and covariance relationships among them [4]. According to Kline [4], the model specification should align with existing theory or empirical findings, ensuring that the relationships between constructs make sense and reflect the research objectives. It is important to consider the underlying assumptions about the structure of the data and to ensure that the model is not overly complex or too simplistic. Kline [4] further emphasizes the need for parsimonious models that balance fit with interpretability. This balance is important, as overly complex models may lead to overfitting, where the model fits the sample data well but does not generalize to new data. Overall, model specification generally starts with a conceptual model, which is then formalized into a statistical model [8] . Theoretical understanding, empirical evidence, and model simplicity are key factors in making these decisions [12]. Therefore, model specification is not just a technical step, but one that requires a strong theoretical foundation and careful judgment to ensure the model's validity and relevance.

3.2. Model Identification

In SEM, model identification refers to the process of determining whether a model has enough information to estimate all of its parameters uniquely. According to Kline [4], a model is considered identified if the number of independent data points, such as the sample size and the number of observed variables, is greater than or equal to the number of parameters to be estimated. Proper model identification is essential for the estimation process to produce meaningful and reliable results. If a model is not identified, it means that there are insufficient data to estimate the parameters, leading to indeterminate solutions. Kline [4] distinguishes between three types of identification: just-identified, under-identified, and over-identified models. A just-identified model has exactly as many parameters as there are data points available, meaning that the model is fully specified but cannot be tested for goodness of fit. An under-identified model has fewer data points than parameters to be estimated, leading to multiple possible solutions. In contrast, an over-identified model has more data points than parameters, which allows for testing the model's fit to the data. Ensuring proper identification is a critical step in the SEM process, as it affects the reliability and interpretability of the estimated parameters.

3.3. Model Estimation

In SEM, model estimation refers to the process of calculating the values of the model parameters, such as path coefficients, variances, and covariances, based on the observed data. According to Kline [4], model estimation is a critical step in SEM because it directly influences the quality and accuracy of the results. The estimation process involves using a chosen estimation method to find the best-fitting values for the model parameters that minimize the discrepancy between the observed data and

the predicted model. Kline [4] highlights that the most commonly used estimation methods in SEM are maximum likelihood estimation (MLE). MLE is a technique for estimating one or more parameters for a given statistic by maximizing the known likelihood, which represents the hypothetical probability of a past event leading to a specific outcome distribution. This method identifies the parameters of the model that are most consistent with the observed data. These parameters include the variances and covariances of latent variables, direct effects (path coefficients) on the dependent variable, and the variances of the disturbances (residual errors). When SEM applies MLE, a crucial assumption is that the observed variables adhere to a multivariate normal distribution. Any deviation from this assumption can result in biased parameter estimates and unreliable standard errors [2]. The method chosen for estimation has significant implications for the quality of the results, and researchers have to carefully consider the characteristics of their data and the assumptions underlying each estimation method.

3.4. Model Fit Evaluation

Model fit evaluation in SEM is the process of assessing how well a specified model aligns with the observed data. According to Kline [4], model fit is a critical component of SEM because it indicates the extent to which the theoretical assumptions of model are supported by the data. The goal of model fit evaluation is to ensure that the relationships specified in the model accurately reflect the underlying data structure, thus providing reliable results and interpretations. Model fit is not a single measure, but rather a combination of various indices that together provide a comprehensive assessment of the model's performance. Kline [4] identifies several categories of fit indices that are commonly used in SEM, including absolute fit indices, incremental fit indices, and parsimony-based fit indices. Absolute fit indices evaluate the overall fit of the model without considering alternative models. Incremental fit indices compare the specified model with a baseline model, typically a null model, which assumes no relationships between variables. Parsimony-based fit indices take into account the model's complexity, penalizing models that are overly complex. Table 1 shows categories of model fit indices and acceptable thresholds.

Table 1. Categories of Model Fit Indices and Acceptable Thresholds

Name of Category	Name of Index	Level of Acceptable and Remarks
Absolut Fit	*p* value	$p > 0.05$ [5]
	GFI	GFI > 0.90 Great; GFI > 0.80 Acceptable [5]
	RMSEA	RMSEA < 0.08 [5]
Incremental Fit	AGFI	AGFI > 0.90 Great; AGFI > 0.80 Acceptable [13]
	CFI	CFI > 0.90 Great; CFI > 0.80 Acceptable [5]
	NFI	NFI > 0.90 Great; NFI > 0.80 Acceptable [5]
	NNFI(/TLI)	NNFI > 0.90 Great; NNFI > 0.80 Acceptable [5]
Parsimonious Fit	Chi-square/*df*	$\chi 2/df < 3$ [7]

3.5. Model Modification

Model modification in SEM refers to the process of adjusting the initial model in order to improve its fit to the observed data. According to Kline [4], model modification is a common step in SEM, especially when the initial model does not fit the data adequately. The goal of model modification is to refine the model so that it better reflects the underlying relationships among the variables, while maintaining theoretical consistency and parsimony. However, Kline [4] emphasizes that modifications should be made thoughtfully, as they can lead to overfitting and a loss of generalizability. Therefore, any modifications made should be theoretically justified and based on empirical evidence. Model modification is often guided by the modification indices (MI), which suggest changes to the model by highlighting areas where parameter estimates could be added or altered to improve the model fit. Modification indices provide a measure of how much the overall model fit would improve if a particular parameter, such as a path or covariance, were freely estimated. Modification indices should be used carefully, as blindly following them can result in a model that fits the data but lacks theoretical meaning or generalizability.

4. LIMITATIONS OF SEM

SEM, while a powerful statistical tool widely used in social sciences, is subject to multiple limitations. Kline [4] systematically critiques the inherent constraints of SEM, which can be categorized into four key areas.

First, SEM imposes strict requirements on data quality and sample size. The stability of parameter estimates heavily relies on large samples, particularly for complex models, to avoid issues such as non-convergence or improper solutions such as negative error variances. Furthermore, traditional SEM methods assume multivariate normality. Violations of this assumption, including skewed data or outliers, may bias results, necessitating corrections through techniques like bootstrapping or robust estimation methods.

Second, model specification is inherently theory-dependent and prone to subjectivity. SEM is fundamentally confirmatory, requiring researchers to predefine all hypothesized relationships among variables. Weak theoretical foundations or omitted critical variables can lead to model misspecification and biased conclusions. Kline [4] specifically criticizes the common practice of post hoc model modifications, such as adding paths based on modification indices, arguing that such adjustments risk capitalizing on random noise and undermining the confirmatory nature of the analysis.

Third, overreliance on and misinterpretation of fit indices remain significant concerns. Although indices such as RMSEA and CFI are often treated as gold standards for evaluating model quality, Kline [4] emphasizes that these metrics are heuristic tools rather than definitive criteria. For instance, chi-square tests in large samples may overreact to trivial specification errors, rejecting models that are practically reasonable, while small samples may fail to detect genuine flaws. An excessive focus on achieving ideal fit values can obscure structural weaknesses in the model, such as omitted confounding variables.

Fourth, balancing model complexity with parsimony poses challenges. Researchers often favor constructing complex models with excessive latent variables or paths, but this risks identification issues, including non-unique parameter estimates, or reduced interpretability. Kline [4] stresses that model design should prioritize theoretical clarity over unnecessary complexity. Additionally, inadequate measurement models, such as insufficient indicators for latent variables, may render models under-identified, requiring rigorous upfront planning to avoid such pitfalls.

To address these limitations, Kline [4] proposes several strategies. First, researchers should anchor their work in robust theoretical frameworks, minimizing reliance on post hoc adjustments. Second,

transparency is critical, multiple fit indices, standardized residuals, and parameter estimates should be comprehensively reported to avoid selective presentation of results. Third, robustness checks, such as bootstrapping or simulation studies, should be employed to validate model stability. Finally, researchers must explicitly acknowledge SEM's inherent limitations in their discussions.

5. CONCLUSION

In conclusion, SEM has proven to be an invaluable tool in social sciences research, providing a sophisticated method for analyzing intricate relationships between variables. Its ability to handle complex models, address measurement errors, and evaluate both observed and latent variables makes it a preferred choice over traditional analytical approaches. However, researchers must be mindful of its limitations, including sample size requirements, assumptions of normality, model fit, and model complexity. By understanding both the strengths and challenges of SEM, researchers can better apply this methodology to advance theory and empirical research in the social sciences. Future research should continue to refine SEM techniques and explore ways to mitigate its limitations, ensuring its continued relevance and utility in the field.

ACKNOWLEDGEMENT

I wish to express my deepest gratitude to my supervisor, Assoc Prof Dr Ali, and my co-supervisor, Prof Dr Faridah, for their unwavering support throughout this project. Their patient and professional guidance have been invaluable in shaping the direction and scope of this research. Their dedication and commitment to fostering my development as a researcher have been both motivating and deeply appreciated. This work would not have reached its present form without their mentorship, for which I am truly thankful.

REFERENCES

[1] Bollen, Kenneth A. "Structural equations with latent variables", Applied John Wiley & Sons, 1989.

[2] Byrne, Barbara M. "Structural equation modeling with Mplus: Basic concepts, applications, and programming", Applied Routledge, 2013.

[3] Jeff Risher, Marko Sarstedt, Christian M. Ringle. "When to use and how to report the results of PLS-SEM", Applied European Business Review, Vol.31, No.1, pp. 2-24, 2019. https://doi.org/10.1108/EBR-11-2018-0203

[4] Rex B. Kline, "Principles and practice of structural equation modeling", Applied Guilford Publications, 2023.

[5] Li-tze Hu, Peter M. Bentler. "Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives", Applied Structural Equation Modeling: A Multidisciplinary Journal, Vol.6, Issue.1, pp 1-55, 1999. https://doi.org/10.1080/10705519909540118

[6] Principles and practice of structural equation modeling. Guilford publications, 2023. Nicole Franziska Richter, Rudolf R. Sinkovics, Christian M. Ringle, Christopher Schlägel. "A critical look at the use of SEM in international business research" Applied International Marketing Review, Vol.33, Issue.3, pp 376-404. https://doi.org/10.1108/IMR-04-2014-0148

[7] Joseph F. Hair Jr, William C. Black, Barry J. Babin, Rolph E. Anderson. "Multivariate data analysis", Applied Cengage, 2019.

[8] Marley W Watkins. "Exploratory Factor Analysis: A Guide to Best Practice", Applied Black psychology, Vol.44, Issue.3, pp 219-246, 2018. https://doi.org/10.1177/0095798418771807

[9] Darmaraj Sakaria, Siti Mistima Maat, Mohd Effendi Ewan Mohd Matore. "Examining the Optimal Choice of SEM Statistical Software Packages for Sustainable Mathematics Education: A Systematic Review", Applied Sustainability, Vol.15, Issue.4, pp 47-71, 2023. https://doi.org/10.3390/su15043209

[10] Damian Gallagher, Lucy Ting, Adrian Palmer. "A journey into the unknown; taking the fear out of structural equation modeling with AMOS for the first-time user", Applied The Marketing Review, Vol. 8, No. 3, pp. 255-275, 2008. https://doi.org/10.1362/146934708X337672

[11] Karl G. Jöreskog. “Structural analysis of covariance and correlation matrices”, Applied Psychometrika, Vol. 43, Issue. 4, pp. 443-477,1978. https://doi.org/10.1007/BF02293808

[12] Kenneth A. Bollen, John Scott Long. “Testing structural equation models”, Applied Sage, 1993.

[13] Daire, H., C. Joseph, and R.M. Michael, “Structural Equation Modelling: Guidelines for Determining Model Fit Structural equation modelling: guidelines for determining model fit”, Applied Electron J Bus Res Methods, Vol.6, pp.53–60, 2008.