Open Access Indonesia Journal of Social Sciences

Journal Homepage: https://journalsocialsciences.com/index.php/OAIJSS

Structural Equation Modelling (SEM) in Research: Narrative Literature Review

Rachmat Hidayat[1*], Patricia Wulandari[2]

[1]Department of Biology, Faculty of Medicine, Universitas Sriwijaya, Palembang, Indonesia

[2]Cattleya Mental Health Center, Palembang, Indonesia

ARTICLE INFO

Keywords:
Data analysis
Predicting power
Social research
Structural equation modelling
Variable

***Corresponding author:**
Rachmat Hidayat

E-mail address:
rachmathidayat@fk.unsri.ac.id

All authors have reviewed and approved the final version of the manuscript.

https://doi.org/10.37275/oaijss.v5i6.141

ABSTRACT

The structural equation modelling (SEM) method has stronger predicting power than path analysis and multiple regression because SEM is able to analyze at the deepest level the variables or constructs studied. This literature review aimed to describe the use of structural equation modelling in research. In general, SEM can be used to analyze research models that have several independent (exogenous) and dependent (endogenous) variables, as well as moderating or intervening variables. SEM provides several benefits and advantages for researchers, including building research models with many variables, examining variables or constructs that cannot be observed or cannot be measured directly (unobserved), testing measurement errors (measurement errors) for observed variables or constructs (observed) and confirmatory factor analysis. Broadly speaking, SEM methods can be classified into two types, namely covariance-based structural equation modelling (CB-SEM) and variance or component-based SEM (VB-SEM), which includes partial least squares (PLS) and generalized structured component analysis (GSCA). This literature review aimed to describe the use of structural equation modelling in research.

1. Introduction

Human nature wants to continue to progress and develop in order to achieve a better quality of life. This also happens in the world of research. Experts in the social or behavioral sciences, including management, consistently develop research methods that can be used to obtain better, perfect, fast, accurate, effective, and efficient quality research results (Burhan, 2011). Experts in the field of social or behavioral sciences, including management, have developed a research method called structural equation modelling (SEM) (Byrne, 2013). At first, the SEM method was only good at the conception level. At that time, the SEM method could not be operationalized due to technological limitations. With the rapid development of computer technology, the SEM method is now becoming increasingly recognized and widely used in behavioral and management research (Capmourteres, 2016). The SEM method is a development of path analysis and multiple regression, which are both forms of multivariate analysis models. In an associative, multivariate-correlational, or causal-effect analysis, the SEM method seems to break the domination of the use of path analysis and multiple regression, which have been used for decades. Compared to path analysis and multiple regression, the SEM method is superior because it can analyze data more comprehensively (Chang, 1981). Data analysis in path

This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License

analysis and multiple regression was only carried out on the total variable score data, which is the sum of the research instrument items. Thus, path analysis and multiple regression are actually only carried out at the level of latent variables (unobserved). In comparison, data analysis in the SEM method can penetrate deeper because it is carried out on each item score of a research variable instrument. Instrument items in SEM analysis are referred to as manifest variables (observed) or indicators of a construct or latent variable (Chen, 2010).

The SEM method has stronger predicting power than path analysis and multiple regression because SEM is able to analyze at the deepest level the variables or constructs studied (Cohen, 2013). The SEM method is more comprehensive in explaining research phenomena. Meanwhile, path analysis and multiple regression are only able to reach the level of latent variables, so they experience a dead end in parsing and analyzing empirical phenomena that occur at the level of items or indicators of latent variables. Judging from the data used, path analysis and multiple regression actually only reach the outer shell of a research model (Cudeck, 1994). In comparison, the SEM method can be likened to being able to reach as well as parse and analyze the deepest entrails of a research model. The SEM method is expected to be able to answer the weaknesses and impasses faced by the previous generation of multivariate methods, namely path analysis and multiple regression (Curran, 2003). The development of SEM methods is becoming increasingly significant in the practice of social, behavioral, and management research, along with advances in information technology (Duncan et al., 2013). Many multivariate statistical methods which were difficult to operate manually in the 1950s, such as factor analysis, multiple regression with more than three independent variables, path analysis, and discriminant analysis, gradually became necessary because of the invention of computer programs such as SPSS (Statistical Package for Social Science), Minitab, Prostat, QSB, SAZAM, etc. The SEM method is currently estimated to be the most dominant multivariate method. Computer programs that can currently be used to process data in SEM research methods include AMOS, LISREL, PLS, GSCA, and TETRAD. This literature review aims to describe the use of structural equation modeling in research (Eisenhauer et al., 2015).

The benefits of SEM in research

In general, SEM can be used to analyze research models that have several independents (exogenous) and dependent (endogenous) variables, as well as moderating or intervening variables (Fan et al., 1999). SEM provides several benefits and advantages for researchers, including building research models with many variables, examining variables or constructs that cannot be observed or cannot be measured directly (unobserved), testing measurement errors (measurement errors) for observed variables or constructs (observed), confirming the theory in accordance with research data (confirmatory factor analysis), being able to answer various research problems in a more systematic and comprehensive analysis set; more illustrative, robust and reliable than the regression model when modeling interaction, non-linearity, measurement error, correlation of error terms, and correlation between multiple independent latent variables; used as an alternative to path analysis and covariate-based time series data analysis; factor, path and regression analysis; explain the complex interrelationships of variables and direct or indirect effects of one or several variables on other variables; and has higher flexibility for researchers to relate the theory with data (Fritz et al., 2007; Grace, 2006).

Types of SEM

As stated above, in general, the SEM method can be classified into two types, namely covariance-based structural equation modelling (CB-SEM) and variance

This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License

or component-based SEM (VB-SEM), which includes partial least squares (PLS) and generalized structured component analysis (GSCA) (Grace, 2008; Grace, 2010). A variant is the deviation of the data from the mean (average) value of the sample data. Variance measures the deviation of data from the mean value of a sample, so it is a measure of metric variables. Mathematically, the variance is the average of the squared differences between each observation and the mean, so the variance is the average squared value of the standard deviation (Haavelmo, 1943). A variable must have a variance that is always positive. If it is zero, then it is not a variable but a constant. Meanwhile, covariance shows a linear relationship that occurs between two variables, namely X and Y. If a variable has a positive linear relationship, then the covariance is positive. If the relationship between X and Y is opposite, then the covariance is negative. If there is no relationship between the two variables, X and Y, then the covariance is zero.

Covariance-based structural equation modelling (CB-SEM)

Covariance-based SEM (CB-SEM) was first developed by Joreskog (1973), Keesling (1972), and Wiley (1973). CB-SEM became popular after the availability of the LISREL III program developed by Joreskog and Sorbom in the mid-1970s. By using the maximum likelihood (ML) function, CB-SEM tries to minimize the difference between the sample covariance matrix and the covariance matrix predicted by the theoretical model so that the estimation process produces a residual covariance matrix with a small value close to zero. Some things that need to be considered in CB-SEM analysis include the following:
a) The assumption of using CB-SEM is like the parametric analysis. The assumptions that must be met are that the observed variables must have a multivariate normal distribution, and the observations must be independent of one another. If the sample is small and not asymptotic, it will give poor parameter estimates and statistical models or even produce a negative variance, which is called the Heywood Case.
b) A small sample size will potentially result in a Type II error, i.e., a bad model will still result in a fit model.
c) CB-SEM analysis requires the form of latent variables whose indicators are reflective. In the reflective model, indicators or manifest are considered variables that are influenced by latent variables according to the classical measurement theory. In the reflective indicator model, indicators in a construct (latent variable) are influenced by the same concept. Changes in one item or indicator will affect changes in other indicators in the same direction. The examples referred to as reflective variables are:

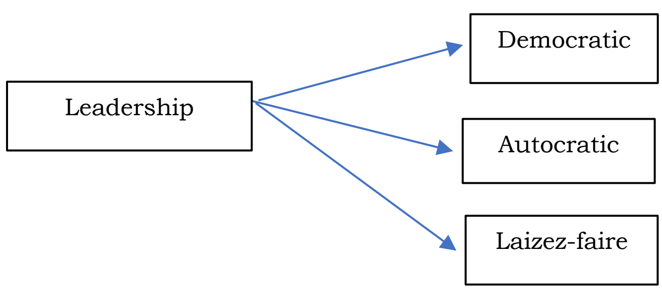

Figure 1. An example of a reflective variable from a latent (construct) variable. Democratic, autocratic, and Laizez-faire are reflective variables of leadership. Reflective variables are variables that stay away from latent (construct) variables, as shown in the blue arrows above.

This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License

Variance-based SEM (VB-SEM)

PLS-SEM

PLS-SEM aims to test predictive relationships between constructs by seeing whether there is a relationship or influence between these constructs (Hair et al., 2013). The logical consequence of using PLS-SEM is that testing can be carried out without a strong theoretical basis, ignoring some assumptions (non-parametric) and the parameter accuracy of the prediction model seen from the value of the coefficient of determination (R^2). PLS-SEM is very appropriate for use in research that aims to develop theory. PLS-SEM was developed to overcome tests that cannot be done with CB-SEM. (Harrington, (2009). For example, in testing formative variables, the examples of formative variables are as follows:

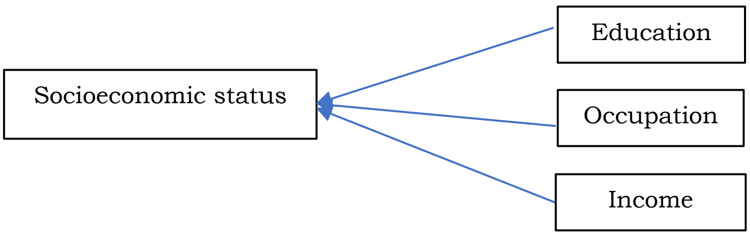

Figure 2. An example of a formative variable from latent (construct) variables. Education, Occupation, and Income are formative variables of socioeconomic status. Formative variables are variables that lead to or influence or form latent (construct) variables, as shown in the blue arrow above.

GSCA

GSCA combines the characteristics found in CB-SEM and PLS-SEM. GSCA can handle latent variables with many indicators, the same as PLS-SEM, requiring goodness of fit model criteria, and indicators and constructs must be correlated like CB-SEM. Until now, the GSCA method is rarely used widely by researchers because this method is relatively new. GSCA has the same goal as PLS-SEM, does not require the assumption of multivariate normality data, and can be tested without a strong theoretical basis with a small number of samples (Hoyle, 2013).

Model covariance-based SEM (CB-SEM) is often called hard modeling, while component-based or variance-based SEM (VB-SEM) modeling is called soft modeling. Hard modeling aims to provide a statement about the causality relationship or provide a description of the mechanism of the causality relationship (cause and effect). This provides an ideal picture scientifically in data analysis. However, the data to be analyzed does not always meet the ideal criteria, so it cannot be analyzed by hard modeling (Hu, 1999). As a solution, soft modeling tries to analyze data that is not ideal. Literally, soft actually means soft or soft, but in the research context, soft is defined as not based on assumptions on the scale of measurement, data distribution, and sample size (Iacobucci, 2010). The main purpose of analysis with hard modeling is to test the causal relationship between those that have been built based on the theory and whether the model can be confirmed with empirical data. In comparison, the main objective of soft modeling analysis aims to find predictive linear relationships between latent constructs. It should be understood that a causality or estimation relationship is not the same as a predictive relationship (Jackson et al., 2009). In terms of causality, CB-SEM looks for invariant parameters that structurally or functionally describe how the world's systems work. The invariant parameter describes the causal relationship between variables in a closed system so that events can be fully controlled. Whereas in Partial Least Square, Variance,

This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License

or Component-Based SEM, the optimal linear relationship between latents is calculated and interpreted as the best available predictive relationship with all the limitations that exist (Joreskog, 1993). So that the existing events can not be fully controlled, if the data to be analyzed meets all the assumptions required by CB-SEM, then the researcher should analyze the data by hard modelling using appropriate software, such as AMOS and LISREL (Kim, 2005).

If the data does not meet all the required assumptions, but the researcher still uses hard modelling or CB-SEM analysis, then several problems may be encountered, an improper solution or an imperfect solution because of the Heywood Case, which is a symptom of a negative variance value; the model becomes unidentified due to indeterminacy; and non-convergence algorithms. If that conditions occur and we still want to analyze the data, then our goal is not to change causality between variables but to find optimal predictive linear relationships using component or variance-based SEM (Lamb et al., 2014).

Based on the objective of empirical research, the quantitative paradigm can be divided into two, namely estimation and prediction. Estimation research is research that aims to test an empirical model with valid and reliable measurements. Testing and measurement are carried out at the indicator level. The hypothesis being tested is the model hypothesis. The measurement criterion for testing the feasibility of the model is called the goodness of fit test (LeCun et al., 2015). For estimation research purposes, CB-SEM is an appropriate technique to use. Prediction research is research that aims to examine the influence between constructs to predict causal relationships. Testing and measurement are carried out at the level of constructs or latent variables (McDonald, 2002). The hypothesis that is done is generally a partial hypothesis. Partial testing criteria with a significance test predicting the relationship between variables using the t-statistic test. PLS-SEM and regression techniques are the right choices of statistical techniques to use (Mulaik et al., 1989; Murtaugh, 2009). Therefore, component or variance-based SEM (PLS and GSCA) is only used if the data we have cannot be solved with covariance-based SEM (CB-SEM).

2. Conclusion

SEM can be used to analyze research models that have several independent and dependent variables as well as moderating or intervening variables.

3. References

Burnham KP, Anderson DR, Huyvaert KP. 2011. AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. Behav Ecol Sociobiol. 65(1): 23–35.

Byrne BM. 2013. Structural equation modeling with AMOS: basic concepts, applications, and programming. Routledge, New York.

Capmourteres V, Anand M. 2016. Assessing ecological integrity: a multi-scale structural and functional approach using structural equation modeling. Ecol Indic.

Chang WY. 1981. Path analysis and factors affecting primary productivity. J Freshwater Ecol. 1(1): 113–20.

Chapin FS, Conway AJ, Johnstone JF, Hollingsworth TN, Hollingsworth J. 2016. Absence of net long-term successional facilitation by alder in a boreal Alaska floodplain. Ecology.

Chaudhary VB, Bowker MA, O'Dell TE, Grace JB, Redman AE, et al. 2009. Untangling the biological contributions to soil stability in semiarid shrublands. Ecol Appl. 19(1): 110–22

Chen Y, Lin L. 2010. Structural equation-based latent growth curve modeling of watershed attribute-regulated stream sensitivity to reduced acidic deposition. Ecol Model. 221(17): 2086–94

Chen F, Curran PJ, Bollen KA, Kirby J, Paxton P. 2008. An empirical evaluation of the use of fixed

This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License

cutoff points in RMSEA test statistic in structural equation models. Socio Meth Res. 36(4): 462–94

Chen J, John R, Shao C, Fan Y, Zhang Y, et al. 2015. Policy shifts influence the functional changes of the CNH systems on the Mongolian plateau. Environ Res Lett. 10(8): 085003

Cohen J. 2013. Statistical power analysis for the behavioral sciences. Academic Press, New York.

Cover TM, Thomas JA. 2012. Elements of information theory. John Wiley & Sons, Hoboken, New Jersey.

Cudeck R, Odell LL. 1994. Applications of standard error estimates in unrestricted factor analysis: significance tests for factor loadings and correlations. Psychol Bull. 115(3): 475–87

Curran PJ. 2003. Have multilevel models been structural equation models all along? Multivar Behav Res. 38(4): 529–69.

Curran PJ, West SG, Finch JF. 1996. The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. Psychol Methods. 1(1): 16–29.

Curran PJ, Bollen KA, Paxton P, Kirby J, Chen F. 2002. The noncentral chi-square distribution in misspecified structural equation models: finite sample results from a Monte Carlo simulation. Multivar Behav Res. 37(1): 1–36.

Duncan TE, Duncan SC, Strycker LA. 2013. An introduction to latent variable growth curve modeling: concepts, issues, and applications, 2nd ed. Psychology Press, New York.

Eisenhauer N, Bowker M, Grace J, Powell J. 2015. From patterns to causal understanding: structural equation modeling (SEM) in soil ecology. Pedobiologia. 58(2): 65–72.

Fan X, Sivo SA. 2005. Sensitivity of fit indexes to misspecified structural or measurement model components: rationale of two-index strategy revisited. Struct Equ Modeling. 12(3): 343–67.

Fan X, Thompson B, Wang L. 1999. Effects of sample size, estimation methods, and model specification on structural equation modeling fit indexes. Struct Equ Modeling. 6(1): 56–83.

Fritz MS, MacKinnon DP. 2007. Required sample size to detect the mediated effect. Psychol Sci. 18(3): 233–9.

Galton F. 1888. Personal identification and description. Nature. 38: 173–7.

Grace JB. 2006. Structural equation modeling and natural systems. Cambridge University Press, New York.

Grace JB, Bollen KA. 2008. Representing general theoretical concepts in structural equation models: the role of composite variables. Environ Ecol Stat. 15(2): 191–213.

Grace JB, Anderson TM, Olff H, Scheiner SM. 2010. On the specification of structural equation models for ecological systems. Ecol Monogr. 80(1): 67–87.

Haavelmo T. 1943. The statistical implications of a system of simultaneous equations. Econometrica 11(1): 1–12.

Hair JF, Hult GT, Ringle C, Sarstedt M. 2013. A primer on partial least squares structural equation modeling (PLS-SEM). Sage, Thousand Oak.

Harrington D. 2009. Confirmatory factor analysis. Oxford University Press, New York Heise DR (1975) Causal analysis. John Wiley & Sons, Oxford.

Hinton GE, Osindero S, Teh YW. 2006. A fast learning algorithm for deep belief nets. Neural Comput. 18(7): 1527–54.

Hoyle RH. 2011. Structural equation modeling for social and personality psychology. Sage, London.

Hoyle RH, Isherwood JC. 2013. Reporting results from structural equation modeling analyses in Archives of Scientific Psychology. Arch Sci Psychol. 1: 14–22.

Hu LT, Bentler PM. 1999. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct Equ Modeling. 6(1): 1–55.

This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License

Iacobucci D. 2010. Structural equations modeling: fit indices, sample size, and advanced topics. J Consum Psychol.

Jackson DL, Gillaspy JA, Purc-Stephenson R. 2009. Reporting practices in confirmatory factor analysis: an overview and some recommendations. Psychol Methods. 14(1): 6–23.

Johnson JB, Omland KS. 2004. Model selection in ecology and evolution. Trends Ecol Evol. 19(2): 101–8.

Jones CM, Spor A, Brennan FP, Breuil MC, Bru D, et al. 2014. Recently identified microbial guild mediates soil N2O sink capacity. Nat Clim Change 4(9): 801–5.

Joreskog K, Sorbom D. 1993. LISREL 8: structural equation modeling with the SIMPLIS command language. Scientific Software International, Chicago.

Kaplan D, Depaoli S. 2012. Bayesian structural equation modeling. In: Hoyle RH (ed) Handbook of structural equation modeling. Guilford, New York.

Kim KH. 2005. The relation among fit indexes, power, and sample size in structural equation modeling. Struct Equ Modeling. 12(3): 368–90.

Kline RB. 2006. Reverse arrow dynamics. Formative measurement and feedback loops. In: Hancock GR, Mueller RO (eds) Structural equation modeling: A second course. Information Age Publishing, Greenwich.

Kline RB. 2010. Principles and practice of structural equation modeling. Guilford Press, New York.

Lamb EG, Mengersen KL, Stewart KJ, Attanayake U, Siciliano SD. 2014. Spatially explicit structural equation modeling. Ecology. 95(9): 2434–42.

LeCun Y, Bengio Y, Hinton G. 2015. Deep learning. Nature. 521(7553): 436–44.

Lee SY. 2007. Structural equation modeling: a Bayesian approach. John Wiley & Sons, Hong Kong.

Lefcheck J. 2015. PiecewiseSEM: Piecewise structural equation modelling in r for ecology, evolution, and systematics. Methods Ecol Evol.

Liu X, Swenson NG, Lin D, Mi X, Umaña MN, et al. 2016. Linking individual-level functional traits to tree growth in a subtropical forest. Ecology.

MacCallum RC, Hong S. 1997. Power analysis in covariance structure modeling using GFI and AGFI. Multivar Behav Res, 32(2): 193–210.

Maddox GD, Antonovics J. 1983. Experimental ecological genetics in Plantago: a structural equation approach to fitness components in P. aristata and P. patagonica. Ecology. 64(5): 1092–9.

McDonald RP, Ho MH. 2002. Principles and practice in reporting structural equation analyses. Psychol Methods. 7(1): 64–82.

Mehta PD, Neale MC. 2005. People are variables too: multilevel structural equations modeling. Psychol Methods. 10(3): 259–84.

Monecke A, Leisch F. 2012. semPLS: structural equation modeling using partial least squares. J Stat Softw. 48(3): 1–32.

Mulaik SA, James LR, Van Alstine J, Bennett N, Lind S, et al. 1989. Evaluation of goodness-of-fit indices for structural equation models. Psychol Bull. 105(3): 430–45.

Murtaugh PA. 2009. Performance of several variable-selection methods applied to real ecological data. Ecol Lett. 12(10): 1061–8.

This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License