International Journal of Quantitative and Qualitative Research Methods

Vol.5, No.4, pp.10-16, September 2017

Published by European Centre for Research Training and Development UK (www.eajournals.org)

STRUCTURE EQUATION MODELING BASIC ASSUMPTIONS AND CONCEPTS: A NOVICES GUIDE.

Sunil Kumar[1] and Dr. Gitanjali Upadhaya[2]

[1]Research Scholar, Department of HRM & OB, School of Business Management & Studies, Central University of Himachal Pradesh

[2]Assistant Professor, Department of HRM & OB, School of Business Management & Studies, Central University of Himachal Pradesh

ABSTRACT: *The modern tools and techniques of research make the decision making easy. The present study focused on structure equation modeling (SEM) technique of research. SEM is one of widely recognized technique in research. This paper outlined basic working of SEM its modeling criteria, assumptions and concepts. A brief idea about second generation structure equation modeling was described in the paper. This study will allow the readers to develop understanding of SEM and its applications in different fields.*

KEYWORDS: Research, Structure Equation Modeling, Assumptions, Concepts, Second Generation.

INTRODUCTION

Decision making is an important task in different spheres of life. The global forces and economic openness propels the decision makers to implement those decisions which are based on the research. To assist decision makers and solve problems researchers have to continuously discover the key techniques to assist mankind. Structure Equation modeling (SEM) establish the relationship between measurement model and structural model based upon the assumptions supported by the theory. It is a combination of factor analysis and linear regression (Ullman 2001). Regression models are additive, but the Structure Equation Models are relational in nature, that makes a difference in the regression and SEM approach of decision making. Structure Equation modeling tries to justifying the acceptance or rejection of proposed hypothesis by analyzing the direct effects and indirect effects of mediators on the relationship of independent variable and dependent variable. The role of controls and moderators also analyzed with the help of SEM. All Structure Equation Models are distinguished by three characteristics (Hair & Black 2012).

- Estimation of multiple and interrelated dependence relationships
- An ability to represent unobserved concepts in these relationships and correct the measurement errors in the estimation process.
- Defining model to explain the entire set of relationship

Jöreskog (1973) outlined a general model divided into two parts: (a) structural part connecting the constructs to each other, and (b) measurement part which connects the observed variables to the latent variables.

Structural mathematical model:

ISSN 2056-3620(Print), ISSN 2056-3639(Online)

$$\eta = B\eta + L\mathfrak{m} + \varepsilon$$

Here η represent endogenous variables, $\mathfrak{m}$ is a vector of exogenous variables, ε is the error or disturbance term vector, and B and L are the regression coefficients of endogenous and exogenous variables

Measurement mathematical model:

The equations for endogenous and exogenous latent factors, which are related to observable variable via measurement equations, are defined as

$$y = {}^\wedge y + \boldsymbol{\delta}$$

$$x = {}^\wedge x + e$$

The $^\wedge y$, $^\wedge x$ are the regression coefficients of observable variables and the $\boldsymbol{\delta}$, e are residual errors.

Assumptions of Structure Equation Modeling:

Normality: Normality of observations is the first and important assumption before building the model and checking its fit indexes. The observations must draw from a continuous and multivariate normal population. But normality of data is a condition which happens rarely in the real. So the researchers use the estimation technique as per the skewness and kurtosis of data in hand. If the variable in the study reveals normality the maximum likelihood (ML) technique of approximation is used to find the estimates of parameters. But if the normality conditions of data are violated the alternative techniques like asymptotic distribution free (ADF) of estimations are used. ADF face problem with models of moderate size. Specifically, with n variables there are $u = 1/2\ n\ (n+1)$. So u is the elements required to build a model in case of non-normal data.

Missing Data: variables in study should be complete in data forms. Simply there is no missing data in any variable. The researcher talked much about the missing data treatment through the missing completely at random (MCAR) approach. This approach assumes that missing data is totally irrelevant in study, but this is not actually same. Muthen et al (1987) advocate new approach when data is missing at random (MAR), instead of using pairwise and list-wise deletion for missing data. In later studies the researchers found that the approach of Muthen and other only applicable when missing data is in small numbers. To answer the complexities in handling the missing data, specifically imputation approach is available when maximum likelihood technique is going to estimate the parameters in SEM.

Measurement and Sampling Errors: The Errors in measurement caused by biased tool and techniques used for collection of information, and errors on the part of respondents effects the model fit. The variance of given dataset also affects the standard error. As the variance increase the standard error decrease, this violates the assumptions of normality in data. (Nevitt, Hancock, & Hancock, 2014) emphasized that increasing variance doesn't affects the estimation of parameters, but it affects the approximation of errors. MacCallum, Tucker, and Briggs (2001) compare maximum likelihood (ML) and OLS techniques of estimation on simulated models, having large number of small factors, and found that OLS is better technique of approximation as compare to ML. This is because OLS makes no distribution assumption. (Chin, Peterson, & Brown, 2008) asked a key question that how perfect the

ISSN 2056-3620(Print), ISSN 2056-3639(Online)

estimations of a model that present imperfectly the real world. Past researches emphasized the role of pre-tests in handling the measurement and sampling errors.

Model Fit Indexes: In available indexes preference is given to NFI, GFI, CFI, RAMSEA, P-CLOSE and Parsimony index values for first look of model fitting indexes. The NFI, GFI and CFI should be ≥.95, RMSEA value should be ≤.60, and higher value of P-Close is required for best fit. One can check the other index values given in the table no. 1. The fitness basically defines the usability of given model drawn from the sample on the population. The parameter estimation of model is only applied on the population if the model fits well as per the population. For fit indices the Chi Square test ($\chi2$) is used. In the case of chi-square sample size makes the difference in the results. The increase sample size (value of n increase) increases the value of chi statistics. The $\chi2$ / degree of freedom should be ≤2. Statistical packages like SPSS Amoss, LISREL and other are available to check the model fit. All these packages are user friendly and make the analysis easily handled by non-statisticians also.

Table No. 1.
Several Fit Indexes and their Cut-off Criteria

Indexes	**Shorthand**	**Acceptable Fit if Data is Continuous**
Absolute Fit:		
Chi-Square	X2	Ratio to Chi Square should ≤ 2.
Consistent AIC	CAIC	Smaller is better.
Bayes information	BIC	Smaller is better.
Comparative Fit: Comparison to baseline model		
Comparative Fit Index	CFI	≥.95
Normed Fit Index	NFI	≥.95
Non-normed Fit Index	NNFI	≥.95
Incremental Fit Index	IFI	≥.95
Parsimonious Fit: Very sensitive to the sample size.		
PNFI, PGFI and other fit index can be used.		
Other Fit Index:		
Goodness of Fit Index	GFI	≥.95
Adjusted Goodness of Fit	AGFI	≥.95
Root Mean Square Residuals	RMR	Smaller is better.
Standardized RMR	SRMR	≤.08
Root Mean Square Error of approximation	RMSEA	≤.06 (Depend upon confidence limit)
P- Close		Higher is Better.

Source: The Journal of Educational Research, Vol. 99, No. 6 (Jul.-Aug., 2006).

Concepts in Structure Equation Modeling:

Causality: Structure Equation Modeling explores the cause and effect relationship between exogenous and endogenous factors by examining the direct and indirect relation. If the empirical study reflects a relationship which is not supported by theory, the researcher must have strong empirical logics and reasons to support the existing unexpected causality. Theory based causality establish a structural model for estimation of constructs relationships. The causal models are applied to data sets where information was collected for independent, intervening, and dependent variables in a single, cross-sectional study (Biddle, & Marlin, 2014).

Structural model and Measurement model: The SEM consists of two parts one the structural and other measurement. Second part is observed, and used to estimate the structural part. The estimates of structural models are observed indirectly. Measurement model gives empirical evidences, while the structural model provide framework to support the hypothesis. Path analysis is a tool to formulate the structural model, but the observed variables under the latent constructs should be unidirectional, and need composite variables.

Validity and Reliability: Validity in SEM measured as convergent and discriminant validity. Reliability reflects the results and output through the structure equation modeling. For checking reliability the Composite/ construct reliability is measured. Without the validity and reliability of the model, it is like garbage in and garbage out.

Convergent Validity (CV): Convergent validity reflects the variable loadings on the construct. CV answer weather the variable contribution to the variance of factor is valid to describe the factor accordingly or not**.** Factor loadings are the first step to observe the convergent validity. The loadings as rule of thumb should be ≥.70 for every observable variable. This also depends on the factor identification. In many studies the above condition is relaxed little is the factor is over identified.

$$\mathrm{C.V} = \sum_{i=1}^{n} \lambda\, i^{\,2}$$

$\lambda i^{\,2}$ in above mathematical equation gives the summated values of all squared factor loadings, and it should be ≥.50.

Discriminant Validity (DV): Discriminant validity differentiates the latent constructs from each other, if the covariance between the two construct is high (≥.60), the constructs are reflecting similarities, and they are not different. One can go for second order modeling by combining the constructs. This will affect the degree of freedom of model. (Chen, 2014) advocates the value of *r* with respect to other constructs to examine the construct validity.

Composite/Construct Reliability (CR): Composite reliability describes the ability of measured variables to present the latent factor. The Cronbach Alpha for reliability is not valid in the SEM. In SEM the unobserved latent factor are predicted by the observed variables, so this is the necessary condition that the variables should be reliable and have high composite reliability (≥.70).

$$C.R = \frac{\left(\sum_{i=1}^{n} \lambda i\right)^2}{\left(\sum_{i=1}^{n} \lambda i\right)^2 + \left(\sum_{i=1}^{n} \delta i\right)}$$

Mediation, moderation and controls: The concept of mediation, moderation and controls are important to establish the direct and indirect relation between independent and dependent variables. The mediation means relationship between independent and dependent variable is best define in presence of third construct, which is known as mediator. The mediation is of three types full, partial and bifurcated between independent and dependent variable. In fig. 1.1 the customer commitment acts as mediator between independent variable product and dependent variable customer loyalty. The diagram shows the direct and indirect relationships.

ISSN 2056-3620(Print), ISSN 2056-3639(Online)

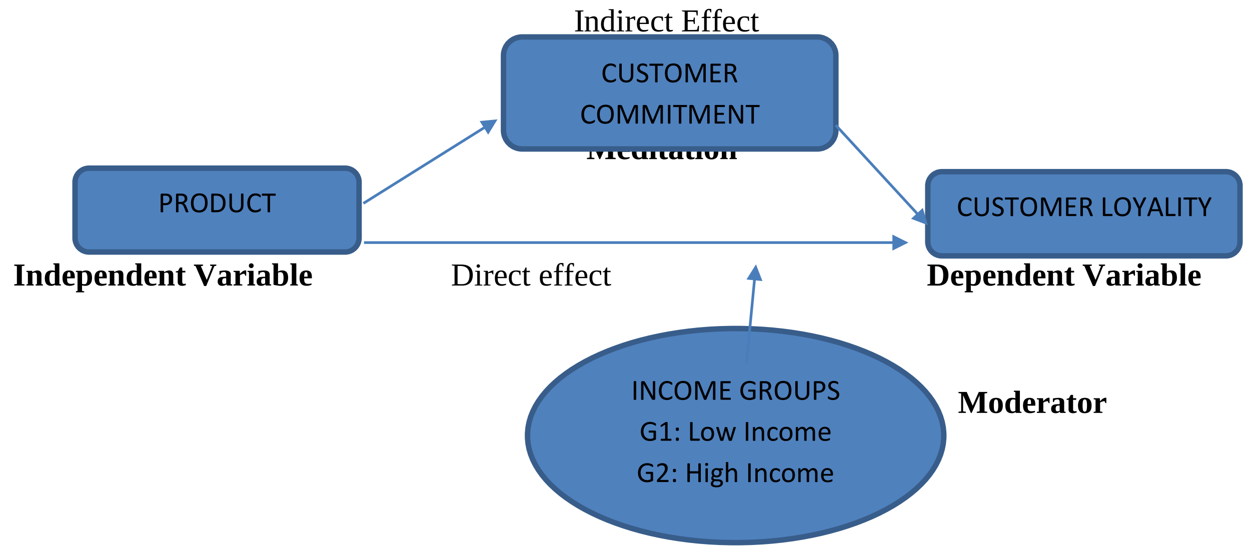

Fig. 1.1

Moderator increase or decrease the degree of relationship between independent and dependent variables**.** In fig. 1.1 the income groups works as the moderator and it impacts the relationship between Independent and Dependent Variables.

Direct and Indirect Effects: The direct and indirect affects with and without mediator present the level of mediation. The kind of mediation measured with direct and indirect effects test the hypothesis proposed. For checking indirect effects minimum sample of 500 is required. Table no.2 shows the direct and indirect effects.

Table No. 2

	Direct effect without mediator	Direct effect with mediator	Indirect effect	Output
PR>CC>CL	0.782**	0.342***	0.851***	Full mediation

(* sig. at 95%, ** sig. at 99 %, *** sig. at 99.9 %)

PR- Product, CC- Consumer commitment, CL- Consumer loyalty.

Direct Effects: In case of direct effect without mediator the $\beta = 0.782$ which is significant at 99 percent. Independent (IV) and dependent variables (DV) has significant relationship. Direct effect with mediator has $\beta = 0.342$ which is significant at 99.9 percent, it reflect that most of the relationship between independent variable and dependent variable through mediator.

Indirect Effects: The $\beta = 0.851$ for indirect effects which is significant at 99.9 percent. It means there is full mediation between independent and dependent variables.

Second Generation Structural Equation Modeling (SEM):

Multilevel Structure equation modeling and growth curve change modeling are new features added to the first generation of SEM. The collection of longitudinal data from subsets of population, measuring the change in the growth of variables over a period of time providing researchers' new idea in the field of structure equation modeling. A general methodology for handling measurement errors and multistage sampling is provided through integrating

ISSN 2056-3620(Print), ISSN 2056-3639(Online)

multilevel modeling and SEM; it explains between-group variation of the within-group variables (Kaplan, 2000).

DISCUSSION

Theory must be at the background for building any model. Variables under the investigation may relate to each other in multiple ways. Lack of theoretical support run out of track even the simplest models. So the researchers should carefully examine the relationship between the constructs. The assumption of normality in building a model is the pre-model building condition, data under investigation should be normal, although (Chen, 2014) advocate the use of sample have skewness ≤ 2 and kurtosis ≤ 7 in maximum likelihood estimation technique, for non-normal data alternative methods of estimation like OLS with large sample size requirements. Residual errors can never be negative, a negative error in the model shows problem in the data of that variable. Alternative model strategy is also suggested by researcher to check the model fit, out of these models best fit should be used for SEM. The model should be just identified or over-identified, so that it will increase the degree of freedom and estimation accuracy.

CONCLUSION

Novice researchers must take care of assumptions and concepts of Structure Equation Modeling, while building a model to check the proposed hypothesis. SEM is more or less an evolving technique in the research, which is expanding to new fields. Moreover, it is providing new insights to researchers for conducting longitudinal investigations.

REFERENCES

Biddle, B. J., Marlin, M. M., Biddle, B. J., & Marlin, M. M. (2014). General Issues Structural Equation Modeling, *58*(1), 4–17.

Chen, M. (2014). Validation of the Wood 's Job Satisfaction Questionnaire Taiwanese Nonprofit Sport Organization among Workers, *94*(3), 437–447. doi:10.1007/sl

Chin, W. W., Peterson, R. a., & Brown, S. P. (2008). Structural Equation Modeling in Marketing: Some Practical Reminders. *The Journal of Marketing Theory and Practice, 16*(4), 287–298. doi:10.2753/MTP1069-6679160402

Hair et al., (2012). Multivariate Data Analysis. 6th ed. NewJersey: Prentice-Hall Inc.

Kaplan, D. (2000). Structural Equation Modeling: Foundations and Extensions . Sage, Newbury Park, CA

K.G. Jöreskog. (1973). A general method for estimating a linear structural equation system, in Structural equation models in the social sciences, A.S. Goldberges and O.D. Duncan, eds., Seminar Press, New York, pp.85-112.

King et al., (2006). Reporting Structural Equation Modeling and Confirmatory Factor Analysis Results: A Review. *The Journal of Education Research,* 99(6), 323-337.

MacCallum, R. C., Tucker, L. R, & Briggs, N. E. (2001). An alternative perspective on parameter estimation in factor analysis and related methods. In R. Cudeck, S. du Toit, &

ISSN 2056-3620(Print), ISSN 2056-3639(Online)

D. Sorbom (Eds.), *Structural equation modeling: Present and future*. (pp. 39-57). Lincolnwood, IL: SSI

Nevitt, J., Hancock, G. R., & Hancock, G. R. (2014). Improving the Root Mean Square Error of Approximation Conditions Nonnormal for in Structural Equation Modeling, *68*(3), 251–268.

Ullman, J.B (2001). Structural equation modeling. In B. G. Tabachnick & L. S. Fidell (Eds.), using multivariate statistics (4th ed.). Needham Heights, MA: Allyn & Bacon.

ISSN 2056-3620(Print), ISSN 2056-3639(Online)