A Short Review on Structural Equation Modeling: Applications and Future Research Directions

Surajit Bag*

**Tega Industries Ltd., India. E-mail: surajit.bag@gmail.com*

Abstract

The application of multivariate techniques is mainly to expand the researchers' explanatory ability and statistical efficiency. The first generation analytical techniques share a common limitation i.e. each technique can examine only a single relationship at a time. Structural Equation Modeling, an extension of several multivariate techniques is the technique popularly used today can examine a series of dependence relationships simultaneously. The purpose of this study is to provide a short review on Structural Equation Modeling (SEM) being used in social sciences research. A comprehensive literature review of article appearing in top journals is conducted in order to identify how often SEM theory is used. Also the key SEM steps have been provided offering potential researchers with a theoretical supported systematic approach that simplify the multiple options with performing SEM.

Keywords: Multivariate Data Analysis, SEM, Path Modeling

Introduction

Contemporary research in the area of social sciences involves analysing datasets consisting of multiple variables; the body of methodology dealing with this type of datasets is called multivariate analysis. Additionally more mathematics is required to derive multivariate statistical techniques for making inference than in a univariate setting.

SEM is a family of statistical models that seek to explain the relationships among multiple variables. In this process, it examines the structure of the interrelationships expressed in a series of equations, similar to a series of multiple regression equations. These equations depict the relationships among constructs involved in the analysis. Constructs are unobservable or latent factors represented by multiple variables.

SEM is known by many names such as covariance structure analysis, latent variable analysis and path modeling. Although SEM models can be tested in many ways, all SEM models are distinguished by three characteristics:

1. Estimation of multiple and interrelated dependence relationships.
2. An ability to represent observable concepts in these relationships and account for measurement
3. Defining a model to explain the entire set of relationships.

The basic steps of SEM are 1) Model specification; 2) Model identification; 3) Data preparation and screening; 4) Estimation of the model; and 5) Model Re-specification, if necessary (Kline,2005).

Literature Review

Here researcher made an attempt to understand, what work has been carried out in the past in the direction of "*Structural Equation Modeling*". In order to understand the evolution of SEM, content analysis was performed on papers published in reputed journals.

SEM has gained popularity over time but due to its complexity, researchers often make mistakes in selecting the right program.

SEM can be classified into covariance based SEM and component based SEM. The first approach has been developed around Karl Joreskog and second approach around Herman Wold under the name Partial Least Squares. Covariance based SEM is usually used with an objective of model validation and requires a large sample. Component based SEM is mainly used for score computation and can be carried out on very small samples (Tenenhaus, 2008).

SEM is a technique used to specify, estimate, and evaluate models of linear models among a set of observed variables in terms of an often smaller number of unobserved variables. SEM may be used to build or test theory. When selecting the SEM, care should be taken to consider a theory's stage of development. Exploratory techniques are well suited for establishing and whether it explains a meaningful amount of variance in an endogenous construct. Because of the components based approach to estimating relationships, exploratory techniques such as PLS are less prone to Type I error and better suited for small, non-normal datasets often collected for initial tests of relationships. Confirmatory techniques may be used to build theory derived from well established set of constructs. Regardless of whether the SEM technique is exploratory or confirmatory, it possesses the ability to integrate measurement and structural models (Roberts, Thatcher and Grover, 2010).

Covariance based SEM is used when the sample size is large, data is normally distributed and the model is correctly specified. PLS SEM becomes a good alternative to Covariance based SEM when the sample size is small, researcher have little available theory, predictive accuracy is paramount and correct model specification cannot be ensured.

Wong (2013) presented a technical note to the step wise guide to the Smart PLS software for beginners. Different software programs are used by researchers in performing structural equation modeling. In Covariance based SEM software packages AMOS, LISREL, EQS and MPlus are commonly used.

In Partial Least Squares, which focuses on the analysis of variance can be carried out using PLS-Graph, Visual PLS, Smart PLS and WarpPLS.

The other approach is component based SEM known as Generalized Structured Component analysis which is implemented through Visual GSCA or a web based application called GeSCA.

The advantages of SEM over traditional MANOVA/ MANCOVA analyses are: 1) estimating and removing both random and correlated measurement errors; and 2) examining the mediating process (Lee, 2011).

Structural Equation Modeling has been widely applied in the area of social sciences.

Jayakumar and Sulthan (2014) used structural equation modeling to bring out the employee perception on training and development program in the industry.

Saxena and Khanna (2013) proposed a model for measuring advertising value through structural equation modeling.

Jayakumar and Sulthan (2013) used structural equation modeling to throw light on different types of stress factors, stress symptoms and their impact of stress on college students.

Lee (2011) demonstrated a comprehensive statistical analysis, SEM in the field of Educational Technology. Explored how interventions affect learning and examine the indirect effect of related psychological constructs.

Thomas and Bhasi (2011) used SEM in the area of information technology for software project risk management.

Singh et al., (2010) used SEM in the area of retail supply chain.

Zukuan et al., (2010) employed the SEM approach to understand the relationship of TQM and organisational performance.

Mohamad et al., (2011) used SEM to study empirically and test a model to examine the relationships among service recovery satisfaction and destination loyalty in the hotel industry.

Silva et al., (2010) used SEM to identify determinants of attitude towards ICT usage among rural administrators.

Biswas (2010) undertook multi-same studies to explore the various Guna constructs of Indian philosophy. The author has made a confirmatory assessment of the two constructs using a SEM.

Grace and Bollen (2008) presented a framework for discussing composites and demonstrate how the use of partially reduced form models can help to overcome some of the parameters estimation and evaluation problems associated with models containing composites.

Yap and Khong (2006) used SEM to model the relationships between critical success factors of business process reengineering implementations, customer service management and perceived business performance in Malaysian banking institutions.

Tempelar et al., (2007) investigated the relationship between attitudes and reasoning abilities by estimating a full structural equation modeling.

Chen et al., (2011) conducted empirical study using SEM on turnover intention by modeling job stress as a mediating variable.

Hackl and Westlund (2000) used SEM for customer satisfaction measurement.

Nachtigall et al., (2003) provide advantages and pitfalls of structural equation modeling so that right application can be performed by other researchers.

Bollen and Pearl (2013) have presented a technical report which presents eight myths about causality and structural equations models for better understanding of researchers.

Afthanorhan et al., (2015) intend to demonstrate a parametric approach using z test to attain the probability level with the help of SmartPls 2.0

Thomas and Bhasi (2011) and Pousette and Hanse (2002) used multigroup SEM approach to test for multigroup invariance in measurement models and structural models between job characteristics, psychosocial intervening variables, health outcomes and sickness absenteeism.

To tackle methods effect Pohl et al., (2008) presented a new approach for modeling this kind of phenomenon, consisting of a definition of method effects and a first model, the method effect model, which can be used for data analysis. This model may be applied to multi trait-multi method or to longitudinal data where the same construct is measured with at least two methods at all occasions.

The main feature of SEM is to compare the model to empirical data. This comparison leads to so called fit statistics assessing the matching of model and data. If the fit is acceptable, the assumed relationships between latent and observed variables (measurement models) as well as the assumed dependencies between the various latent variables (structural model) are regarded as being supported by the data. In some cases, only the fit of a measurement model is of interest. In other cases, parameters of the structural model may be of interest. Even though researchers use the term effect, this does not mean that a SEM is a causal model. Although under specific circumstances, SEM can represent causal relationships, a well fitting SEM does not necessarily have to contain any information on causal dependencies at all. Hence the testing the fit of a SEM is not tests of causality.

Singh (2009) has argued that in the social science literature very few studies report the correct set of model fit indices (FIs), with little justification. Effort was put to reduce some of the confusion surrounding the appropriate use of SEM model fit indices.

Six Stages in Structural Equation Modeling

Here the six stages in structural equation modeling are presented.

Defining Individual Constructs

A good measurement theory is a necessary condition to obtain useful results from SEM. Hypothesis tests involving structural relationships among constructs will be no more reliable or valid than the measurement model, in explaining how these constructs are validated. It entirely depends on how the researcher selects the items to measure each construct which sets the foundation for the entire remainder of the SEM analysis. The researcher must invest significant time and effort early in the research process to make sure the measurement quality will enable valid conclusions to be drawn.

Developing the Overall Measurement Model

In this stage, each latent construct to be included in the model is identified and the measured indicator variables are assigned to latent constructs. Although this identification and assignment can be represented by equations, it is simpler to represent the process with a diagram. There are three types of relationships: measurement relationships between indicators/items and constructs; correlation relationship among the constructs; and error terms for the items.

Design a Study to Produce Empirical Results

After the basic model specified in term of constructs and measured variables/indicators, the researcher must turn attention to issues involved with research design and estimation.

Research design includes decision making on the type of data to be analyzed, either covariances or correlations; the impact and remedies for missing data; the impact of sample size.

The researcher must be careful to specify the type of data (Metric or Non Metric) being used for each measured variable so that appropriate measure of association can be calculated. Also researcher must still choose between correlation versus covariance based on interpretive and statistical issues.

Researcher must also make several important decisions regarding the missing data.

Selecting the sample size in SEM is more critical than other multivariate techniques because some of the statistical algorithms used by SEM programs are unreliable with small sample size. Five considerations affecting the required sample size for SEM include: multivariate normality of the data; model complexity; the amount of missing data; the average error variance among the reflective indicators.

Once the model is specified, researchers must choose the estimation method, the mathematical algorithm that will be used to identify estimates for each free parameter. Several options are available such as Ordinary Least Sqaures (OLS) regression. Maximum Likehood estimation (MLE) is more efficient and unbiased when the assumption of multivariate normality is met. The potential sensitivity of MLE to non-normality however created a need for alternative estimation techniques. Methods such as weighted least squares (WLS), generalised least squares (GLS), and asymptotically distribution free (ADF) estimation became available. All of the alternative estimation techniques have become widely available. Various software programs are available today such as AMOS, EQS and LISREL to test structural models.

Assessing the Measurement Model Validity

Measurement model validity depends on establishing acceptable levels of goodness-of-fit for the measurement model and finding specific evidence of construct validity.

Multiple fit indices should be used to assess a model's goodness of fit and should include:

- The Chi square value and the associated degree of freedom
- One absolute fit index (i.e., GFI, RMSEA, or SRMR)
- One incremental fit index (i.e. CFI or TLI)
- One goodness of fit index (GFI, CFI, TLI etc.)
- One badness of fit index (RMSEA, SRMR, etc)

Thumb rules for use of model fit indices (FI):-

- Regarding the overall fit, use the FIs cut-offs for continuous data as: RMSEA<.06, TLI>.95, CFI>.95, SRMR<.08
- For categorical variables, use the above cut-off values, except SRMR; also WRMR <.90 works well for continuous and categorical data and WRMR $\leq$1.0 even for moderately non-normal continuous data.
- For non-normal continuous data when N>250, the SB based CFI cut-off value is 0.95 and SRMR at 0.07 (acceptable Type I and Type II error). When N$\geq$500, the TLI_{ML} and CFI_{ML} at the suggested values were acceptable within normal data.
- The power of TLI, CFI and RMSEA to detect models with mis-specified loadings is higher than their power to detect models with mis-specified covariances. YaunBentler statistic should be used when N is in the range 60-120.
- Q plot should be discussed, as the standardized residuals that depart excessively from the Q-plot line indicate that a mis-specified model.
- Only CN, NNFI and RMSEA are not significantly related to study characteristics. NNFI is the most suitable index as it was not significantly related to study characteristics such as sample size, number of indicators per latent variable, number of latent variables, number of estimated paths, and degrees of freedom.

Specifying the Structural Model

This is a critical step in developing a SEM model. This step involves specifying the structural model by assigning relationships from one construct to another based on the proposed theoretical model. Structural model specification focuses on adding single headed, directional arrows to represent structural hypothesis in that researcher's model.

Assessing the Structural Model Validity

This is the final stage and involves efforts to test the validity of the structural model and its corresponding hypothesized theoretical relationships.

Conclusions

SEM is a confirmatory rather than exploratory approach to test the relationships. SEM accounts for measurement errors in the course of model testing. It can incorporate observed variables as well as latent variables. It tests a priori relationships rather than allowing the technique or data to define the nature of relationship between variables.

This paper has presented a brief tutorial overview of SEM procedure. While in this brief note it was not possible to provide the detail required for a complete understanding of SEM technique, it is hoped that reader will pursue references noted here in developing theories and testing models using SEM.

Apart from reading the papers listed in the references beginners can also learn SEM from the following sources:-

- "Introduction to Structural Equation Modeling taught by Randall Schumacker" http://www.statistics.com/sem
- http://davidakenny.net/cm/causalm.htm
- Introduction to Structural Equation Modelling using Mplushttp://www.utrechtsummerschool.nl/courses/social-sciences/introduction-to-structural-equation-modelling-using-mplus
- Ralph O. Mueller & Gregory R. Hancock (2008), "Best Practices in Structural Equation Modeling", Best practices in quantitative methods, Chapter 32,DOI: http://dx.doi.org/10.4135/9781412995627, Print ISBN: 9781412940658 | Online ISBN: 9781412995627

Future Research Directions

Future research directions should put efforts on hybrid modeling approach such as moderated-mediation and mediated-moderation in designing SEM research for empirical studies.

References

Afthanorhan, A., Nazim, A., & Ahmad, S., (2015). A parametric approach using Z-test for comparing 2 means to multi-group analysis in partial least square structural equation modeling. *British Journal of Applied Science & Technology*, 6(2), 194-201

Biswas. M. (2010). Personality and organization citizenship behavior: an Indian argument An application of Structural Equation Modelling using PLS Algorithm, Vilakshan. *XIMB Journal of Management*, March issue, 77-102

Biswas, M. (2010). In search of personality inventory for Indian managers: An application of structural Equation Modelling. *Journal of Services Research*, 10(1), 101-123

Bollen, K. A., & Paearl, J. (2013). Eight myths about causality and structural equations models, In SL. Morgan (Ed.), *Handbook of Causal Analysis for Social Research*, Chapter 15, 301-328, Springer

Chen, M. F., Lin, C. P., & Lien, G. Y. (2011). Modelling job stress as a mediating role in predicting turnover intention. *The Services Industries Journal*, 31(8), 1327-1345

Grace, J. B., & Bollen, K. A. (2008). Representing general theoretical concepts in structural equation models: The role of composite variables. *Environmental and Ecological Statistic*, 15, 191-213

Hackl, P., & Westlund, A. H. (2000), On structural equation modeling for customer satisfaction measurement. *Total Quality Management*, 11(4/5&6), 820-825

Jayakumar, G. S., & Sulthan, A. (2013). Stress symptoms: Structural equation modeling. *SCMS Indian Journal of Management*, (July-September), 95-109

Jayakumar, G. S., & Sulthan, A. (2014). Modelling : Employee perception on training and development. *SCMS Indian Journal of Management*, (April-June), 57-70

Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). New York: Guilford Press

Lee, H.W. (2011). *An application of latent variable structural equation modeling for experimental research in educational technology*, TOJET, 10(1), 15-23

Mohamad, M., Mohamad, M., Mamat, I., & Mamat, M. (2014). Modelling positive development, life satisfaction and problem behavior among youths in Malaysia. *World Applied Sciences Journal*, 32(2), 231-284

Mohamad, M., Abdullah, A. R., & Mokhlis, S. (2011). Examining the influence of service recovery satisfaction on destination loyalty: A structural Equation Modelling. *Journal of Sustainable Development*, 4(6), 3-11

Nachtigall, C., Kroehne, U., Funke, F., & Steyer, R. (2003). Why should we use SEM? Pros and Cons of Structural Equation Modelling. *Methods of Psychological Research Online*, 8(2), 1-22

Ogasawara, H. (2008). Some properties of pivotal statistics based on the asymptotically distribution free theory in Structural Equation Modeling. *Communication in Statistics-Simulation and Computation*, 37, 1931-1947

Pohl, S., Steyer, R., & Kraus, K. (2008). Modelling methods effects as individual causal effects. *Journal of Royal Statistical Society*, 171(Part 1), 41-63

Pousette, A., & Hanse, J. J. (2002). Job characteristics as predictors of ill-health and sickness absenteeism in different occupational types- a multigroup structural modeling approach. *Work & Stress*, 16(3), 229-250

Roberts, N., Thatcher, J. B., & Grover, V. (2010). Advancing operations management theory using exploratory structural equation modeling techniques. International *Journal of Production Research*, 15(1), 4329-4353

Saxena, A., & Khanna, U. (2013). Advertising on social network sites: A structural equation modeling approach. *Vision*, 17(1), 17-25

Silva, J. L., Samah, B. A., Shaffril, H. A. M., Hassan, M. A., & Badsar, M. (2010). Determinants of attitudes towards information and communication technology usage among rural administrators in using structural equation modeling. *American Journal of Applied Sciences*, 8(5), 481-485

Singh, R., Sandhu, H. S., Metri, B. A., & Kaur, R. (2010). Relating organized retail supply chain management practices, competitive advantage and organizational performance. *VISION-The Journal of Business Perspective*, 14(3), 173-190

Singh, R. (2009). Does my structural equation model represent the real phenomenon?: a review of the appropriate use of structural equation modeling (SEM) model fit indices. *The Marketing Review*, 9(3), 199-212

Tempelaar, D. K., Loeff, S. C. V. D., & Gijselaers, W. H. (2007). A structural equation model analyzing the relationship of students' attitudes toward statistics, prior reasoning abilities and course performance. *Statistics Education Research Journal*, 6(2), 78-102

Tenenhaus, M. (2008). Component based structural equation modeling. *Total Quality Management*, 19(7-8), 871-886

Thomas, S., & Bhasi, M. (2011). A software model for project risk management Vilakshan. *XIMB Journal of Management*, (March issue), 71-84

Wong, K. K. K. (2013). Partial least squares structural equation modeling techniques using Smart PLS. *Marketing Bulletin*, 24, 1-32. Retrieved from http://marketing-bulletin.massey.ac.nz

Yap, B. W., & Khong, K.W. (2006). Examining the effects of customer service management on perceived business performance via structural equation modeling. *Applied Stochastic models in Business and Industry*, 22, 587-605, DOI: 10.1002/smb.648

Zakuan, N. M., Yusof, S. M., Laosirihongthong, T., & Shaharoun, A. M. (2010). Proposed relationship of TQM and organizational performance using structured equation modeling. *Total Quality Management*, 21(2), 185-203