Review Article

Perspectives of Structural Equation Modeling (SEM)

Dr. S. Prabu Shankar
Assistant Professor
Department of Education
Institute of Advanced Study in Education (Autonomous)
Saidapet, Chennai - 600 015.

ABSTRACT

Structural Equation Modeling (SEM) is basically a multivariate statistical analysis, profoundly linear and cross-sectional modeling which includes focussed dimensions of other analysis and methods such as canonical correlation, principal component analysis, factor analysis, path analysis, multiple regression, discriminant function analysis, growth analysis, trend analysis, bi-variate and inter-correlations as special statuses. Structural equation modeling (SEM) relies on sequence of other statistical methods that infer composite relationships between independent variables and dependent variables.

Structural Equation Modeling is often referred as simultaneous equation modeling, analysis of covariance structures or casual modeling or casual analysis or path analysis or confirmatory factor analysis or analysis of moments, because of its varied applications at abstract instances where multivariate techniques may be applied. Educational research, significant as any other research area with several complex variables, dimensions and structures, proportions and magnitudes, more importantly involving social samples that make studies still abstract in all aspects require numerous analyses to arrive at an apparent logical supposition (Maccallum, 1989). Structural Equation Modeling (SEM) will create possibilities to find out causal relationships by systematic experimentations, assumptions and suppositions by means of multivariate statistical analysis hence contributing in numerous ways to derive unbiased estimates for the relations between latent constructs between the independent and dependent variables.

Keywords: Structural Equation Modelling (SEM), Latent variables, Measurement model, Structural model, Causal modelling

INTRODUCTION

Structural equation modeling is widely used to analyze structural relationships between the variables. This model takes in to consideration the combination of multiple statistical methods and analysis to analyze the structural relationship between measured variables and latent constructs. Structural equation modeling is chosen to estimate the multiple and interrelated dependence in a single analysis. In this analysis, two types of variables are used endogenous variables and exogenous variables. Endogenous variables are equivalent to dependent variables and are equal to the independent variable.

Structural Equation Modeling is primarily concerned with 'Linear Structural Relations' (LISREL) between the variables. 'Structural relation' is connected with the handling of relationships between latent variables usually formulated by linear regression equations, graphically expressed by so-called path diagrams using or multiple linear regression, but with a system of regression equations. SEM's important characteristic is the capability to deal with latent variables, i.e. non- observable measures like true-score variables or factors underlying observed variables. Latent variables are variables that are not directly observed but are rather inferred (through a mathematical model) from other variables that are observed (directly measured).Latent variables are connected to observable variables by a measurement model (Schumacher & Lomax, 2010). SEMs consist of a structural model representing the relationship between the latent variables of interest, and measurement models representing the relationship between the latent variables and their manifest or observable indicators.

Perspectives of Structural Equation Modeling

Structural Equation Modeling aims at assimilating a set of relationships by providing consistency and comprehensive description singularities present in the variables. Two independent models namely a) measurement model and b) structural model that provides insight on how the study is influenced by the variable based on the constructs. i.e., the measurement model represents the theory that specifies how measured variables come together to represent the theory and structural model represents the theory that shows how constructs are related to other constructs. SEM is often referred as casual modeling as it assesses the casual relationships between the variables or constructs based on certain important assumptions. These assumptions vary in their characteristics, features and the variation in the statistical treatments with regard to finding out the casual relationships (Bollen, 1989). The following are the conditions under which SEM may be applied in the research process in order to arrive at casual relationships.

- Multivariate normal distribution, which is generalization of the one-dimensional (univariate) normal distribution to higher dimensions. The maximum likelihood method is used and assumed for multivariate normal distribution. Small changes in multivariate normality can lead to a large difference in the chi-square test.
- Linearity, a statistical relationship or function that can be graphically represented as a straight line, as in two quantities that are directly proportional to each other. A linear relationship is assumed between endogenous and exogenous variables.
- Sequence, a set of related events, movements, or items that follow each other in a particular order. There should be a cause and effect relationship between endogenous and exogenous variables and a cause has to occur before the event.
- Model identification, includes the optimal design of experiments for efficiently generating informative data for fitting such models as well as model reduction. Comparisons must be greater than the estimated parameters or models should be over identified or exact identified. Under identified models are not considered.
- Sample size, determination is the act of choosing the number of observations or replicates to include in a statistical sample. The sample size is an important feature of any empirical study in which the goal is to make inferences about a population from a sample. Most of the researchers prefer a 200 to 400 sample size with 10 to 15 indicators. As a rule of thumb, that is 10 to 20 times as many cases as variables.

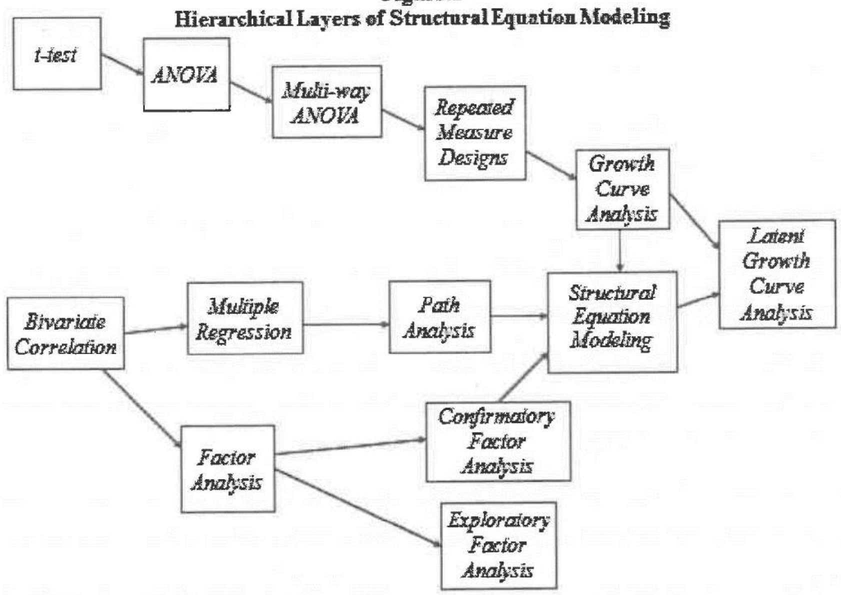
Scope of Structural Equation Modeling (SEM)

SEM can abstractly be applied to any research method involving one or more independent variables or one or more dependent variables. The chief goal of Structural Equation Modeling is to determine and rationality of a causal process or a construct or a model. This function of Structural Equation Modeling proves that SEM is a confirmatory technique. The feature of selecting sample from a population based on a sampling technique is similar in SEM like in other models or research process. The condition is that there should be a covariance matrix to serve as dataset, which is based on the sample of collected measurements of the variables.

The possibility of applying SEM in social science research methods depends on whether the proposed study involves a population covariance matrix that is consistent with the sample covariance matrix. Kaplan (2000) observes that in this case one must specify a priori a model that will undergo validation testing; there are many questions SEM can answer.

The purpose of SEM is to examine a set of relationships between one or more independent variables and one or more dependent variables. The primary condition with regard to the variable for applying SEM is that both independent variables and dependent variables should either be continuous or discrete. Independent variables are usually considered either predictor or causal variables because they predict or cause the dependent variables (the response or outcome variables). SEM endorses whether the statistical model selected for the study is adequate or not. Parameters are estimated and compared with the sample covariance matrix. Structural equation modeling is referred as 'causal modeling' or 'analysis of covariance structures' as it involves multiple statistical methods and analysis may be applied to find out casual relationships. Path analysis and confirmatory factor analysis (CFA) are special types of SEM.

Figure:1
Hierarchical Layers of Structural Equation Modeling



Goodness of fit statistics can be calculated that ascertains whether the selected model is appropriate or needs further revision. SEM can also be used to compare multiple schemes that are specified a priori. SEM can determine the amount of inconsistency in the dependent variables which is accountable in determining the independent variables. SEM establishes the reliability of each measured variables, irrespective of whether it is a dependent or independent variable and further it mediates and moderates the covariance to achieve indirect effects. SEM regulates group differences in order to fit separate structural equation models for different groups and compare results. Possibility to include both random and fixed effects in the models and thus include hierarchical modeling techniques is possible in the analysis.

Precincts of Structural Equation Modeling

Pearl (2000) states that "modern SEM is far from the original causality modeling theme, mainly for the following two reasons, researchers have tried to build scientific 'credibility' of SEM by isolating (or removing) references to causality and causal relationships do not have commonly accepted mathematical notation". But SEM proves to be a confirmatory technique irrespective of the many limitations of application in finding the casual modeling techniques namely· path analysis, factor analysis, multiple regression, covariance matrix etc.,The major limitations of applying SEM in determining causal relationship between independent variables and dependent variables is that the researcher must be very careful with the study· design when using SEM for exploratory work.

a. Ability to model constructs as latent variables
b. SEM is a confirmatory approach. Need to have established theory about the relationships. It cannot be used to explore possible relationships when there are more variables. If there is not enough theoretical background the model will not deliver goodness fit.
c. Ability to model constructs as latent variables. SEM compares the performance of a model across multiple populations.
d. SEM applications needs to be have a large sample size to get stable estimates of the covariances/ correlations.
e. SEM is often thought of strictly correlational but can be used with experimental data.

f. Causal modeling' referring to SEM is distorted as there is nothing causal, in the sense of inferring causality, about the use of SEM.
g. SEM's ability to analyze more complex relationships produces more complex models: Statistical language has turned into jargon due to vast supply of analytic softwares (LISREL, EQS, AMOS) (Mueller, 1996).
h. Analysis of research reports methodologically based on SEM, usually a LISREL model, one notices that they lack review which is a prerequisite to parameter estimation.
i. Specification of a full model a priori and test that model based on the sample and variables included in your measurements. Also it is vital to know the number of parameters needs to be estimated - including co-variances, path coefficients, and variances. It is essential to know all relationships that need to be specified in the model.
j. Works with multiple, related equations simultaneously. Allows reciprocal relationships.
k. SEM will auto-estimate missing data and run the model simultaneously.

SEM has the ability to assess complex relationships between multivariate data, sample size is an important issue. Important assumptions with regard to sample selection for a structural equation modeling are that you need more than 200 observations, or at least 50 more than 8 times the number of variables in the model. A larger sample size is always desired for SEM (Maccallum, 1989).

As applied in the multivariate statistical methodologies, most of the estimation techniques used in SEM require multivariate normality. The data need to be examined for uni-variate and multivariate outliers. Transformations on the variables can be made. However, there are some estimation methods that do not require normality.SEM techniques only look at first-order (linear) relationships between variables. Linear relationships can be explored by creating bivariate scatterplots for all the variables (Schumacker & Lomax, 2010).

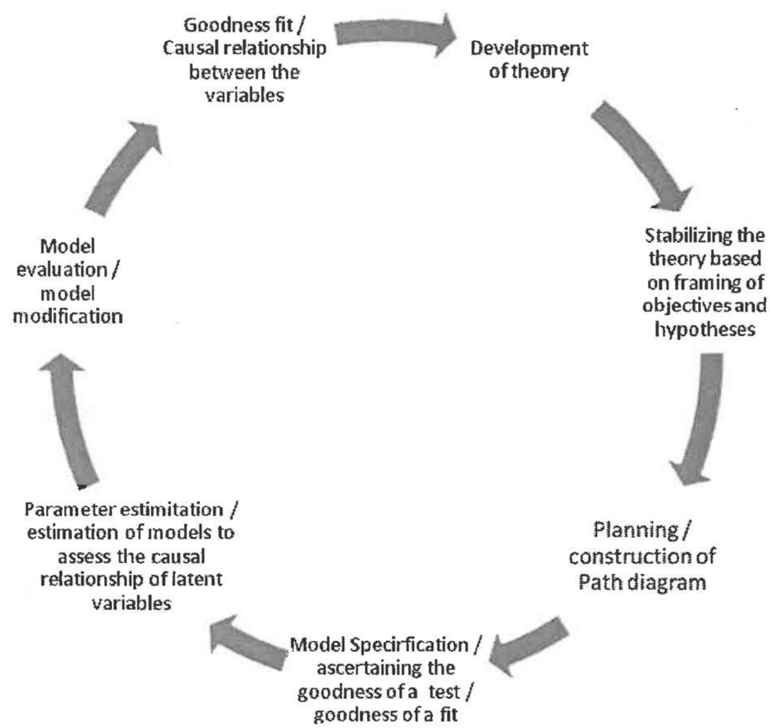
The residuals of the co-variances (not residual scores) need to be small and centered about zero. Multi-collinearity among the independent variables for manifest variables can be an issue. Most models other than structural equation modelling will inspect the determinant of a section of your covariance matrix, or the whole covariance matrix. A very small determinant may be indicative of extreme multi-collinearity(Schumacker & Lomax, 2010).

Inter-disciplinary applications of SEM

- SEM may extensively be applied in social and behavioural sciences, education, management, psychology, economics, chemistry, physics and biology.
- In cognitive sciences static and dynamic longitudinal structural analyses of cognitive changes in old age.
- In neurosciences unified structural equation modeling approach for the analysis of multi subject, multi variate functional Magnetic resonance imaging (MRI) data.
- In health and medicine applications of SEM to health outcomes research and SEM of inflammation and metabolic dysfunction in children.
- In business and commerce to evaluate the intention of logistics to use services in shipping.
- In social psychology use of a risk assessment instrument in child productive services, SEM extends its applications in developing social influences and exposure to media on adolescent deviations.
- Structural Equation Modeling like path analysis looks at the relationships among latent variables. It is useful because it accounts for the unreliability of measurement so it offers more un-biased parameters. Also lets you test virtually any theory or research.
- In theory testing it assesses the strength of prediction or association in models with multiple dependent variables and further extends its application in model fit.
- Mediation or tests of indirect effects in the experimental research process.
- SEM applications in ascertaining group differences, multiple-sample analysis, longitudinal models, multi-level nested models.

Process of Structural Equation Modeling

Figure 2: SEM Process



Advantages of SEM

- The model assesses the goodness of a fit, finds the causal relationship between the variables, finds the structural relationship between the variables.
- SEM is a graphical modeling interface, further SEM serves as a graphical model builder that is useful for creating the interrelated models.
- SEM tests models with multiple dependencies; path analysis, multiple regression, factor analysis, ANOVA are used to describe the directed dependencies among a set of variables.
- (SEM) is a methodology for representing, estimating the covariance/correlation modeling, however, relies on several statistical tests to determine the causality usually caused by linear dependency among observed variables.

- SEM tests coefficient across multiple relationships between groups; As a consequence, the multiple regression model by itself cannot be tested. To examine this further, considering two SEM models to determine the dependency between measures and to test a hypothesized hierarchical relation between the variables will serve in finding out the causality among the variables.
- SEM handles complex data structures / variable constructs; handles time series designs with auto-correlated error, non-normal data, incomplete data.
- SEM involves in finding out the structured modeling variances between indirect variables.

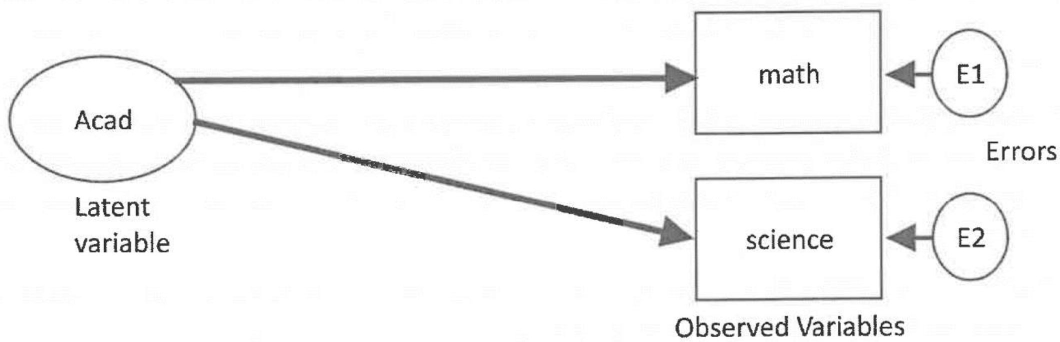
Types of diagrams symbols used in SEM

- **Rectangles : Observed variables**

[(Endogenous and Exogenous variables) (Independent variable that affects a model without being affected by it, and whose qualitative characteristics and method of generation are not specified by the model builder. An exogenous variable is used for setting arbitrary external conditions, and not in achieving a more realistic model behavior. An endogenous variable is a dependent variable generated within a model and, therefore, a variable whose value is changed (determined) by one of the functional relationships in that model)].

- **Circles : Error terms** (As a result of this incomplete relationship, the error term is the amount at which the equation may differ during empirical analysis. The error term is also known as the "residual", or the "remainder" term.)
- **Ovals : Latent variables** (latent variables are variables that are not directly observed but are rather inferred from other variables that are observed.
- **Single headed arrow** : represents prediction/ regression coefficient/ factor loading.
- **Double headed arrow** : represents correlation.

Figure 3: Example of SEM diagram



Summary

SEM is widely useful when there is a need to deal with latent (unobserved) constructs where there is a strong theoretical background to the data which is referred to as a priori hypothesis ascertaining complex relationships with a large sample. SEM are systems of linear equations that describe a network of relations among variables.

Further SEMs are implied systems of non-linear equations that describe patterns of variances and covariances among variables. SEM has a wide range of applications and it analyses the following vital research hypotheses; it questions about causal process, it questions of measurements and it questions about causal process when variables are not well measured. Further SEM implies to render solutions where proposed causal explanations are made explicit, tests of fit allow implausible models to be rejected and competing models can often be compared, and one may emerge as more plausible given the data.

Structural equation modeling as an enterprise or by itself is a phenomena which is very difficult to characterize. It is a model of causality, covariance, correlation etc. The major components of SEM includes canonical correlation, principal component analysis, factor analysis, path analysis, multiple regression, discriminant function analysis, growth analysis, trend analysis, bi-variate and inter-correlations thus extending its applications to a wide range of analysis that precisely assess and interpret data.

Glossary of terms:

Analysis of Variance (ANOVA) is a statistical method used to test differences between two or more means. It may seem odd that the technique is called "Analysis of Variance" rather than "Analysis of Means."

Analysis of covariance (ANCOVA) is a general linear model which blends ANOVA and regression.

Behavioural sciences: is the systematic analysis and investigation of human and animal behaviour through controlled and naturalistic observation and disciplined scientific experimentation. It attempts to accomplish legitimate, objective conclusions through rigorous formulations and observation.

Causal modelling: A causal model is an abstract model that describes the causal mechanisms of a system. The model must express more than correlation because correlation does not imply causation.

Confirmatory factor analysis (CFA): Confirmatory factor analysis (CFA) is a statistical technique used to verify the factor structure of a set of observed variables. CFA allows the researcher to test the hypothesis that a relationship between observed variables and their underlying latent constructs exists.

Construct: In a scientific theory, particularly within psychology, a hypothetical construct is an explanatory variable which is not directly observable.

Covariance matrix (also known as dispersion matrix or variance-covariance matrix) is a matrix whose element in the i^{th} j^{th} position is the covariance between the /hand t elements of a random vector. A random vector is a random variable with multiple dimensions.

Dependent variable: a variable (often denoted by y) whose value depends on that of another. A dependent variable is what you measure in the experiment and what is affected during the experiment. The dependent variable responds to the independent variable. It is called dependent because it "depends" on the independent variable.

Endogenous and Exogenous variables: As with endogenous variables, the status of the variable is relative to the specification of a particular model and causal relations among the independent variables. An exogenous variable is by definition one whose value is wholly causally independent from other variables in the system.

Goodness of fit: The goodness of fit of a statistical model describes how well it fits a set of observations. Measures of goodness of fit typically summarize the discrepancy between observed values and the values expected under the model in question.

Independent variable: a variable (often denoted by x) whose variation does not depend on that of another. An independent variable is the variable that is changed or controlled in a scientific experiment. Independent variables are the variables that the experimenter changes to test their dependent variable. The effect on the dependent variable is measured and recorded. An independent variable is a variable that is manipulated to determine the value of a dependent variable s. The dependent variable is what is being measured in an experiment or evaluated in a mathematical equation and the independent variables are the inputs to that measurement.

Latent growth modelling is a statistical technique used in the structural equation modeling (SEM} framework to estimate growth trajectory. It is a longitudinal analysis technique to estimate growth over a period of time. It is also called latent growth curve analysis. The latent growth model was derivedfrom theories of SEM.

Linear Structural Relations: LISREL, an acronym for linear structural relations, is a statistical software package used in structural equation modeling (SEM)for manifest and latent variables.

Latent variables: are variables that are not directly observed but are rather inferred (through a mathematical model)from other variables that are observed (directly measured).

Multiple regression: is an extension of simple linear regression. It is used when we want to predict the value of a variable based on the value of two or more other variables. The variable we want to predict is called the dependent variable (or sometimes, the outcome, target or criterion variable).

Path analysis: Path analysis is a straightforward extension of multiple regression. Its aim is to provide estimates of the magnitude and significance of hypothesised causal connections between sets of variables. This is best explained by considering a path diagram.

Priori: deductive, relating to or derived by reasoning from self-evident propositions - compare a posteriori, presupposed by experience.

Univariate and Multivariate Outliers: A univariate outlier is a data point that consists of an extreme value on one variable. A multivariate outlier is a combination of unusual scores on at least two variables. Both types of outliers can influence the outcome of statistical analyses. Outliers exist for four reasons. Incorrect data entry can cause data to contain extreme cases. A second reason for outliers can be failure to indicate codes for missing values in a dataset. Another possibility is that the case did not come from the intended sample. And finally, the distribution of the sample for specific variables may have a more extreme distribution than normal.

References:

- ♣ Bollen, K. (1989). Structural Equations with Latent Variables. New York: John Wiley & Sons.
- ♣ Browne, M. W:; Cudeck, R. (1993).Alternative ways of assessing model fit. California: Sage Publishers.
- ♣ Hox, J.J. (1995). Amos, Eqs and Lisrelfor Windows: A comparative review. Structural Equation Modeling. 2, 79-91.
- ♣ Kaplan, D. (2000}. Structural Equation Modeling: Foundations and Extensions.SAGE,
- ♣ Advanced Quantitative Techniques in the Social Sciences series, vol. 10.
- ♣ Kaplan, David.(2007). Structural Equation Modeling.Sage. pp. 1089-1093.
- ♣ Ke-Hai Yuan, Wai Chan, George A. Marcoulides& Peter M. Bentler. (2015). Assessing Structural Equation Models by Equivalence Testing With Adjusted Fit Indexes. Routledge.319-330.
- ♣ Kline, Rex (2011). Principles and Practice of Structural Equation Modelling. (Third ed.). Guilford.
- ♣ Long, J. (1983).Confirmatory Factor Analysis. California: Sage.
- ♣ Maccallum, R.C. (1989}. Specification searches in covariance structure modelling. Psychological Bulletin, 100, 533-541.
- ♣ Mueller, R.(1996). Basic Principles of Structural Equation Modelling. New York: Springer.
- ♣ Prabu Shankar, S. (2014).Applications of Value Added Modelling in Educational Research.
- ♣ Edutracks,13, 7-9.
- ♣ Pearl, Judea. (2000). Causality: Models, Reasoning, and Inference. London: Cambridge University Press.
- ♣ Schumacker, E.Randall& Lomax, G. (2010).A Beginner's Guide to Structural Equation. London: Taylor and Francis Group, Routledge Publishers.
- ♣ Sen, A. &Srivastava, M. (2011).Regression Analysis - Theory, Methods, and Applications.4th ed. Springer: Berlin.
- ♣ SAS/STAT. User's Guide. Version 8.

- ♣ Tabachnick, B ., &Fidell, L. (1996). Using Multivariate Statistics. New York: Harper Collins.
- ♣ Westland, J. Christopher (2015).Structural Equation Modelling: From Paths to Networks. New York : Springer.
- ♣ www2.sas.com/proceedings/sugi31/200-31.pdf
- ♣ www-personal.umd.umich.edu/~delittle/.../Endogenous%20variable.htm