Structural equation modeling VS multiple regression

The first and second generation of multivariate techniques

AmirAlavifar, Mehdi Karimimalayer

Dept of mechanical and manufacturing

Engineering faculty of university Putra Malaysia

Serdang, Malaysia

Prof.Mohd Khairol Anuar

Dept of mechanical and manufacturing

Engineering faculty of university Putra Malaysia

Serdang, Malaysia

Abstract- Structural Equation Modeling (SEM) or path analysis is very powerful multivariate technique that is specialized versions of other analysis methods and enables researchers in measurement of direct and indirect effects and performing test models with multiple dependent variables and also using of several regression equations simultaneously.

SEM also enables statistical analysts to handle difficult data including; time series with auto-correlated error, non-normal data and even incomplete data.

In addition since the power in statistical theory is defined as the likelihood of rejecting the null hypothesis give that null hypothesis is false, in the context of structural equation modeling, the null hypothesis is defined by the specification of fixed and free elements in relevant parameter matrices of the model equations.

This paper depicts that how SEM works and what has distinct SEM from other statistical tools and techniques in testing and estimating causal relations by using a combination of statistical data and qualitative causal assumptions.

Key works: structural equation modeling, multivariate regression

I- INTRODUCTION

Sewall Wright (1921) defines Structural equation modeling (SEM) as a statistical method in order to test and estimate casual relationships by using casual assumptions and statistical data(Bartholomew, 1999).After Sewall Wright the economist TrygveHaavelmo(1943) and the cognitive scientist Herbert Simon (1953) have stated similar definition of SEM in their own researches. (Bartholomew, 1999)

The both type of confirmatory and exploratory modeling can be used by SEM. In other word SEM is suitable for both theory extension and theory testing. In confirmatory modeling which

Begins with a hypothesis as a causal model representative, the model concepts must be operationalized in order to permit testing of the relations of concepts in the model. Then the model will be tested against gathered measurement data to specify does the model fit the data or not? (Bartholomew, 1999)

II- ADVANTAGES OF SEM OVER REGRESSION

The most important and crucial part of the research study is a proper selection of methodology(Davis, 1996)(Stevens, 2002). The second generation of multivariate technique that was applied for validity and reliability assessment of model measures was called SEM. The first generation methods such as multiple regressions were suitable for assessing constructs and relations between constructs. The first purpose of regression analysis is prediction while the intent of a correlation is to evaluate the relationship between the dependent and independent variables.(Tabachnick, 2001)

One of the best variance predictor in an interval dependent variable is multiple regressions which is an approach to determine the model of relationship between dependent variable as Y and independent variables as X.If the explanatory variable is one variable it's called simple regression and for more explanatory variables it would be multiple regressions.

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

In this linear equation $\beta 0$ and $\beta 1$ are parameter estimates, when the independent variable, X, changes by one unit the value $\beta 1$ shows the amount of dependent variable, Y, changes while the other independent variables remain fixed.

Multiple regressionspredict the same standard errors and coefficients as gathered using OLS (ordinary least Squares) regressions.

The assumptions for multiple regression includes suitable specification of the model, linear relationships, near interval or interval data with limited range and the same level or relation through the range of independent variables.

In current research meeting such a strict assumption of multivariate regression are not possible or practical, however the main reason of choosing second generation of multivariate method in current researches is the ability of evaluating model construct relationships simultaneously.

In statistics, linear regression is an approach to modeling the relationship between a scalar variable y and one or more explanatory variables denoted X. The case of one explanatory variable is called simple regression. More than one explanatory variable is multiple regressions. Evaluation must be done in sequential steps which are carried out by some software such as LISREL, AMOS, DEPATH, EQS and RAMONA.

Therefore analysis of all the variables in the same time would be possible by second generation multivariate techniques in compared with the first generations which performed analysis separately. In addition measurement error is not accumulated in a residual error term (Fornell, 1984).

Structural equation modeling has been used for a various research Issues. There are a lot of applications embedded in SEM including:

- Causal modeling or path analysis
- Confirmatory factor analysis (CFA)
- Second order factor analysis.
- Covariance structure models.
- Correlation structure models (Jöreskog, 1989).

III- FUNDAMENTAL THEORY OF STRUCTURAL MODELING

One of the main concepts in applied statistics in medium level is the effect of Additive and Multiplicative transitions in a series of numbers, that is to say that if any numbers in list multiplied by a constant amount of k then the average of numbers would be multiplied by the same k. hence standard deviation would be multiplied by absolute value of k. the point is that a series of numbers x were related to another series of numbers y with the equation y=4x then the variance of y would be 61 times more than variance of x therefore, with the comparison between variances of x and y the equation y=4x can be tested indirectly. This theory can be generalized to some correlated variables with a number of linear equations in different approaches. However its rules might become more complicated and the calculations become more difficult but the general message remains unchanged. Which means by study of variance and covariance of variables the theory of "variables are correlated with a series of linear relations" can be examined. (Pearl, 2000) (Westland, 2010)

Development of models of latent variables identifies the convergence of independent research methods in econometrics, psychometrics, biology and many of other known methods which collects them in vast framework. Concepts of latent variables (versus observed variables) and error in variables has a long time background. In econometrics simultaneous oriented effects of some variables on another variable have studied with label of equation models. In psychometrics it measured as factor analysis and developed credit theory and it is also the principle of many measurement researches in psychometric. Regarding to biology usually follows by a similar pattern with simultaneous equation models for path analysis. (Tabachnick & Fidell, 2001)

IV- LISREL APPLICATIONS:

LISREL approach whilst calculates unknown coefficients of structural linear equations also designed to embedding models include latent variables, measurement errors in every dependent and independent variables, feedbacks, and Codependency and simultaneity.

However this method can be applied in some circumstances for confirmatory factor analysis methods, multivariable regression analysis, path analysis, and time-based data in economical models, reversible and irreversible models for cross-sectional data, structural covariance models and multi-sample analysis. (Princeton, 2010)

A- LISREL software

LISREL is a software product distributed by scientific software international (www.ssicentral.com). This software can assess and construe factor loadings, variances, and latent variable errors with measured correlation and covariance. This pattern can be used for exploratory factor analysis (EFA), second order factor analysis, Confirmatory factor analysis (CFA), and also path analysis. (Princeton, 2010)

B- Exploratory and Confirmatory factor analysis

Factor analysis can be eitherConfirmatory or Exploratory. Based on the objective of data analysis, each of these approaches can be implemented. In Exploratory Factor Analysis the researcher is looking for experimental data to discover and identity indicators and terms without any imposed model. On the other hand, exploratory analysis not only has proposal or under covering value but also it can be made structures, models or hypothesizes. (Diana D.Suhr, 2003)

Exploratory analysis can be implemented while there was no predecessor evidential matter for hypothesis formation. In fact it would used to analyze variable covariance. Hence, exploratory analysis is a method to edit and produce a theory and not an approach to test a theory.

Exploratory factor analysis mostly applied to measure latent resources of variance and covariance of observed measurements. Researchers believe that exploratory factor analysis can be useful in primitive stages of experiments. Besides, the more knowledge about the nature of social or psychic is, then the less it can be used as a helpful technique or even it can be deterrent.

From the other point of view, most of the studies can be kind of exploratory or confirmatory because they include some known and unknown variables .known variables has to be selected with high accuracy to increase information of unknown variables. It is more pleasant that edited hypothesis by exploratory analysis method, can be confirmed or rejected by exposing to more accurate statistical method. Exploratory analysis needs samples with extremely high volume.

In confirmatory factor analysis, researcher is looking to provide a model which interprets and describe experimental data with less parameter. This model is based on pre-experimental information about data structures. It could be formed as:

C- A theory or hypothesis

A classified plan for items or some tests adapted with form, shape and contents

Significant distinguish between exploratory and confirmatory is exploratory method is the most advantageous method to indicate common variances of a correlation matrix while confirmatory methods (Hypothesis Testing) determine whether data with factorial structural are coordinated or not.(Hoyle, 1995) (Hu L. & Bentler, 1999)

D- Fitting tests:

Although different types of tests which named generally as fitting indexes are in a progressive and evolutionary stage, but there is no public agreement about even one test. Thus, different papers are presenting different indexes and even various versions of famous programs such as SEM like LERSIL, EQS Amos, also use lot of fitting indexes. These indexes are categorized with different methods which absolute, relative and adjusted are dominant categories. Some of different indexes are presented below:

E- GFI and AGFI index:

Goodness-Of-Fit indicates, evaluating relative amount of variances or covariance commonly with the model. If the range of GFI is between zeros to one, its amount should be more than 0.09.

Adjusted Goodness of Fit Index (AGFI) is applicable for degree of freedom. This index is equal with application of Mean Squares instead of sum of squares in numerator and the denominator GFI (-1). The amount of this index is also between zero and one. GFI and AGFI indexes which suggested by Jöreskog and Sorbom (1989) are not related to volume of sample. (Daire Hoopan, 2008)

F- RMSEA index:

Root Mean Square Error of Approximation for good models is equal or less than 0.05.

Models with RMSEA more than 1.0 have weak fitting.

G- Chi square:

Quantity of chi square is highly related to the volume of sample and bigger sample increases the quantity of chi square.

H- NFI and IFC index:

NFI (normed fit index) or (Bentler&Bonett index) is acceptable for amounts more than 0.09 and shows the fitness of model. CFI index (Comparative Fit Index) more than 0.09 is acceptable which shows the fitness of the model. This index can test the proposed model by comparing with an independent model with no relation between variables. CFI is same as NFI but it has penalty for volume of sample group.

V- CONCLUSION

This paper seeks to highlight the deferent aspects of the first and second multivariate methods and at the end concludes the structural equation modeling as a second generation representative is more applicable and practical in current research problems because of some of its capabilities such as having more flexible assumptions, measurement error reduction by using CFA, having attractive interface and more visual, testing total model rather than individual coefficients, and at the end ability of manage difficult data including ; time series with auto-correlated error , incomplete and non-normal data.

LISREL or Structural equation modeling is a powerful multi-variables analytical technique from multi-variables regression family or an extension of General liner modeling which enables researchers to examine some set of regression equations simultaneously. Structural equation modeling is a general approach for examination of observed and latent variables that sometimes is called structural covariance analysis, causal modeling or LISREL but the common term is that structural equation modeling of SEM.

Except stated index in this paper, there are some other indexes in output of LISREL such asAIC (Akaike's Information Criterion), CAIC (an extension of the AIC, which more strongly penalizes models for lack of parsimony), and ECVA to determine the model with best fitness through different models, for example the model with minimum amount of AIC, CAIC, ECVA has more fitness. Some indexes are highly volume dependant and they have significance in high volume samples.

ACKNOWLEDGMENT

We acknowledge Brain Storm of technology (ToofanZehni) co. for preparation some data and supporting our study.

References:

Daire Hoopan, J. C. (2008). structural equation modeling: guidelines for determining modeling fit. Electronic Journal of Business Research Methods Volume 6 Issue 1 , (53-60).

Diana D.Suhr, P. (2003). exploratory and confimatory factor analysis. Statistics and data analysis .

Hoyle, R. H. (1995). The structural equation modeling approach: Basic concepts and fundamental issues. CA:Sage publication , 1-15.

Hu L. & Bentler, P. (1999). Cutoff criteria for fit indexes in covariance structure analysis. Conventional criteria versus new alternatives, structure equation modeling , 1-55.

Pearl, J. (2000). Causality: Models, Reasoning, and Inference. Cambridge : Cambridge University Press. ISBN 0 521 77362 8.

Princeton. (2010). LISREL: A General Computer Program for Estimating a Linear Structural Equation System Involving Multiple Indicators of Unmeasured Variables . NJ: Educational Testing Service , RB-72-56.

Tabachnick, B., & Fidell, L. (2001). Using Multivariate Analysis- ISBN 0-321-05677-9. Boston: Allyn and Bacon.

Westland, J. C. (2010). "Lower bounds on sample size in structural equation modeling. Electron. Comm. Res. Appl. 9 (6): 476–487. doi:10.1016/j.elerap.2010.07.003.

Akaike, H. (1974), "A New Look at the Statistical Model Identification," IEE Transactions on Automatic Control, 19 (6), 716-23.

Bagozzi, R.P., Yi, Y., and Phillips, L.W. (1991), "Assessing Construct Validity in Organizational Research," Adminstrative Science Quarterly, 36 (3), 421-58.

Barrett, P. (2007), "Structural Equation Modelling: Adjudging Model Fit," Personality and Individual Differences, 42 (5), 815-24.

Crowley, S.L. and Fan, X. (1997), "Structural Equation Modeling: Basic Concepts and Applications in Personality Assessment Research," Journal of Personality Assessment, 68 (3), 508-31.

Diamantopoulos, A. and Siguaw, J.A. (2000), Introducing LISREL. London: Sage Publications.

Fan, X., Thompson, B., and Wang, L. (1999), "Effects of Sample Size, Estimation Methods, and Model Specification on Structural Equation Modeling Fit Indexes," Structural Equation Modeling, 6 (1), 56-83.

Miles, J. and Shevlin, M. (1998), "Effects of sample size, model specification and factor loadings on the GFI in confirmatory factor analysis," Personality and Individual Differences, 25, 85-90.

SAS® Language and Procedures, Version 6, First Edition. Cary, N.C.: SAS Institute, 1989.

SAS® Online Doc 9. Cary, N.C.: SAS Institute.

SAS® Procedures, Version 6, Third Edition. Cary, N.C.: SAS Institute, 1990.

SAS/STAT® User's Guide, Version 6, Fourth Edition, Volume 1. Cary, N.C.: SAS Institute, 1990.

SAS/STAT® User's Guide, Version 6, Fourth Edition, Volume 2. Cary, N.C.: SAS Institute, 1990.

Schumacker, R. E. & Lomax, R. G. (1996). A Beginner's Guide to Structural Equation Modeling. Mahwah, New

Jersey: Lawrence Erlbaum Associates, Publishers.

Thorndike, R. M., Cunningham, G. K., Thorndike, R. L., & Hagen E. P. (1991). Measurement and evaluation in psychology and education. New York: Macmillan Publishing Company.

Truxillo, Catherine. (2003). Multivariate Statistical Methods: Practical Research Applications Course Notes. Cary, N.C.: SAS Institute.