Research Journal of Recent Sciences ____ ISSN 2277-2502
Vol. **3(1),** 57-63, January **(2014)** *Res.J.Recent Sci.*

Structural Equation Models and Its Application

Usman M.[1], Iqbal M.[1], Qamar Z.[1] and Shah S.I.A.[2]
[1]Department of Statistics, University of Peshawar, Khyber Pakhtokhawa, PAKISTAN
[2]Department of Mathematics, Islamia College University, Peshawar, Khyber Pakhtokhawa, PAKISTAN

Available online at: www.isca.in, www.isca.me
Received 31st October 2013, revised 9th November 2013, accepted 18th December 2013

Abstract

This study describe the general aspects of the Structural Equation Models as well as some extensions which have been proposed; these extensions looking for relax the underlying assumptions and generalize the technique. Then, based on a simplified model we consider the three stages to make an application; after we present path analysis which is a way to analyse Structural Equations Models and a special case called Confirmatory Factor Analysis. Finally we make an application based on the study monitoring the Future as an example of Confirmatory Factor Analysis.

Keywords: Structural equations, factor analysis, covariance matrix, observed, unobserved.

Introduction

The Structural Equation Models (SEM) is a full setting which permits modeling the relations amongst observed and unobserved variables, in this framework the causality is an underlying condition. In the history, the starting point of the SEM is not clear because has been applied in sociology, psychology and economics; one of the first approaches was the path analysis, which is a method to analyze SEM. The method includes a graphical representation or path diagram, based on it and considering some writing rules an equation system can be proposed. The equation system establishes the relations between the covariance of the observed variables and parameters of the model, finally we consider the classification of the effects.

With the time many extensions of the initial concept has been proposed, trying to give a broader scope and relaxing some of the assumptions. In this study we present the general model and some extensions; then, focus on a simplified model we describe the fitting process and finally make an application[1].

Material and Methods

General Model and Some Extensions: The Structural Equation Models are developed to scrutinize the hypotheses of no relations between observed variables and non-observed (latent) variables. This type of model contains a system of structural equations to express the relationship between variables. There are three kinds of variables which we use in SEM, which are the responses, the predictors and the latent or unobserved variables. In this chapter we present the general model assuming continuous latent variables and some extensions proposed based on the general model.

General Model Continuous Latent Variable: The model includes a random dependent variable vector y and a random independent variable vector x; each variable can be continuous or categorical. The observed variables in the vectors x and y are assumed to be generated by a set of continuous latent variables.

Suppose η be a system of linear structural equations consist of m latent dependent (endogenous) variables and ξ be a system of n latent independent variables, then

$$\eta = B\eta + \Gamma\xi + \zeta \qquad (1)$$

Where B (m x m) is a regression matrix with zero in the diagonal and $I - B$ non singular, ξ (n x 1) is the vector of latent independent variables, Γ (m x n) is a regression matrix of latent independent variables and ζ is the vector of residuals (latent errors); in the model, $E(\xi) = \kappa$, with covariance matrix of ξ denoted by Φ, $E(\zeta) = 0$ and the covariance matrix of ζ is denoted by Ψ .

Also assuming linear relations for y^*, a set of p latent response variables, and for x^*, a set of q latent variables

$$x^* = \Lambda_x \xi + \delta$$
$$y^* = \Lambda_y \eta + \varepsilon \qquad (2)$$

Where Λ_y (p x m) is a matrix of coefficients relating the latent response variable y^* to η and Λ_x (q x n) is a matrix of coefficients relating x^* to the latent variable ξ. ε (p x 1) and δ (q x 1) are random vectors of residuals (errors measurement) for y^* and x^* respectively with mean vector zero and variance covariance matrices are Θ_ε and Θ_δ. Assuming ε uncorrelated with η and ξ, and δ uncorrelated with ξ δ. Also δ, ε and ζ are mutually uncorrelated.

Now we establish the relation amongst observed and un-observed latent variables. When the observed variables in the vectors x and y are continuous, the identity transformation is used, that is $x^* = x$, $y^* = y$ and it will be explained later there are less considerations to specify the model.

On the other hand, when the observed variables are categorical we assume a monotonic relation between the latent and the predicted variable. That is, suppose that z^* is a latent variable based on the observed variable z which has C categories, then z^* is defined as follows:

$$z = \begin{cases} C-1, & \text{if } \tau_{c-1} < z^* \\ C-2, & \text{if } \tau_{c-2} < z^* \le \tau_{c-1} \\ \vdots & \\ 1, & \text{if } \tau_1 < z^* \le \tau_2 \\ 0, & \text{if } z^* \le \tau_1 \end{cases} \qquad (3)$$

The system describe throughout the equations 1 and 2 is called a Structural Equation Model. This system can be considered a general system in the sense that the observed variables can be categorical or continuous and one different structure is allowed for each set of observed variables x and y.

For the categorical dependent variables Muthén distinguishes two cases related with the identifying the distribution of observedvariables[1].In the first case, the joint distribution of the latent variables (y^*, x^*) is completely specified, under this approach we must to estimate the following parameters arrays: τ_y, τ_x, Λ_y, Λ_x, Θ_ε, Θ_δ, B, Γ, κ, and Ψ. Under this approach the assumption of multivariate normality for the latent variable is required due to the number of parameters involved.

Now, in the second case x is fixed, that is we do not impose any model structure to x, the conditional distribution $f(y^*/x)$ is specified, and the parameters arrays to estimate are: τ_y, Λ_y, Θ_ε, B, Γ, and Ψ. Considering this case, the multivariate normality assumption for the latent variable can be relaxed.

For the estimation Muthén mention basically two approaches, one based on the Maximum Likelihood Estimation and the second one based on Weighted Least Squares, the last one due to the heavy computations that involve Maximum Likelihood when the variables are categorical[1].

The previous general model can be extended to more than one populations, in fact the general model presented by Muthén is considering that the observational units come from g different populations where for each population can be formulated a SEM[1].

Some extensions to the general model have been proposed relaxing the multivariate normality assumption for latent variables that is, considering some of the continuous distribution or even discrete latent variables, or including additional levels, that is multilevel structural equations models. In the following section we describe some of the generalizations which have been proposed based on a general multilevel models.

Some Extensions: Consider a general multilevel model framework, which unifies factor and random coefficient models[2]. Under this approach the response model for the level-2 units is given by:

$$y_j = X_j\beta + \Lambda_j\eta_j + \varepsilon \qquad (4)$$

The subscript j refers to the level-2 units (clusters in random effects model). The interpretation of the terms involved in the equation 4 is different depending on the context (factor model or the random model). For instance, in random effects models the matrix Λ_j is the design matrix of the random effects, denoted by Z_j whereas in Factor model is called by factor loading matrix and is denoted by Λ.

Then the following model represents the structural relation of the latent variables

$$\eta_j = B\eta_j + \Gamma w_j + \zeta_j \qquad (5)$$

As in the previous section, the equation 5 denotes a linear model for the latent variables; in order to relax this assumption some non-linear approaches have been proposed. Under this approach the latent explanatory variables ξ_j can determine the latent response variables η_j through a structural model given by:

$$\eta_j = B\eta_j + \Gamma\alpha_j + \zeta_j \qquad \alpha_j \equiv g(\xi_j) \qquad (6)$$

The multivariate normality is considered for ξ_j and ζ_j, where $g(\xi_j)$ is a deterministic vector function. As it was mention before in the general model Muthén assumed multivariate normality for the latent variables, another possible extension of the general model is to assume that latent variables follow another continuous distribution or even a discrete distribution; then the structural model is defined throughout the probabilities that one unit belong to a specific latent category, for the discrete case the probability can be modeled using a multinomial log it model which may depend on a linear function of the covariates. In this case mathematical expectation of the latent variable is zero[1].

Other issue is the specification of the distribution of the latent errors ζ (also called disturbances). For continuous disturbances, the most common distribution is multivariate normality with mean zero and not necessary the same covariance matrix in each level. Despite of the inference is in many cases robust to the departures of the normal error some authors have proposed flexible parametric distribution such us mixture of normal distribution among them[3].

For discrete disturbances, again a multinomial log it models can be used, and finally a mixed continuous and discrete distribution. One possibility would be include both types of distribution for the latent variables in the response model or the second one is include only discrete latent variables in the response model and continuous latent variable in the structural model.

Until now we have briefly described the general Structural Equation Model and some extensions; to applied SEM distinguishes four stages[2], the first one is the identification of the model, then the estimation, after the predicting of the latent variables and the model selection.

However, we are going to focus in a simplified scenario describing three of the four stages to fit a model, one way to analyze a SEM and one special case.

SEM For Continuous Observed Variables: In this context the response variables y are also called endogenous and the explanatory variables x are the exogenous variables. In this section we assume the join distribution of the vector $z = (y \quad x)$ which include both sets of observed vectors is multivariate normal with order $N = p + q$; we consider continuous latent variables with $E(\eta) = 0$. Under this framework our main interest is explain the variability of the observed variables z as a functional form of the parameters and obtain the smallest difference between the variability of the model and the observed variability. The three phases considered for this model are identification, estimation and evaluating of the fit.

Identification is talking about whether each and every model parameter could be estimated by variance covariance information of observed data. If there are single unique estimate for each parameter, then the model is called just identified model. If one or more parameters have more than one estimate, then the model is called over identified model. Of the observed variance covariance information are not sufficient for estimation of parameters, then it called under identified model.

Estimation: The parameter estimation of SEM is based on the covariance matrix (Σ) of the observed variables. If the specified SEM is correct and with the known population parameters then $\sum$ will become $\sum(\theta)$, where Σ can be determine from the free model parameters in terms of its functional form.

But in real, $\sum$ should be estimated by observed covariance matrix S .From a process of iterations with a set of initial values, the covariance matrix can be estimated close enough to the observed matrix. After each iteration, resulted matrix is compared with the observed variance covariance matrix. Numbers of criterion are used in this purpose.

Maximum likelihood method, which assumes multivariate normality, is the one widely used which is minimizing the function[4]:

$$F_{ML} = log|\Sigma(\theta)| + tr(S\Sigma^{-1}(\theta)) - log|S| - (p+q)$$

and it can be shown that asymptotically $(N-1)F_{ML}$ is distributed as chi square with $\frac{1}{2}(p+q)(p+q+1) - t$ degrees of freedom where t is the number of free parameters.

There are several other methods developed for this purpose as un-weighted least squares, generalized least squared, etc.

Evaluation of Fit: "A model is said to fit the observed data to the extent that the covariance matrix it implies is equivalent to the observed covariance matrix (elements of the residual matrix are near zero"[5].

Several methods are used to assess the goodness of fit of SEM. Here we describe two overall measures of fit; Goodness of Fit Index (GFI) and Adjusted Goodness of Fit Index (AGFI) which is proposed by Joreskog and Sorbom[6]. Both indexes fall between zero and one and the values closer to one implies the better fit.

$$GFI = 1 - \frac{F[S, \Sigma(\hat{\theta})]}{F[S, \Sigma(\theta)]}$$

Numerator is the minimum value of the fitting function F for the formulated model and the denominator is the minimum value of the fitting function F when no model is formulated. Thus, GFI is measured "how much better the model fit as compared to no model at all"[7].

$$AGFI = 1 - \left(\frac{c}{df_h}\right)(1 - GFI)$$

C = number of non-redundant variances and co-variances of observed variable, df_h = degrees of freedom of the hypothesized model

Moreover, squared multiple correlation coefficients could be used to assess the reliability of observed variables in the system of structural equations. SEM is used in various kinds of scenarios. In this particular report we will consider most widely used path analysis and confirmatory factor analysis by presenting specific features and applications.

Path Analysis: One method to analyse a system of structural equation is Path Analysis. Three major steps could be identified in path analysis.

Path diagram: Decomposition of covariance and correlations: Identifying direct, indirect and total effects very first step in graphical representation of relationships between endogenous, exogenous and latent variables is the path diagram. here are some specific notations for path diagrams. Observed (endogenous and exogenous) variables are representing by boxes. Latent variables are representing by circles. And moreover, the structural relationship between variables is figured out as follows;

$X \rightarrow Y$: X is structurally influenced Y, but not vice versa
$X \leftarrow Y$: Y is structurally influence X, but not vice versa
$X \leftrightarrow Y$: X structurally influence Y and Y structurally influence X

In path analysis, we consider all variables as observed variables. As an example, following system contains two predictor variables and three response variables. Path diagram and relevant system of structural equations are below.

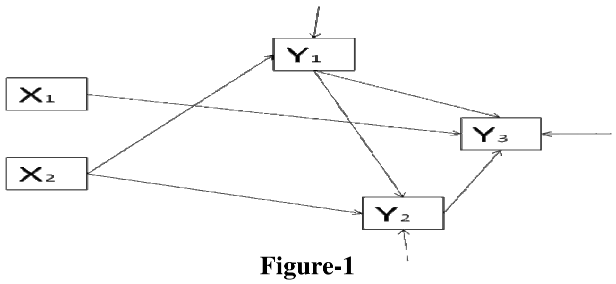

Figure-1
Path Diagram for example 1
Source[8]: Structural Equation Modelling

The system of structural equations could be defined as follows.

$$Y_1 = \gamma_{12} X_2 + \varsigma_1$$
$$Y_3 = \beta_{21} Y_2 + \gamma_{22} 2 + \varsigma_2$$
$$Y_3 = \beta_{31} Y_1 + \beta_{32} Y_2 + \gamma_{31} X_1 + \varsigma_2$$

The matrix representation of the system is:

$$Y = BY + \Gamma X + \varsigma$$
$$\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ \beta_{21} & 0 & 0 \\ \beta_{31} & \beta_{32} & 0 \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix} + \begin{pmatrix} 0 & \gamma_{12} \\ 0 & \gamma_{22} \\ \gamma_{31} & 0 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} + \begin{pmatrix} \varsigma_1 \\ \varsigma_2 \\ \varsigma_3 \end{pmatrix}$$

It is important to note that one predictor variable could be an explanatory variable in a regression equation of another predictor variable. The measurement errors of endogenous variables are assumed to distribute as normal with zero mean and the variance covariance matrix ψ . And exogenous variables are assumed to measure without error.

In the path analysis, covariance or correlation between two variables could be decomposed as a function of the parameters in the system. In broad, the covariance of X and Y can be decomposed into the sum of products of structural coefficients of all the variables with direct path to Y and the covariance of these variables with X. This is known as the first low of path analysis[8].

As an example, for above system of equations, the covariance of X_2 and Y_2 can be expressed as follows.
$\delta^2_{y_2 x_2} = \gamma_{22} \delta^2_{X_2} \beta_{21} \delta^2_{y_1 x_2}$. Since
$\delta^2_{y_1 x_2} = \gamma_{12} \delta^2_{X_2}$.Then$\delta^2_{y_2 x_2} = \gamma_{22} \delta^2_{X_2} + \gamma_{12} \beta_{21} \gamma_{12} \delta^2_{X_2}$

The influence of one variable on other variable could be divided in to two parts such as direct effect and indirect effect. As an example, for the same system of equation, the direct effect of X_2 on Y_2 is γ_{22}. The indirect effect of X_2 on Y_2 is coming via Y_1. It can be given by $\gamma_{12}\beta_{21}$. Total effect is the sum of direct and indirect effects.

Confirmatory Factor Analysis: Major reason of factor analysis is to explain the relationship between number of variables in terms of a small number of underlined, but unobserved random variables, called factors or latent variables[9]. Exploratory factor analysis (EFA) is one of the two widely use techniques, while the other one is Confirmatory Factor Analyses (CFA).

There are some differences between two methods. EFA does not have any detailed initial model, that is, observed or latent variables are not specified initially and each of the latent variables related to the all observed variables. Whereas a detailed and identified initial model is required for CFA; here we discuss about the elements of CFA. Path diagram is an important step in model formulation. Similar notations are valid in this case as we discussed earlier.

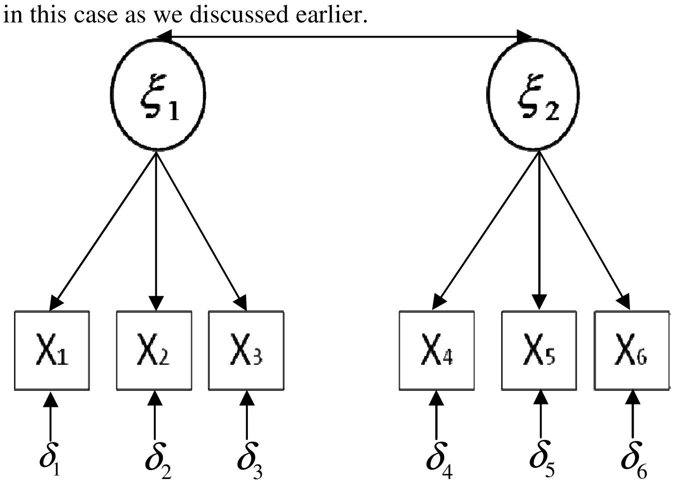

Figure-2
Path Diagram for a CFA Model
Source[4]: Structural Equations with Latent Variables p. 227.

The figure-2 is the path diagram for a CFA model which contains six observed variables and two latent variables. ξ_1 is the unobserved latent variable which is represented by observed X_1, X_2 and X_3. And ξ_2 is the latent variable for X_4, X_5 and X_6. In addition, δ_i are measurement errors of exogenous variables. In CFA we assume that exogenous variables are measured with an error which is distributed normally with zero mean. Unlike the path analysis model, in CFA we are dealing with two kinds of variability[8]. First is the variability associated with latent variables and other one is associated with δ_i.

The measurement model specification:

$$X_1 = \lambda_{11}\xi_1 + \delta_1$$
$$X_2 = \lambda_{21}\xi_1 + \delta_2$$
$$X_3 = \lambda_{31}\xi_1 + \delta_3$$
$$X_4 = \lambda_{42}\xi_2 + \delta_4$$
$$X_5 = \lambda_{52}\xi_2 + \delta_5$$
$$X_6 = \lambda_{62}\xi_2 + \delta_6$$

The matrix notation of the model is:

$$X = \Lambda_X\xi + \delta$$

Or,

$$\begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \\ X_6 \end{pmatrix} = \begin{pmatrix} \lambda_{11} & 0 \\ \lambda_{21} & 0 \\ \lambda_{31} & 0 \\ 0 & \lambda_{42} \\ 0 & \lambda_{52} \\ 0 & \lambda_{62} \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} + \begin{pmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \\ \delta_6 \end{pmatrix}$$

X is a (q X 1) matrix of observed variables. Λ_X is the (q X s) matrix of model parameters. And also assume that ξ is normally distributed vector with zero mean and covariance matrix given by $\Phi = \begin{bmatrix} \delta_1^2 & \delta_{12} \\ \delta_{12} & \delta_2^2 \end{bmatrix}$. And Θ_δ is the 4 X 4 covariance matrix of δ. Parameter estimation methods and testing the goodness of the model is similar which we discussed in Estimation and Evaluation of Fit.

Results and Discussion

An Application for the Confirmatory Factor Analysis(Monitoring the Future Study): The data set comprises of the factors that determine the use of alcohol and marijuana, collected from 1608 students of high school during 2001. The data was taken from Monitoring. The Future (MTF) study, carried out by Institute for Social Research at the University of Michigan. The survey has done with 8^{th}- and 10^{th}-grade students using four questionnaires which each have score questions about demographics and drug use. Consider the following as observed variables:

ALCLIFS(X_1): Mention the numbers occasions have you had alcoholic beverages to drink in your lifetime?
ALC12MOS(X_2): Mention the numbers occasions have you had alcoholic beverages to drink in the past 12 months?

ALC30DS(X_3): Mention the numbers occasions have you had alcoholic beverages to drink in the past 30 days?
XMJLIFS(X_4): Mention the numbers occasions have you used marijuana in your lifetime?
XMJ12MOS(X_5): Mention the numbers occasions have you used marijuana in the past 12 months?
XMJ30DS(X_6): Mention the numbers occasions have you used marijuana in the past 30 days?
TICK12MO(X_7): Within the last 12 months, how many times have you received a ticket (or been stopped and warned) for moving violations?
ACCI12MO(X_8): Within the last 12 months, how many times you were involved in an accident while driving?
Three latent variables are used as alcohol usage (AlcUseξ_1), marijuana usage (MarjUseξ_2) and social characteristics (Socialξ_3) of the student. Measurement model for the study can be formulated as below.
Model 1:

$$X_1 = \lambda_{11}\xi_1 + \varsigma_1 \qquad X_4 = \lambda_{41}\xi_2 + \varsigma_4$$
$$X_7 = \lambda_{71}\xi_3 + \varsigma_7$$
$$X_2 = \lambda_{21}\xi_1 + \varsigma_2 X_5 = \lambda_{51}\xi_2 + \varsigma_5 X_8 = \lambda_{81}\xi_3 + \varsigma_8$$
$$X_3 = \lambda_{31}\xi_1 + \varsigma_3 X_6 = \lambda_{61}\xi_2 + \varsigma_6$$

In the matrix notation;

$$X = \Lambda\xi + \varsigma$$

In the system, each equation gives a linear relationship between observed variables and unobserved (latent) variables with estimable random error term. Structural coefficients, λ_{ij} which are quantifying the structural relationship, have to be estimated. Moreover, latent variables are assumed to be associated each other and independent error terms also assumed. LISREL 8.80 has used as a statistical software for the analysis.

In the analysis of model 1, we found another important and common scenario in SEM, which is the estimated error variances, became negative!!! This is known as "Heywood Cases" which implies the misspecification of the model (improper solutions).

Due to the above drawback of the first model, a reduced model is analysed for the illustration purpose for the same data set.

Model 2:

$$X_1 = \lambda_{11}\xi_1 + \varsigma_1 \qquad X_4 = \lambda_{41}\xi_2 + \varsigma_4$$
$$X_7 = \lambda_{71}\xi_3 + \varsigma_7$$

$$X_3 = \lambda_{31}\xi_1 + \varsigma_3 X_6 = \lambda_{61}\xi_2 + \varsigma_6 X_8 = \lambda_{81}\xi_3 + \varsigma_8$$

System considers the all assumptions and definitions made for model 1.

Path diagram with estimated loadings is given below.

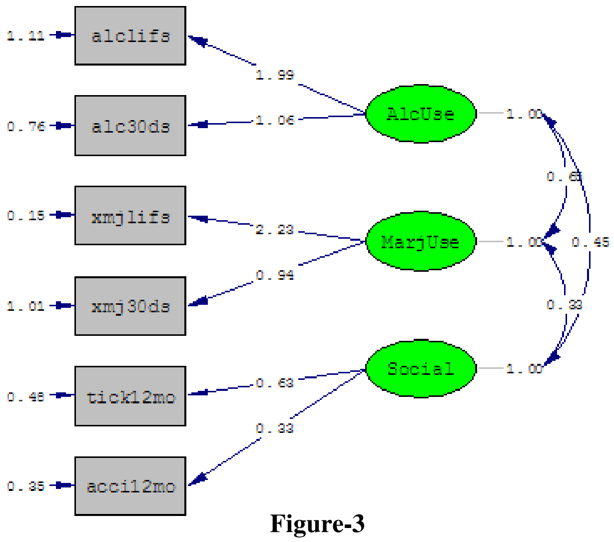

Figure-3
Path Diagram for Model 2

GFI and AGFI are used to examine the fit of the model. Both indexes give values close to 1 (0.9824 and 0.9382, respectively) which implies the better fit of the data. Most of the exogenous variables obtained a considerably large squared multiple correlation coefficients (60% as an average) which imply the reliability of each variable in the system (Table-A, Appendix). The variable XMJLIFS(X_4) has perfect reliability (R^2 = 0.97). Variables ALCLIFS(X_1), ALC30DS(X_3) and XMJ30DS(X_6) have moderate and acceptable reliabilities (R^2 is 0.78, 0.59 and 0.47 respectively)[7]. Table-1 Presents the parameter estimators for factor loadings and their standard errors.

Table-1
Un-standardized Factor Loading Estimations

Variable	Factor Loading	Estimate	Std. Error	T Value
ALCLIFS(X_1)	λ_{11}	1.99	0.05	39.44
ALC30DS (X_3)	λ_{31}	1.06	0.037	28.46
XMJLIFS(X_4)	λ_{42}	2.23	0.045	50.03
XMJ30DS(X_6)	λ_{62}	0.94	0.072	12.99
TICK12MO(X_7)	λ_{73}	0.63	0.051	12.41
ACCI12MO(X_8)	λ_{83}	0.33	0.028	11.74

Table-2
Squared Multiple Correlation Coefficients

Variables	Error Variance	R^2
ALCLIFS(X_1)	1.11	0.78
ALC30DS(X_3)	0.76	0.59
XMJLIFS(X_4)	0.15	0.97
XMJ30DS(X_6)	1.01	0.47
TICK12MO(X_7)	0.48	0.45
ACCI12MO(X_8)	0.35	0.24

Table-3
Standardized Estimates for Factor Loadings

Variables	Factor Loading	Estimate
ALCLIFS(X_1)	λ_{11}	0.8811
ALC30DS(X_3)	λ_{31}	0.7728
XMJLIFS(X_4)	λ_{42}	0.9842
XMJ30DS(X_6)	λ_{62}	0.6833
TICK12MO(X_7)	λ_{73}	0.6767
ACCI12MO(X_8)	λ_{83}	0.4878

Table-4
Correlation Matrix of Latent Variables

	AlcUse	MarjUse	Social
AlcUse	1	-	-
MarjUse	0.65	1	-
Social	0.45	0.33	1

Table-5
Correlation Matrix of Observed Variables

	ALC LIFS (X_1)	ALC30DS (X_3)	XMJ LIFS (X_4)	XMJ30DS (X_6)	TICK12MO (X_7)	ACCI12MO (X_8)
ALCLIFS (X_1)	5.06	-	-	-	-	-
ALC30DS (X_3)	2.10	1.88	-	-	-	-
XMJLIFS (X_4)	2.92	1.46	5.11	-	-	-
XMJ30DS (X_6)	1.12	0.75	2.09	1.89	-	-
TICK12MO(X_7)	0.51	0.39	0.44	0.20	0.88	-
ACCI12MO(X_8)	0.27	0.18	0.26	0.12	0.21	0.46

Table-1 shows that all un-standardized estimates for factor loadings are significant at 5% level. Standardized structural coefficients, which are presented in the table-3 in appendix, are used to assess the relative importance of observed variables on the latent variables they related[7]. The variable XMJLIFS(X_4) is the most reliable and strongest indicator for the latent variable "Marijuana Usage". As an average all observed variables seems to be considerably reliable indicators for their related latent variables except variable ACCI12MO for the latent variable "Social Behaviours". The table-4 in the appendix shows the considerable correlations between latent variables.

As a whole, considered three factors are seems to be significant to represent the information which contained by 6 observed variables and can be used in further analyses, but considering seriously the amount of information loss.

Conclusion

The nature of the causality considering three conditions isolation, association and direction of the causality[4].In order to establish the causal relation between two variables let say x_1 and y_1, we assume that the disturbance or latent error is unrelated with the explanatory factor, this is known as pseudo-isolation; another requirement is the association between the two observed variables (x_1 and y_1), and finally we require establish the direction of the causality for the three factors, x_1 and y_1 and the latent variables, which variable is the cause and which is the affected. When we propose a SEM we are taking into account these conditions; and we can check if our model really "fit" the data, but this is not enough as stated:

"If a model is consistent with reality then the data should be consistent with the model. But if the data are consistent with a model, this does not imply that the model corresponds to reality"[4].

The models are an approximation to the reality, and there is not "formal" statistical test to check the three requirements of the causal model are well described in the model.

Structural Equation Models is a broad topic with applications in different fields; some special cases are random effects models and factor analysis. Many extensions of the general model have been proposed however, still there is a big research area especially in multilevel models when non-contiguous responses are considered.

References

1. Muthén B., Latent variable structural equation model with categorical data, *Journal of Econometrics,* 22, 43-65**(1983)**
2. Skrondal A. and Rabe-Hesketh S., Generalized Latent Variable Modeling, Multilevel, Longitudinal and Structural Equation Models, Boca Raton, FL: Chapman & Hall/CRC **(2004)**
3. Verbeke G. and Lesaffre E., A linear mixed model with heterogeneity in the random-effects population, *Journal of the American Statistical Association,***91**, 217–221**(1996)**
4. Bollen K.A., Structural Equations with Latent Variables, NY: John Wiley & Sons **(1989)**
5. Hoyle, R.H. (ed.), Structural Equation Modeling. SAGE Publications, Inc. Thousand Oaks, CA. **(1995)**
6. Joreskog K.G. and Sorbom D., Lisrel user's guide, Version V, Chicago: International education services **(1981)**
7. Mueller R.O., Basic Principles of Structural Equation Modeling: An Introduction to LISREL and EQS, New York, NY: Springer **(1996)**
8. Shkedy Z., Structural Equation Modeling. An Introduction: Path Analysis and Confirmatory Factor Analysis, Diepenbeek: University of Hasselt; unpublished course notes **(2004)**
9. Johnson R.A. and Wichern D.W., Applied Multivariate Statistical Analysis (5th ed.), Upper Saddel River, NJ: Prentice Hall **(2002)**