Data_analysis4.calm

Structural Equation Modelling
"Causal" Modelling

R. C. Gardner
Department of Psychology
University of Western Ontario

Purpose

The label "causal modelling" has been applied to one application of structural equation modelling where some variables are hypothesized to be independent and others to be dependent variables in a set of regression equations. The label is inappropriate, however, in that causes and effects are not identified in the procedure, but rather a model is tested in which some variables are considered to be regressed on others. The model is said to be a good fit if the covariance matrix based on the parameters estimated is similar to the covariance matrix for the raw data. Causal modelling is an extension of path analysis where at least some of the variables are latent variables (or factors) defined by a number of measured variables.

Terminology

There are many terms and concepts associated with structural equation modelling. Following are some, most often used in "causal modelling", though many of them also apply to the other forms of structural equation modelling, path analysis and confirmatory factor analysis.

Cause	Equivalent to regression. A variable (A) is said to be caused by another variable (B) if it is significantly regressed on that variable. Note this is not equivalent to cause as generally understood.
Indicator variable	A measured variable
Latent variable	A factor that represents an aggregate of two or more indicator variables.
Exogenous variable	A (latent or indicator) variable that has no cause in the model (i.e., is not regressed on any other variable). It can correlate with other exogenous variables or can "cause" (endogenous) variables.
Endogenous variable	A (latent or indicator) variable that is caused by exogenous or endogenous variables.
Measurement model	The model linking the indicator variables to the latent variables. The index of this link is a factor loading.
Structural model	The model linking endogenous variables to other endogenous and exogenous variables by means of regression coefficients, and exogenous variables to other exogenous variables by means of correlation coefficients. In most cases, the variables are latent variables as defined by the measurement model.

Basic Mathematics

The matrix equation is as follows:

$$\Sigma \quad = \quad \begin{matrix} \Lambda_Y A(\Gamma\Phi\Gamma' + \Psi)A'\Lambda'_Y + \theta_\varepsilon & \Lambda_Y A\Gamma\Phi\Lambda'_X \\ \Lambda_X \Phi\Gamma' A'\Lambda'_Y & \Lambda_X \Phi\Lambda'_X + \theta_\delta \end{matrix}$$

where:

Σ = Sigma	The reproduced covariance matrix based on the parameters estimated
$A = (I - \beta)^{-1}$	Inverse of an Identity matrix minus the Beta matrix
A`	The transpose of the A matrix
β = Beta	Matrix of regression coefficients of endogenous variables (η) on endogenous variables (η)
Γ = Gamma	Matrix of regression coefficients of endogenous variables (η) on exogenous variables (ξ)
ϕ = Phi	Matrix of correlation coefficients among (latent) exogenous variables
Ψ = Psi	Covariance matrix for the errors (ς) (disturbance terms) in the equations
Λ_Y = Lambda Y	Matrix of factor loadings of Y indicator variables on endogenous (η) latent variables
θ_ϵ = Theta epsilon	Covariance matrix for the errors in the measurement of Y indicator variables
Λ_X = Lambda X	Matrix of factor loadings of X indicator variables on exogenous (ξ) latent variables
θ_δ = Theta delta	Covariance matrix for the errors in the measurement (δ) of X indicator variables

Specifying the Model

In structural equation modelling, the objective is to estimate values in the various matrices that are consistent with a particular theoretical model, and determine how well the obtained covariance matrix (S) agrees with the covariance matrix, $\sum$, computed from the estimated values. The model is assessed in terms of two measurement models in which the exogenous indicator variables (X's) and the endogenous indicator variables (Y's) are linked to latent variables, and a structural model in which the latent variables are linked to each other.

The model is essentially a matrix representation of a series of regression equations. It is most frequently shown in the form of a diagram, and in fact it is possible to define the model for **AMOS** in terms of a path diagram. As indicated above, the aim of structural equation modelling is to define a model that accounts as much as possible for the covariances among the measures. In this regard, reference is often made to measures of goodness of fit. There are many such measures but the majority of them are defined in terms of a χ^2 statistic assessing the degree of departure of the hypothesized to the empirical model. The number of degrees of freedom for the χ^2 statistic is:

$$df = \frac{p(p+1)}{2} - k$$

where:

p = the number of indicator variables
k = the number of parameters estimated in the model

Model Equations

Measurement Model for X

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

Measurement Model for Y

$$Y = \Lambda_Y \eta + \epsilon$$

Structural Model

$$\eta = \beta\eta + \Gamma\xi + \zeta$$

where:

η = eta
ϵ = epsilon
ξ = ksi
δ = delta
ς = zeta
Γ = gamma
β = beta
Λ = lambda

Identification

Identification is a critical condition for structural equation modelling, and although it is fairly straightforward to define, it can often be very difficult to determine. Identification refers to whether or not the estimates in the model are uniquely determined. A necessary precondition for identification is that the degrees of freedom are greater than 0. If the degrees of freedom are equal to 0, the model is said to be *just identified*, and the model perfectly reproduces the sample covariance matrix (i.e., $\chi^2 = 0$). Such a model has little value because there is in truth no estimate of error for the model and its generality cannot be determined. If there are fewer than 0 degrees of freedom, the model is said to be *under-identified* and the parameters cannot even be estimated. This is a very minimal condition, however, and there are other factors that can influence the identifiability of the model. **AMOS** can detect many of them and if it does it prints the statement **“This model is inadmissible”**. Often this is associated with negative variance estimates, but sometimes the program detects a potential problem at which it prints that “some parameters may not be identified”. (It is even possible that such a warning may not be given and the model may still not be identified). It is possible that a model that obtains such warnings can be modified to overcome the problem, but this isn’t always the case.

Generally, AMOS guards against obvious errors. Thus, for each latent variable, it is necessary to fix the loading of an indicator variable at 1 (this is automatically done if the instructions given below for running AMOS are followed). It is also necessary to have 1's associated with the measurement errors and with the disturbance terms.

Model Modification

As indicated above, structural equation modelling is a confirmatory analysis. The model is proposed, parameters are estimated, and the degree of fit of the model to the data is assessed. It is possible, however, to modify the model after the fact to determine whether the fit can be improved, and to assess whether the improvement resulted in a significant reduction of the residuals. Such modifications should be done cautiously because it will generally be the case that the model can be improved, at least on the sample data. Such model trimming should be well based on a strong theoretical structure, and at a minimum the revised model should be tested on a fresh and independent sample of data.

Comparing Models

Models based on the same data file can be compared, providing one model is nested in the other. That is one model should contain all the paths of the other plus one or more others. Thus, when performing model modification, one or more paths can be eliminated from a model and the resulting model can be said to be nested in the first. Subtracting the χ^2 value assessing the fit for the first model (the smaller χ^2) from that for the second (the larger χ^2) results in a difference that is itself distributed as χ^2 with degrees of freedom equal to the difference in degrees of freedom for the two values. Of course, the same rationale can be applied if some paths are added to a model, and the two models can be compared.

Measures of Fit

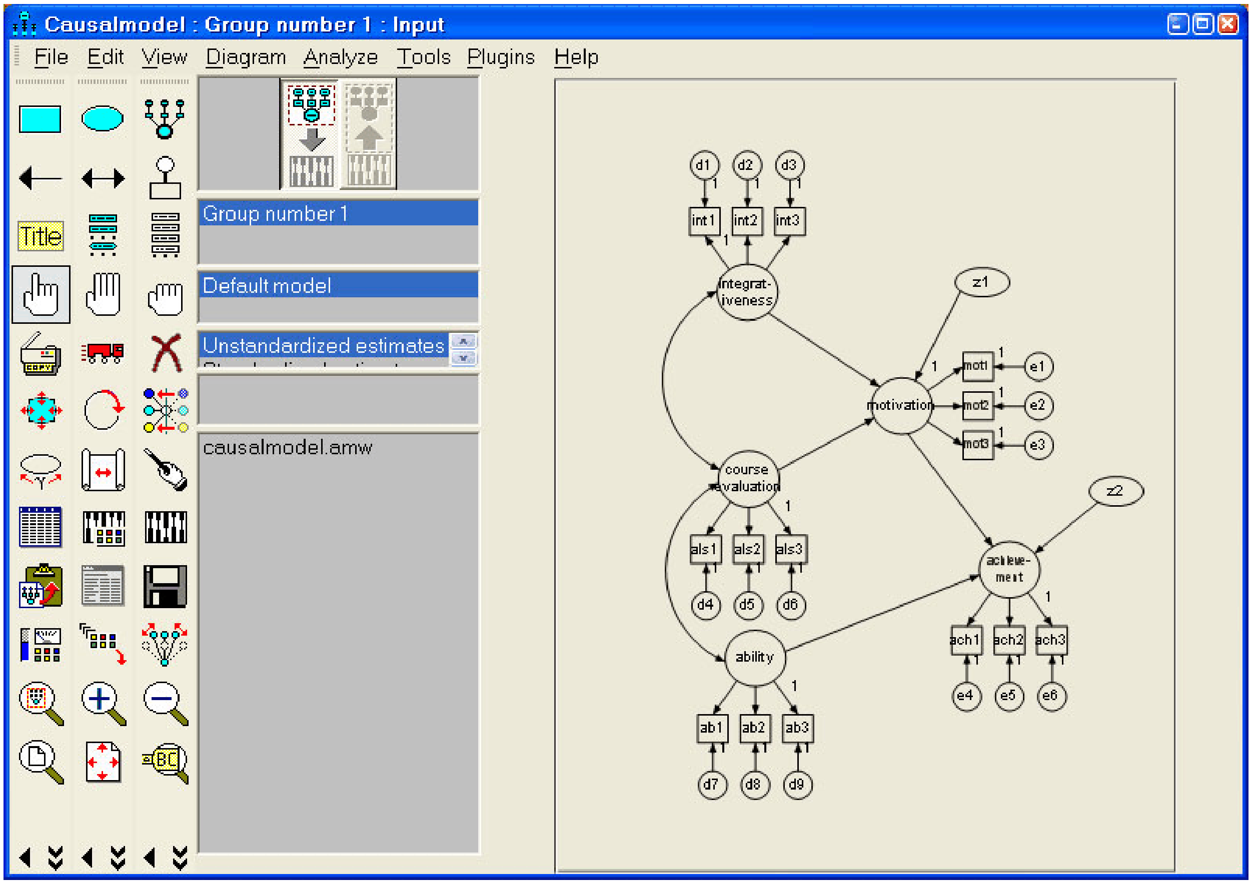
There are many measures of goodness of fit of the model. One measure is common to many of them. That is, the χ^2 measure of goodness of fit. Generally, a good model would be one where the χ^2 is roughly equal to the degrees of freedom, but a commonly accepted index is $\chi^2/df < 2$. A χ^2 that is not significant indicates that the fit is fairly good, though there are other measures that have also been proposed. These can be classified in terms of:

a. *Comparative Fit*	For these indices, the χ^2 for the model (or some function of it) is compared with the χ^2 (or some function of it) for a model assuming independence. Examples are NFI,NNFI,IFI,CFI,RMSEA.
b. *Absolute Fit*	These indices contrast the χ^2 for the model with the degrees of freedom for the model. An example is MFI.
c. *Proportion of Variance*	These indices assess the proportion of variance for the covariance matrix based on the estimated parameters with the original covariance matrix. Examples are GFI, AGFI.
d. *Parsimony Fit*	These indices take into account the degree of parsimony or complexity of the model. Examples are PGFI, AIC, CAIC.
e. *Residual-Based* Fit	These indices are based on the difference between the values in the original covariance matrix (S) and those in the covariance matrix (Σ) calculated on the basis of the estimated parameters. An example is RMR.

Running AMOS

AMOS can be run from SPSS (if it is installed) or directly from the AMOS program. If running from SPSS, ***Click on AMOS*** and you are presented with the Graphics Editor (see below) in which you are to draw the model using the tools presented there. If running from AMOS, ***Click on Graphics*** and you will be presented with the Graphics Editor.

To enter the data, ***click on File*** on the tool bar at the top and ***select Data Files***. This presents a window (Data Files). ***Click on File Name***. Select the data file from the appropriate directory and ***Click on OK***. This will present you with the following **AMOS Graphics Editor**:



To draw the causal model, you would proceed as follows:

1. (Note, if the Graphics Editor is not clear, ***click on File*** and ***select New***). Begin by drawing the large circles (latent variables). ***Click on the circle*** (row 1, column 2 in the icons at the left), move the cursor to where you want to draw the circle, hold down the left mouse key, and move the mouse to draw it. Release the left mouse key and you have your first circle. You can draw more circles, or you can copy this one.
2. To copy a circle, ***select the copy machine*** (row 5, column 1 of the icons). Move the cursor to the existing circle, ***click on it*** (it turns red), and holding the mouse key down, move the cursor to copy the circle in another place. Release the mouse key, and this will produce the circle. Repeat the process as often as you want. You could also use Step 1 to draw the small circles or the squares, and duplicate them in the same way.
3. To draw the small squares (indicator variables) and the small circles (measurement errors), ***select the icon in row 1, column 3***. Move the cursor to the latent variable (large circle) of interest, and ***click the left mouse key*** for as many indicator variables you need for that latent variable.
4. If the arrangement of the indicator variables is not optimal, rotate them around the latent variable. ***Select the icon indicating rotation*** (row 6, column 2), move the cursor to the

latent variable and ***click the left mouse key*** for each 1/4 turn.

5. To draw bidirectional arrows (correlations), ***select the two headed arrow*** (row 2, column 2). Move the cursor to the circle (latent variable or measurement error) of interest, hold down the left mouse key, drag the mouse to the other circle, and release the key. Do this for all bidirectional arrows.
6. To label the latent variables (the circles), ***right click on the circle of your choice and on Object Properties***. This presents a window (Variable Name). Type in the label (the name you want to appear in the figure) and the variable name (the name of the variable in your data file). They need not be the same. When you have finished with one variable, you can close the window and double click on another variable, or if you prefer you can leave the window open and double click on another variable. (Sometimes, you may have to move the window to see the other variables). You can use the same procedure to label the small circles (errors of measurement) and the small squares (the indicator variables), though for the indicator variables, a better procedure is to ***select the icon in row 3, column 3*** . This will open a window listing all the variables in your data file. You then click on a variable and drag it to the appropriate square, depositing the label.
7. It is necessary to indicate errors in the equation for each "dependent variable" (i.e., latent variable with an arrow leading to it) by drawing a small circle and labelling it (i.e., e1, e2, etc...), and drawing a directional arrow from the small circle to the latent variable. It is also necessary to indicate a regression weight of 1 for this error. This is done by ***right clicking*** on the arrow, ***clicking on parameters*** in the Object Properties window and typing the value 1 under ***regression weight***. Failure to do so will result in the element being unidentified.
8. Often there are additional forms of output you might want to obtain such as the Standardized Estimates. To add these to your output **click on View/Select** in the tool bar, choose ***Analysis Properties*** on the drop down menu, select ***Output***, and ***click on Standardized Estimates***. You might also want to click on Modification Indices (and indicate a relatively large minimum value, such as 30.)
9. To run AMOS, ***click on Analyze*** in the tool bar, then ***click on "Calculate estimates"***. Alternatively you could ***click on the abacus*** (row 8, column 3). If this is the first time you drew this model, you will be instructed to type in the file name. Type in the name (it will add the extension AMW) and ***click on the abacus*** again. If there are no errors, it will run. Otherwise you will be given some message. To see your output, move to the tool bar, ***select View/Select***, and then ***click on Text Output*** in the drop down menu.
10. Two other tools you might use are:

    (a) The large X (row 5, column 3). This allows you to erase something. ***Click on it***. Move your cursor to the item you want to delete and click on it.

    (b) The truck (row 5, column 2). This allows you to move something. ***Click on it***. Move to the item you want to move, click on it, and drag it to where you want it placed.