Data_analysis3.calm

Structural Equation Modelling

Confirmatory Factor Analysis

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

Purpose

Confirmatory factor analysis is a form of structural equation modelling. In this case, however, the purpose is to evaluate an hypothesized factor structure of the relationships among a series of individual difference variables by determining how well the model accounts for the covariances among the variables. In confirmatory factor analysis, the relationships among the variables are hypothesized to be due to a series of underlying factors, which may or may not be correlated. It is possible to permit a variable to define more than one factor, though a parsimonious model would typically have any one indicator variable defining only one factor.

Terminology

Indicator Variable.	A measured variable.
Latent Variable.	A factor that represents an aggregate of indicator variables.
Measurement Model	The model linking the indicator variables to the latent variables. The index of this link is a factor loading.

Basic Mathematics

The matrix equation is as follows:

$$\sum = [\ \Lambda_X \phi \Lambda'_X + \theta_\delta\ ]$$

where:

$\sum$ = Sigma	The reproduced covariance matrix based on the parameter estimates
Λ_X = Lambda X	Matrix of factor loadings of X indicator variables on the factors (latent variables)
Λ'_X	The transpose of Lambda X.
ϕ = Phi	Matrix of correlations among the factors (latent variables)
θ_δ = Theta Delta	Variance/covariance matrix of the errors of measurement (δ) of the indicator variables (X)

Specifying the Factor Structure

A model is proposed in which the indicator variables are hypothesized to be determined by underlying factors and measurement errors reflecting to a large extent the discrepancy between the variable and the factor. The model is specified by indicating which variables define each factor. This is described in detail below in the section **Running AMOS**. In essence, you specify directional arrows (regressions) from the factor to the relevant indicator variable(s). When specifying the factor, its variance must be defined. This can be done in a few ways, but the most direct one is to set the scale by fixing one factor loading at 1. In addition, it is necessary to estimate the correlations among those factors that you assume are correlated with one another, as well as the errors of measurement for each of the indicator variables.

Output

In reports, the model is most frequently presented in the form of a diagram, and in fact as we shall see, it is possible to define the model for AMOS in terms of the diagram (see below). The output consists of the estimated parameters. There are three different types of estimates. One form consists of the regression coefficients linking the factors to the indicator variables. In standard score form (which is usually how they are presented in reports), these are the structure coefficients. If the factors were orthogonal, they would also correspond to the correlations of the factors with the variables (the structure coefficients, as is the case with factor analysis). A second type of estimate is comprised of the variances of the errors of measurement, and if estimated the covariances (or correlations among the measurement errors). It is seldom the case that the variances are presented on the path diagram, but often the correlations between measurement errors are shown. The third type of estimate is composed of the correlations among the factors (latent variables). AMOS provides Standard Errors and Critical Ratios to evaluate the significance of each estimate. The critical ratios are the equivalent of standard normal deviates, and a value greater than 2.00 is considered significant (recall that $Z = \pm 1.96$ is significant at the .05 level, two-tailed).

The output also presents a χ^2 statistic assessing the degree of departure of the obtained covariance matrix from the covariance matrix calculated from the hypothesized structure. The value of the degrees of freedom for the χ^2 statistic is:

$$df = p(p+1)/2 - k$$

where:

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

There are also many measures of goodness of fit, which fall into one of five different types (comparative fit indices, absolute fit indices, proportions of variance accounted for, indices that adjust for parsimony, and measures of residual variation in the fitted covariance matrix). Typically, it is the case that one statistic from each type is presented in reports, but as yet there does not appear to be much agreement on which are favoured.

Model Equation

Measurement Model for X

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

where:

X = An indicator variable
Λ_X = Lambda X, the factor loading
ξ = Ksi, the factor (latent variable)
δ = delta, measurement error

Important Considerations

There are two concepts that are important for confirmatory factor analysis. One is identification, and the other is admissibility. Often it is possible that by respecifying the model or altering parts of it, some of the difficulties described can be eliminated.

Identification

Identification refers to whether or not an estimate in the model is uniquely determined in that only one value exists for the sample of data given the proposed model. If an estimate is unidentified, the model is also unidentified. There are a number of things that can result in an unidentified model. Some are simply operational in that mistakes are made in instructing AMOS how to define the model. These are easily avoided. Others are due to mis-specifications of the model, limitations in the data, or complications in the nature of the model. These are less easy to detect, but some of them are identified by the AMOS program and warnings given.

Two of the more obvious operational impediments to identification are (a) failure to set the scale of a factor, and (b) failure to identify a link between measurement error and an indicator variable. Impediment (a) can be avoided by fixing the loading for one indicator variable at 1 for each latent variable, while impediment (b) can be avoided if the measurement error for each indicator is estimated. Both are done automatically if item 3 in the instructions presented below is followed. This shows for (a) on the figure with a “1" associated with one indicator variable for each latent variable, and for (b) with “1's” associated with each measurement error.

Admissibility

If AMOS detects a difficulty with identification, it will indicate which estimate(s) may not be identified, and also print a warning to the effect that “***This model is inadmissible***”. It is also the case that on some occasions, the nature of the data and the model combine to produce estimates that are outside of a meaningful range, for example a negative variance, and the same warning message is printed.

Model Modification

Once the model is proposed, the parameters are estimated, and the degree of fit of the model to the data is assessed, it is possible that one might consider modifying the model. If some loadings or some correlations among the factors are not significant, it may well mean that the model is not confirmed. It is possible, however, to change the model, but model modification should not be taken lightly. Modifications should be done cautiously, because it will generally be the case that the model can be improved, at least on the sample of 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.

One way to identify possible additions to a model is by considering the modification indices. These are estimates of how much the χ^2 will be reduced if the parameter associated with the modification index is estimated (this applies only to the addition of the one new estimate to the existing model with no other changes). Modification Indices can be added to the output by selecting View/Select at item 7 in the instructions for Using AMOS below, clicking on Output, and choosing modification indices. One should also indicate a minimum value in order to limit output (a value of 10 is reasonable - - a $\chi^2 = 6.64$ is significant at the .01 level with 1 degree of freedom), but this value may differ depending upon the interests of the researcher.

Comparing Models

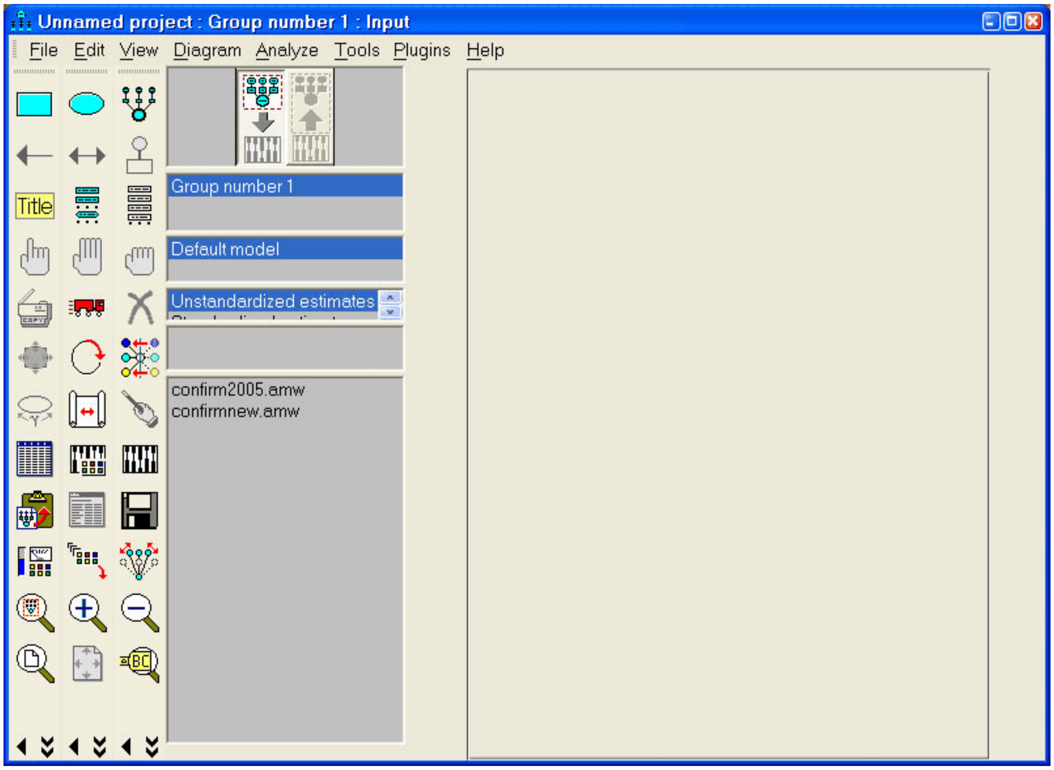
Models based on the same data file can be compared, providing one model is nested in the other. That is, one model should contain all the elements (i.e., factor loadings, error variances) of the other plus one or more other ones. Thus, when performing model modification, one or more correlations can be eliminated from a model, and the resulting model can be said to be nested in the first. Subtracting the χ^2 assessing the fit for the second model from that for the first, 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 first and second χ^2 values. Of course, the same rationale can be applied if some correlations are added to a model, and the two models can be compared.

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 next page) 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 confirmatory factor 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. 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.)
8. 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 (K)*** 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.
9. 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.