Structural equation modeling of paired comparison and ranking data using Mplus

Albert Maydeu-Olivares
Faculty of Psychology, University of Barcelona
Marketing Department, Instituto de Empresa
Ulf Böckenholt
Faculty of Management, McGill University

Supplementary materials for the Psychological Method Website for the manuscript

"Structural equation modeling of paired comparison and ranking data"

Address correspondence to: Albert Maydeu-Olivares. Faculty of Psychology. University of Barcelona. P. Valle de Hebrón, 171. 08035 Barcelona (Spain). E-mail: amaydeu@ub.edu.

Introduction

This document details how to estimate Thurstonian choice models using Mplus 2 (Muthén & Muthén, 2001). It is accompanied by three actual Mplus input files which can be used to reproduce the ranking and paired comparisons example provided in the body of the paper. Also, the two data files used in these examples are also provided. The first input file (ranking.inp) can be used to fit the unrestricted, Case III, and Case V models to the career ranking data example. The second input file (pc1.inp) can be used to fit these same models to the cars paired comparison example. Finally, the third input (pc2.inp) file can be used to fit an exploratory factor model with two factors to the cars paired comparison example (with or without restrictions on the latent preference means).

Estimation of basic Thurstonian models using Mplus

This section describes how to estimate the three basic Thurstonian models described in the body of the paper (the unrestricted, Case III, and Case V models). Both paired comparison and ranking models are discussed.

Estimating these models is similar to the task of estimating a confirmatory factor analysis model where all the indicators are dichotomous variables. However, there are a few important differences that may be summarized as follows:

1) The number of latent variables in Thurstonian paired comparisons models is equal to n, the number of objects.
2) The 'factor loadings matrix' $\mathbf{A}$ is a fixed matrix.
3) The 'factor' means $\boldsymbol{\mu}_t$ are estimated.
4) Some of the 'factor' variances (i.e., some diagonal elements in $\boldsymbol{\Sigma}_t$) are estimated

To estimate a Thurstonian ranking model, the additional constraint needs to be added that all variances of the unique factors in the factor analysis model are zero.

Although Thurstonian scaling models are defined in terms of their mean and covariance structure, because of the binary structure of the data, they are estimated from the thresholds and tetrachoric correlations. In Mplus 2 there are two ways to estimate covariance structures from matrices of sample tetrachoric correlations. One approach is to use in Equation (16) in the body of the paper

$$\boldsymbol{\Theta} = \mathbf{I} - diag\left(\boldsymbol{\Lambda}(\mathbf{I}-\mathbf{B})^{-1}\boldsymbol{\Psi}(\mathbf{I}-\mathbf{B})^{-1\prime}\boldsymbol{\Lambda}'\right). \tag{I}$$

Here, $\mathbf{I}$ denotes an identity matrix and $diag(\mathbf{X})$ denotes a diagonal matrix whose diagonal elements are equal to the diagonal elements in matrix $\mathbf{X}$.

When using this approach, the diagonal elements of $\boldsymbol{\Theta}$ are not free parameters, but functions of the remaining parameters of the original covariance structure. The alternative approach is to use

$$\mathbf{D}\left(\boldsymbol{\nu} + \boldsymbol{\Lambda}(\mathbf{I}-\mathbf{B})^{-1}\boldsymbol{\alpha}\right), \qquad \mathbf{D}\left(\boldsymbol{\Lambda}(\mathbf{I}-\mathbf{B})^{-1}\boldsymbol{\Psi}(\mathbf{I}-\mathbf{B})^{-1\prime}\boldsymbol{\Lambda}' + \boldsymbol{\Theta}\right)\mathbf{D} \tag{II}$$

where $\mathbf{D} = \left(\mathrm{Diag}(\boldsymbol{\Sigma})\right)^{-\frac{1}{2}}$, is a diagonal matrix whose diagonal elements are the inverses of the standard deviations of $\mathbf{y}^*$ under the model.

The following steps summarize the application of Mplus to the estimation of Thurstonian models with both unrestricted and diagonal covariance structures:

1) Declare all observed variables as categorical.

2) Declare that the analysis involves a mean structure.

3) Select the variance normalization (II). This is denoted in Mplus 2 as ‘theta’

parameterization. The default variance normalization (I) does not yield correct results for Thurstonian scaling models.

4) Choose an estimator for the last stage in the estimation process. For paired comparisons and ranking modeling, diagonally weighted least squares (DWLS) estimation is recommended. In Mplus, this estimator is denoted as WLSM when goodness of fit testing is performed using T_M, and by WLSMV when goodness of fit testing is performed using T_{MV}. Thus, the parameter estimates and standard errors will be identical when using WLSM or WLSMV.

5) Fix the intercepts ν in Equation (16) in the body of the paper to zero. This is because by default these are free parameters in Mplus.

6) Free the 'factor' means $\boldsymbol{\alpha}$ (except for the last one). This is because by default these parameters are fixed at zero in Mplus.

7) Set the 'factor loadings' equal to the fixed constants given by the $\mathbf{A}$ matrix

8) Specify the diagonal elements of $\boldsymbol{\Theta}$ to be free parameters. If a model with equal pair specific error variances is desired, the diagonal elements of $\boldsymbol{\Theta}$ can be set equal.

9) Choose a model for the covariance matrix of the discriminal processes. If an unrestricted model is desired, then the 'factor' variances for the first and last factors are fixed to 1, and all the covariances involving the last 'factor' are fixed to zero. If a Case III model is desired, the 'factor' variances are free parameters, except for the last one, which is fixed to 1, and all the 'factor' covariances are fixed at 0. Finally, if a Case V model is desired, all 'factor' variances are fixed to 1, and all 'factor' covariances are fixed to 0.

Ranking models are estimated similarly. First, the observed rankings (or orderings) are transformed into binary paired comparisons. The steps 1) to 9) can be followed with the exception of 8) which requires that all of the elements of $\boldsymbol{\Theta}$ must be set to zero. When using WLSM in fitting ranking data the degrees of freedom reported by the program are incorrect, as MPLUS assumes that there are no redundancies among the sample thresholds and tetrachoric correlations. To test the fit of a ranking model estimated using WLSM one needs to obtain manually a *p*-value for the value of the goodness of fit statistic using the number of degrees of freedom printed in the Mplus output minus r – as given in Equation (12) in the body of the paper. It is not possible to correct the mean and variance adjusted statistic in this way. Therefore, WLSMV should not be employed when fitting ranking data using Mplus.

Estimation of Thurstonian factor models using Mplus

This section describes how to estimate the exploratory factor models (with or without restrictions on the latent preference means) described in the body of the paper. Models for paired comparison data as well as for ranking data are discussed.

Estimating these models is similar to the task of estimating a second order factor analysis model where all the indicators are dichotomous variables, the structure for the first order factors is confirmatory, and it is exploratory for the second order factors. Thus, Thurstonian factor models for paired comparisons can be specified within the framework of Equation (16) in the body of the paper by letting $\boldsymbol{\nu} = \mathbf{0}, \boldsymbol{\alpha} = \boldsymbol{\mu}_t$, and $\boldsymbol{\Theta} = \boldsymbol{\Omega}^2$. In addition, we let $\boldsymbol{\Lambda} = \left(\mathbf{A} \mid \mathbf{0}\right)$, $\mathbf{B} = \begin{pmatrix} \mathbf{0} & \boldsymbol{\Lambda}_t \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$, and $\boldsymbol{\Psi} = \begin{pmatrix} \boldsymbol{\Psi}_t^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix}$. When the variances of the unique factors are set equal to each other $\boldsymbol{\Psi}_t^2$ is to be specified as in Equation (30) of the body of the paper. Thurstonian factor models where the mean preferences depend on the common factors require the additional specification of $\boldsymbol{\alpha} = \left(\mathbf{1}\gamma \mid \boldsymbol{\mu}_{\xi}\right)$.

To fit a m-factor model, two of the above steps need to be replaced. Step 9) is changed to

9) Invoke m additional factors whose indicators are the first order factors (i.e., the discriminal processes). Specify the loadings matrix for the second order factors to be in echelon form. In addition, set all the factor loadings for the last object equal to zero. The unique variances are to be set equal to each other except for the last one which is fixed at one See Equation (30) in the body of the paper. The variance of these second order factors are set to one, and they are uncorrelated. Finally, to fit a m-factor model with the mean preferences depending on the common factors, Step (6) needs also to be replaced by

6) Set all the first order factors means equal to each other, except the last 'factor' mean which is to be fixed at 0, and set all the second order factor means free.