Multiple group analysis using multivariate format: A DSEM Example

Tihomir Asparouhov & Bengt Muthén

January 20, 2022

In this note we illustrate how multiple-group analysis can be estimated by setting up a wide format single group model. Instead of analyzing the data as in the usual format facilitated by the GROUPING option, see Mplus User's Guide example 5.14, we can estimate the model by setting up a multivariate model. This approach can be used in some special instances where the GROUPING variable is not available or implemented in Mplus or when there are multiple grouping variables. An example is multiple-group DSEM where the GROUPING option is not yet available. The approach can be used to create MODEL TEST involving parameters from the different groups or to create parameter comparison using MODEL CONSTRAINT. The approach can also be used to constrain parameter to be equal across groups (parameter invariance across groups).

Let Y be the vector of variables in the model and g be the grouping variable for $g = 1, ..., G$. The variables in the multivariate-format multiple-group model are $(Y_1, Y_2, ..., Y_G)$, where Y_g represents the variable Y in group g. The models for Y_g are specified as in the original model. The model must also impose independence between Y_i and Y_j. The multivariate-format multiple-group approach has two steps: the data preparation step and the analysis step. The data preparation step creates G copies of the original data in terms of columns in the data set. In this step also the vector $(Y_1, Y_2, ..., Y_G)$ specifies missing values for all variables but the Y_g variable for the actual group the observation is in.

We illustrate the approach using a bivariate AR(1) DSEM model similar to Mplus User's guide example 9.32. The model we want to estimate is given in Figure 1. Next, we illustrate how the Figure 1 model can be estimated

in multiple group settings (the GROUPING option is not available yet for DSEM models). This data contains 100 clusters. Group 1 consists of the first 50 clusters and group 2 consist of the remaining 50 clusters. The data preparation stage is illustrated in Figure 2. This is a TYPE=BASIC analysis where the variables are replicated G times and are specified as missing for all but one copy of the variables. All of these manipulations are implemented in the DEFINE command. The final analysis step is illustrated in Figure 3. The model in this step is G copies of the model in Figure 1. To ensure that all these model copies are independent of each other, the model specification: y1-z1 with y2-z2@0 is added.

The approach used here is also similar to the approach illustrated in Section 2 of

Hamaker, E.L., Asparouhov, T, & Muthén, B. (2021). Dynamic structural equation modeling as a combination of time series modeling, multilevel modeling, and structural equation modeling. To be published as Chapter 31 in: The Handbook of Structural Equation Modeling (2nd edition); Rick H. Hoyle (Ed.); Publisher: Guilford Press.

and annotated in details in

https://ellenhamaker.github.io/DSEM-book-chapter/PrePost.html

Note, however, that in this example the two groups are not held independent from each other because the the groups are formed as two different time points of the same sample. Apart from that, however, the methodology is the same.

Figure 1: Bivariate AR(1) DSEM model

```
DATA:   FILE = ex9.32.dat;

VARIABLE:       NAMES = y z subject;
        CLUSTER = subject;
        LAGGED = y(1) z(1);

ANALYSIS:       TYPE = TWOLEVEL; ESTIMATOR = BAYES;

MODEL:  %WITHIN%
        y ON z&1;
        z ON z&1;
        y ON z&1;
        z ON y&1;
        %BETWEEN%
        y WITH z;
```

Figure 2: Data preparation step for two-group Bivariate AR(1) DSEM model

```
DATA:   FILE = ex9.32.dat;

VARIABLE:       NAMES = y z subject;
usevar=subject g y1 z1 y2 z2;

define:
if (subject le 50) then g=1 else g=2;
y1=y; z1=z; y2=y; z2=z;
if (g==1) then y2=_missing;
if (g==1) then z2=_missing;
if (g==2) then y1=_missing;
if (g==2) then z1=_missing;

ANALYSIS:       TYPE = basic;

savedata: file is ex9.32double.dat;
```

Figure 3: Two-group Bivariate AR(1) DSEM model

```
DATA:   FILE = ex9.32double.dat;

VARIABLE:
    NAMES = subject g y1 z1 y2 z2;
    usevar = y1-z2;
    CLUSTER = subject;
    LAGGED = y1-z2(1);
    missing=*;

ANALYSIS:       TYPE = TWOLEVEL; ESTIMATOR = BAYES;

MODEL:

        %WITHIN%
        y1 ON y1&1;
        z1 ON z1&1;
        y1 ON z1&1;
        z1 ON y1&1;
        y2 ON y2&1;
        z2 ON z2&1;
        y2 ON z2&1;
        z2 ON y2&1;
        y1-z1 with y2-z2@0;

        %BETWEEN%
        y1 WITH z1;
        y2 WITH z2;
```