Fitting MNLFA models in Mplus

This supplement to Bauer (2016) illustrates the fitting of MNLFA models in Mplus.[1] An earlier demonstration of how to fit MNLFAs in the NLMIXED procedure of SAS was provided in the supplementary materials of Bauer & Hussong (2009)[2]; however, our experience is that Mplus is much more computationally efficient for fitting these models. Here, we begin by describing how to fit a univariate (single-factor) MNLFA in Mplus, consistent with Step 2 of our recommended model-fitting sequence. We then proceed to show the extension to multidimensional MNFA in a two-factor model, consistent with Step 3. The input scripts shown here reference the delinquency data from Bauer (2016).

Unidimensional MNLFA

Example code for fitting an initial unidimensional MNLFA is shown below. This relatively simple model was fit to the violent delinquency items and included moderation of the factor mean and variance as a function of age and sex, but did not yet include differential item functioning (DIF). The majority of the syntax within the `TITLE`, `DATA`, `VARIABLE`, `DEFINE`, `ANALYSIS`, and `SAVEDATA` commands is standard and greater explanation of these commands can be found in the Mplus user guide.[3] Here we will elaborate primarily on options related specifically to fitting the MNLFA.

```
TITLE:
 MNLFA with factor mean and variance dependent on predictors;

DATA:
 FILE IS delinquincy.dat;

VARIABLE:
 NAMES ARE id school h1ds1-h1ds15 h1fv1-h1fv9 male ageyrs;
 CATEGORICAL ARE H1FV1-H1FV8 H1DS6 H1DS14;
 USEVARIABLES ARE  H1FV1-H1FV8 H1DS6 H1DS14
                   Male Agecent Agecent2 MaleAge MaleAge2;
 CONSTRAINT = Male Agecent Agecent2 MaleAge MaleAge2;
 MISSING = .
 IDVARIABLE IS id;
 CLUSTER IS school;

DEFINE:
 Agecent = Ageyrs - 15;
 Agecent2 = Agecent**2;
 MaleAge = Male*Agecent;
 MaleAge2 = Male*Agecent2;
```

[1] These instructions were developed using Mplus 7.3. Note that in some earlier versions of Mplus, nonlinear constraints on the model parameters as a function of exogenous predictors could not be combined with the use of categorical indicators, limiting the ability to fit MNLFAs within this software package. We thank Linda Muthén for her assistance in removing this restriction in current versions of Mplus, enabling the use of Mplus in fitting a wide variety of MNLFAs

[2] http://supp.apa.org/psycarticles/supplemental/met_14_2_101/met_bauer0079_supp.pdf

[3] http://statmodel.com/ugexcerpts.shtml

```
ANALYSIS:
 TYPE = COMPLEX;
 ESTIMATOR = MLR;
 LINK = LOGIT;

MODEL:
  VLNT BY H1FV1-H1FV8* H1DS6* H1DS14*;
  [VLNT@0];
  VLNT (v_vlnt);
  VLNT ON Agecent Agecent2 Male MaleAge MaleAge2;

MODEL CONSTRAINT:
 NEW (v_vlnt1*0 v_vlnt2*0 v_vlnt3*0 v_vlnt4*0 v_vlnt5*0);
 v_vlnt = EXP(v_vlnt1*Agecent + v_vlnt2*Agecent2 +
          v_vlnt3*Male + v_vlnt4*MaleAge + v_vlnt5*MaleAge2);

SAVEDATA:
 FILE IS
 C:\Users\dbauer\Desktop\AddHealth\MyStuff\scoreV.dat;
 SAVE=FSCORES;
```

For clarification, we will briefly describe a few options that are related to our specific data application.

First, two options were included to account for the fact that individuals are clustered within schools in our data, namely:

- `CLUSTER IS school;` (in the `DEFINE` section)
- `TYPE = COMPLEX;` (in the `ANALYSIS` section)

These commands would not be needed for an MNLFA conducted on a simple random sample. The additional option `ESTIMATOR = MLR;` in the `ANALYSIS` section requests robust standard errors to accompany the maximum likelihood estimates.

Second, two options are included because the delinquency items are binary:

- `CATEGORICAL ARE H1FV1-H1FV8 H1DS6 H1DS14;` (in the `DEFINE` section)
- `LINK = LOGIT;` (in the `ANALYSIS` section)

These options indicate that a logistic regression specification is to be used when relating the items to the factor.

Third, the options within the `DEFINE` command are used for data management purposes, creating new versions of the predictors:

- A centered version of the Age variable (`Agecent`), equal to zero at 15-years of age
- A squared Age variable (`Agecent2`), for the inclusion of quadratic effects
- Products of `Male` (coded 1 when Sex = Male; else 0) with `Agecent` (`MaleAge`) and `Agecent2` (`MaleAge2`), for the inclusion of linear and quadratic age interactions with sex.

Now we turn to the options and commands associated with fitting the MNLFA to this data. An important aspect of the MNLFA is that the values of basic model parameters, such as the factor mean and variance, and/or item intercepts and loadings, are moderated by (depend on) external predictors. For this model, we shall permit the factor mean and variance to depend on age and sex. In particular, using the centered version of age, our model includes a linear moderation function for the factor mean:

$$\alpha_i = \alpha_0 + \gamma_1 Age_i + \gamma_2 Age_i^2 + \gamma_3 Male_i + \gamma_4 Male_i \times Age_i + \gamma_4 Male_i \times Age_i^2 \qquad (1)$$

and a log-linear moderation function for the factor variance:

$$\psi_i = \psi_0 \exp\left(\beta_1 Age_i + \beta_2 Age_i^2 + \beta_3 Male_i + \beta_4 Male_i \times Age_i + \beta_4 Male_i \times Age_i^2\right) \qquad (2)$$

To identify the scale of the violent behavior latent factor, we shall set $\alpha_0 = 0$ and $\psi_0 = 1$, implying that the factor has a mean of zero and variance of one within the reference group (i.e., where all predictors are zero, in this case, for 15-year-old girls).

The linear moderation function for the violent behavior factor mean can be implemented simply by regressing the factor on the exogenous predictors, much like a MIMIC model specification. Hence, in the `MODEL` command we include the following line:

```
VLNT ON Agecent Agecent2 Male MaleAge MaleAge2;
```

This line implements Equation (1). The identification constraint $\alpha_0 = 0$ is set by the line `[VLNT@0];`

The log-linear moderation function for the variance of the violent behavior factor requires a few more steps, as its implementation makes use of the nonlinear constraints feature within Mplus.

- Within the `VARIABLE` command, we include the option

  ```
  CONSTRAINT = Male Agecent Agecent2 MaleAge MaleAge2;
  ```

  to indicate that these predictors will be used to moderate the variance function.

- Within the `MODEL` command, we specify the label `(v_vlnt)` for the factor variance within the line:

  ```
  VLNT (v_vlnt);
  ```

  That is, `v_vlnt` stands for ψ_i in Equation (2).

- We include the `MODEL CONSTRAINT` command and `NEW` option:

  ```
  MODEL CONSTRAINT:
   NEW (v_vlnt1*0 v_vlnt2*0 v_vlnt3*0 v_vlnt4*0 v_vlnt5*0);
  ```

  `NEW` is used to define each unique parameter within Equation 2. Thus, `v_vlnt1` stands for β_1, `v_vlnt2` stands for β_2, etc. The parameter ψ_0 is not included as it is fixed to 1 to set the scale of the latent factor and will not be estimated. For each parameter to be estimated a start value of zero has been assigned (`*0`). Use of informative start values can speed model convergence, and in practice we usually use the final estimates from simpler models as the start values for more complex models. As this is the first model fit to this data, we simply used start values of zero for each parameter.

- Within the `MODEL CONSTRAINT` command, the line

```
v_vlnt = EXP(v_vlnt1*Agecent + v_vlnt2*Agecent2 +
          v_vlnt3*Male + v_vlnt4*MaleAge + v_vlnt5*MaleAge2);
```

  directly expresses Equation (2), with the implicit identification constraint that $\psi_0 = 1$.

This completes the specification of the model.

Let us now consider a more complex model, which also allows for DIF for one item. For this example we shall allow DIF on item FV2: “Pulled a weapon on you”, as allowing for DIF in this item resulted in the largest improvement in model fit. The input script is shown below.

```
TITLE:
 MNLFA for violent factor where one item has DIF;

DATA:
 FILE IS delinquincy.dat;

VARIABLE:
 NAMES ARE id school h1ds1-h1ds15 h1fv1-h1fv9 male ageyrs;
 CATEGORICAL ARE H1FV1-H1FV8 H1DS6 H1DS14;
 USEVARIABLES ARE  H1FV1-H1FV8 H1DS6 H1DS14
                   Male Agecent Agecent2 MaleAge MaleAge2;
 CONSTRAINT = Male Agecent Agecent2 MaleAge MaleAge2;
 MISSING = .
 IDVARIABLE IS id;
 CLUSTER IS school;

DEFINE:
 Agecent = Ageyrs - 15;
 Agecent2 = Agecent**2;
 MaleAge = Male*Agecent;
 MaleAge2 = Male*Agecent2;

ANALYSIS:
 TYPE = COMPLEX;
 ESTIMATOR = MLR;
 LINK = LOGIT;

MODEL:
  VLNT BY H1FV1*1.863 H1FV2*2.397 H1FV3*1.644
          H1FV4*2.398 H1FV5*2.690 H1FV6*2.263
          H1FV7*3.140 H1FV8*4.111 H1DS6*2.284 H1DS14*1.972;
  [H1FV1$1*3.646 H1FV2$1*4.131 H1FV3$1*6.028
   H1FV4$1*5.810 H1FV5$1*2.371 H1FV6$1*4.224
   H1FV7$1*6.884 H1FV8$1*10.396 H1DS6$1*3.351 H1DS14$1*2.877];
  [VLNT@0];
  VLNT (v_vlnt);
```

```
  VLNT ON Agecent*-.067 Agecent2*-.018 Male*.597 MaleAge*.043
         MaleAge2*.015;

  VLNT BY H1FV2 (L);
  H1FV2 ON Agecent Agecent2 Male MaleAge MaleAge2;

MODEL CONSTRAINT:
 NEW (v_vlnt1*-.013 v_vlnt2*-.056 v_vlnt3*-.195 v_vlnt4*.100
      v_vlnt5*.022 L0*2 L1*0 L2*0 L3*0 L4*0 L5*0);

 v_vlnt = EXP(v_vlnt1*Agecent + v_vlnt2*Agecent2 +
          v_vlnt3*Male + v_vlnt4*MaleAge + v_vlnt5*MaleAge2);

 L = L0 + L1*Agecent + L2*Agecent2 + L3*Male + L4*MaleAge +
     L5*MaleAge2;
```

In this model we retain the factor mean and variance specification of Equations (1) and (2) and, to speed estimation, we have used estimates from the prior model as start values for the corresponding estimates in this model (following the `*` symbols). The current model additionally includes DIF via linear moderation of both the intercept and loading of item FV2. For the intercept, we thus implement the following moderation equation:

$$\begin{aligned}\nu_{(\mathrm{FV2})i} = \nu_{(\mathrm{FV2})0} + \kappa_{(\mathrm{FV2})1} Age_i + \kappa_{(\mathrm{FV2})2} Age_i^2 + \kappa_{(\mathrm{FV2})3} Male_i + \\ \kappa_{(\mathrm{FV2})4} Male_i \times Age_i + \kappa_{(\mathrm{FV2})5} Male_i \times Age_i^2\end{aligned} \tag{3}$$

And, likewise, for the factor loading we specify:

$$\begin{aligned}\lambda_{(\mathrm{FV2})i} = \lambda_{(\mathrm{FV2})0} + \omega_{(\mathrm{FV2})1} Age_i + \omega_{(\mathrm{FV2})2} Age_i^2 + \omega_{(\mathrm{FV2})3} Male_i + \\ \omega_{(\mathrm{FV2})4} Male_i \times Age_i + \omega_{(\mathrm{FV2})5} Male_i \times Age_i^2\end{aligned} \tag{4}$$

As in a MIMIC-type model, the linear moderation function for the intercept can be implemented simply by regressing the item on the exogenous predictors.[4] Hence, in the `MODEL` command we include the following line:

```
H1FV2 ON Agecent Agecent2 Male MaleAge MaleAge2;
```

This line implements Equation (3).

[4] Note, however, that for binary items Mplus uses a slightly different parameterization in which a threshold parameter is estimated and the baseline intercept $\nu_{(\mathrm{FV2})0}$ is set to zero. The Mplus parameterization is equivalent to the one presented here; the only practical consequence is that to obtain the estimate of $\nu_{(\mathrm{FV2})0}$ one must multiply the obtained threshold estimate by -1.

To implement moderation of the factor loading, we again make use of nonlinear constraints:

- Within the `MODEL` command, we specify the label `(L)` for the factor loading within the line:

  ```
  VLNT BY H1FV2 (L);
  ```

  That is, `L` stands for $\lambda_{(\text{FV2})i}$ in Equation (4).

- Within the `NEW` option of the `MODEL CONSTRAINT` command we now indicate the additional new parameters associated with moderation of the factor loading:

  ```
  MODEL CONSTRAINT:
   NEW (v_vlnt1*-.013 v_vlnt2*-.056 v_vlnt3*-.195 v_vlnt4*.100
        v_vlnt5*.022 L0*2 L1*0 L2*0 L3*0 L4*0 L5*0);
  ```

  Here, `L0` corresponds to $\lambda_{(\text{FV2})0}$, and `L1`, `L2`, …, `L5` correspond to $\omega_{(\text{FV2})1}, \omega_{(\text{FV2})2}, \ldots, \omega_{(\text{FV2})5}$ from Equation (4).

- Within the `MODEL CONSTRAINT` command, the line

  ```
  L = L0 + L1*Agecent + L2*Agecent2 + L3*Male + L4*MaleAge +
       L5*MaleAge2;
  ```

  directly expresses Equation (4).

This completes the specification of the model. Simplification or elaboration to trim terms or allow DIF for multiple items is straightforward.

Multidimensional MNLFA

We now show the input script associated with the two-factor final model fit to the delinquency data. This script is necessarily more complex, including moderation of the mean and variance of two factors (non-violent and violent delinquency), DIF on multiple items simultaneously, and moderation of the factor covariance. Again, to speed convergence, start values were included based on prior model estimates.

```
TITLE:
 Two-factor final MNLFA;

DATA:
 FILE IS delinquincy.dat;

VARIABLE:
 NAMES ARE id school h1ds1-h1ds15 h1fv1-h1fv9 male ageyrs;
 CATEGORICAL ARE h1ds1-h1ds15 h1fv1-h1fv8;
 USEVARIABLES ARE h1ds1-h1ds3 h1ds6 h1ds8-h1ds10 h1ds13-h1ds15 h1fv1-
                 h1fv8 Male Agecent Agecent2 MaleAge MaleAge2;
 CONSTRAINT = Male Agecent Agecent2 MaleAge MaleAge2;
 MISSING = .
 IDVARIABLE IS id;
 CLUSTER IS school;

DEFINE:
```

```
 Agecent = Ageyrs - 15;
 Agecent2 = Agecent**2;
 MaleAge = Male*Agecent;
 MaleAge2 = Male*Agecent2;

ANALYSIS:
 TYPE = COMPLEX;
 ESTIMATOR = MLR;
 LINK = LOGIT;

MODEL:
  NVLNT BY H1DS1*1.796 H1DS9*2.052 H1DS10*2.282 H1DS13*1.859
           H1DS15*1.255;
  [H1DS1$1*3.733 H1DS9$1*4.992 H1DS10$1*5.215 H1DS13$1*2.459];

  VLNT BY H1FV4*2.387 H1FV5*2.625 H1FV6*2.254 H1FV7*3.135 H1FV8*4.140;
  [H1FV4$1*5.850 H1FV5$1*2.388 H1FV6$1*4.271 H1FV7$1*6.954
   H1FV8$1*10.529];
  [VLNT@0 NVLNT@0];

  VLNT (v_vlnt);
  NVLNT (v_nvlnt);
  VLNT with NVLNT (cov);

  NVLNT ON Agecent*-.078 Agecent2*-.060 Male*.330 MaleAge*.089;
  VLNT ON Agecent*-.089 Agecent2*-.032 Male*.663 MaleAge*.025
          MaleAge2*.022;

!NVLNT DIF ITEMS;

  NVLNT BY H1DS2 (LDS2);
  H1DS2 ON Agecent*-.099 Male*-.690;
  [H1DS2$1*2.826];

  NVLNT BY H1DS3*1.363;
  H1DS3 ON Agecent*.326 Agecent2*-.042 Male*-.866 MaleAge*-.258;
  [H1DS3$1*-.685];

  NVLNT BY H1DS8*1.279;
  H1DS8 ON Agecent*.108 Agecent2*-.059;
  [H1DS8$1*2.917];

  NVLNT BY H1DS15*1.255;
  H1DS15 ON Male*-.477;
  [H1DS15$1*-.081];

!VLNT DIF ITEMS;

  VLNT BY H1FV1 (LFV1);
  H1FV1 ON Agecent*.126 Agecent2*.086 Male*-.839;
  [H1FV1$1*3.722];
```

```
  VLNT BY H1FV2 (LFV2);
  H1FV2 ON Agecent*.341 Agecent2*.114;
  [H1FV2$1*4.588];

  VLNT BY H1FV3 (LFV3);
  H1FV3 ON Agecent*.553 Agecent2*.122;
  [H1FV3$1*6.750];

  VLNT BY H1DS6 (LDS6);
  H1DS6 ON Agecent*.004 Male*.282 MaleAge*.173;
  [H1DS6$1*3.416];

  VLNT BY H1DS14 (LDS14);
  H1DS14 ON Agecent*-.076 Male*-1.294 MaleAge*.578;
  [H1DS14$1*2.544];

MODEL CONSTRAINT:
 NEW (v_vlnt1*.013 v_vlnt3*-.281
      v_nvlnt1*-.050 v_nvlnt3*.135
      z0*.865 z1*-.051 z3*-.206
      LDS2_0*1.894 LDS2_3*1.256
      LFV1_0*2.376 LFV1_1*.055 LFV1_2*-.078
      LFV2_0*2.721 LFV2_1*-.050 LFV2_2*-.092
      LFV3_0*1.808 LFV3_1*-.341
      LDS6_0*2.195 LDS6_1*-.150
      LDS14_0*1.756 LDS14_1*-.089 LDS14_3*.814 LDS14_4*-.329);

 LDS2 = LDS2_0 + LDS2_3*Male;

 LFV1 = LFV1_0 + LFV1_1*Agecent + LFV1_2*Agecent2;

 LFV2 = LFV2_0 + LFV2_1*Agecent + LFV2_2*Agecent2;

 LFV3 = LFV3_0 + LFV3_1*Agecent;

 LDS6 = LDS6_0 + LDS6_1*Agecent;

 LDS14 = LDS14_0 + LDS14_1*Agecent + LDS14_3*Male +
         LDS14_4*MaleAge;

 v_nvlnt = EXP(v_nvlnt1*Agecent + v_nvlnt3*Male);

 v_vlnt = EXP(v_vlnt1*Agecent + v_vlnt3*Male);

 cov = SQRT(EXP(v_vlnt1*Agecent + v_vlnt3*Male))*
       SQRT(EXP(v_nvlnt1*Agecent + v_nvlnt3*Male))*
       (EXP(2*(z0 + z1*Agecent + z3*Male))-1)/
       (EXP(2*(z0 + z1*Agecent + z3*Male))+1);
```

Note that, although this script is considerably longer, all of the features within this code represent straightforward extensions from the prior examples. Moderation of the factor means is accomplished by regressing the factors on the covariates within the `MODEL` command:

```
NVLNT ON Agecent*-.078 Agecent2*-.060 Male*.330 MaleAge*.089;
VLNT ON Agecent*-.089 Agecent2*-.032 Male*.663 MaleAge*.025
        MaleAge2*.022;
```

Moderation of the factor variances is accomplished through the `MODEL CONSTRAINTS` command:

```
v_nvlnt = EXP(v_nvlnt1*Agecent + v_nvlnt3*Male);
v_vlnt = EXP(v_vlnt1*Agecent + v_vlnt3*Male);
```

Likewise, intercept DIF is accomplished by regressing items directly on the predictors within the `MODEL` command, for example:

```
H1DS2 ON Agecent*-.099 Male*-.690;
H1DS3 ON Agecent*.326 Agecent2*-.042 Male*-.866 MaleAge*-.258;
```

And moderation of the factor loadings is incorporated via the `MODEL CONSTRAINTS` command, for example:

```
LDS2 = LDS2_0 + LDS2_3*Male;
LFV1 = LFV1_0 + LFV1_1*Agecent + LFV1_2*Agecent2;
```

The only feature in this script which we have not seen in prior models is the moderation of the factor covariance, incorporated via the lines (within `MODEL` and `MODEL CONSTRAINTS`, respectively):

```
VLNT with NVLNT (cov);
```

```
cov = SQRT(EXP(v_vlnt1*Agecent + v_vlnt3*Male))*
      SQRT(EXP(v_nvlnt1*Agecent + v_nvlnt3*Male))*
      (EXP(2*(z0 + z1*Agecent + z3*Male))-1)/
      (EXP(2*(z0 + z1*Agecent + z3*Male))+1);
```

Although this expression may appear complicated, it simply implements in one step the sequence of equations within the appendix of Bauer (2016). Specifically, we define the moderation function for Fisher's z to be

$$\zeta_i = \zeta_0 + \upsilon_1 Age_i + \upsilon_3 Male_i \tag{5}$$

where ζ_0, υ_1, and υ_3 are represented by `z0`, `z1`, and `z3` in the Mplus script. Then, we invert Fisher's z transformation to obtain the implied factor correlation:

$$\rho_i = \frac{\exp\left[2\left(\zeta_0 + \upsilon_1 Age_i + \upsilon_3 Male_i\right) - 1\right]}{\exp\left[2\left(\zeta_0 + \upsilon_1 Age_i + \upsilon_3 Male_i\right) + 1\right]} \tag{6}$$

Note that the last two lines of the covariance specification in the Mplus script directly map onto Equation (6). Finally, to obtain the covariance we simply multiply the correlation by the square roots of the factor variances (i.e., standard deviations), as shown in the top two lines of the covariance specification in the Mplus script. These lines include the same expressions that were used to specify the factor variances.

References

Bauer, D.J. (2016). A more general model for testing measurement invariance and differential item functioning. *Psychological Methods*.

Bauer, D.J., & Hussong, A.M. (2009). Psychometric approaches for developing commensurate measures across independent studies: traditional and new models. *Psychological Methods*, *14*, 101-125.