Mplus estimators: MLM and MLR

Yves Rosseel
Department of Data Analysis
Ghent University

First Mplus User meeting – October 27th 2010
Utrecht University, the Netherlands

(with a few corrections, 10 July 2017)

NOTE: 2 corrections on slide 2 and 7 (and again on slide 18):

- slide 2 (and 18):

$$W = 2D'(\hat{\Sigma}^{-1} \otimes \hat{\Sigma}^{-1})D$$

  should be

$$W = \frac{1}{2}D'(\hat{\Sigma}^{-1} \otimes \hat{\Sigma}^{-1})D$$

- slide 7 (and 18):

$$U = (W^{-1} - W^{-1}\Delta(\Delta' W^{-1}\Delta)^{-1}\Delta' W^{-1})$$

  should be

$$U = (W - W\Delta(\Delta' W\Delta)^{-1}\Delta' W)$$

Estimator: ML

- default estimator for many model types in Mplus
- likelihood function is derived from the multivariate normal distribution
- standard errors are based on the covariance matrix that is obtained by inverting the information matrix
- in Mplus versions 1–4, the default was to use the *expected* information matrix:

$$\begin{aligned} n\text{Cov}(\hat{\theta}) &= A^{-1} \\ &= (\Delta' W \Delta)^{-1} \end{aligned}$$

  - Δ is a jacobian matrix and W is a function of Σ^{-1}
  - if no meanstructure:

$$\begin{aligned} \Delta &= \partial\hat{\Sigma}/\partial\hat{\theta}' \\ W &= \frac{1}{2} D'(\hat{\Sigma}^{-1} \otimes \hat{\Sigma}^{-1}) D \end{aligned}$$

- in Mplus versions 5–6, the default is the *observed* information matrix (because the default: `TYPE=GENERAL MISSING H1`):

$$\begin{aligned} n\text{Cov}(\hat{\theta}) &= A^{-1} \\ &= [-\text{Hessian}]^{-1} \\ &= \left[-\partial F(\hat{\theta})/(\partial\hat{\theta}\partial\hat{\theta}')\right]^{-1} \end{aligned}$$

  where $F(\theta)$ is the function that is minimized

- overall model evaluation is based on the likelihood-ratio (LR) statistic (chi-square test): T_{ML}
  - (minus two times the) difference between loglikelihood of user-specified model H_0 and unrestricted model H_1
  - equals (in Mplus) $2 \times n$ times the minimum value of $F(\theta)$
  - test statistics follows (under regularity conditions) a chi-square distribution
  - Mplus calls this the “Chi-Square Test of Model Fit”

What if the data are NOT normally distributed?

- in the real world, data may never be normally distributed
- two types:
  - categorical and/or limited-dependent outcomes: binary, ordinal, nominal, counts, censored (WLSMV, logit/probit)
  - continuous outcomes, not normally distributed: skewed, too flat/too peaked (kurtosis), ...
- in many situations, the ML parameter estimates are still consistent (if the model is identified and correctly specified)
- in fewer situations, the ML procedure can still provide reliable inference (SE's and test statistics), but it is hard to identify these conditions empirically
- in practice, we may prefer *robust* procedures

Three classes of robust procedures in the SEM literature

1. ML estimation with 'robust' standard errors, and a 'robust' test statistic for model evaluation
   - bootstrapped SE's, and bootstrapped test statistic
   - Satorra-Bentler corrections (Mplus: estimator=MLM)
   - Huber/Pseudo ML/sandwich corrections (Mplus: estimator=MLR)
2. GLS (Mplus: estimator=WLS) with a weight matrix (Γ) based on the 4th-order moments of the data
   - Asymptotically Distribution Free (ADF) estimation (Browne, 1984)
   - only works well with large/huge sample sizes
3. case-robust or outlier-robust methods: cases lying far from the center of the data cloud receive smaller weights, affecting parameter estimates, SE's and model evaluation
   - only available in EQS(?)

Estimator MLM

- Mplus 6 User's Guide page 533:

  > "MLM – maximum likelihood parameter estimates with standard errors and a mean-adjusted chi-square test statistic that are robust to non-normality. The MLM chi-square test statistic is also referred to as the Satorra-Bentler chi-square."

- parameter estimates are standard ML estimates
- standard errors are robust to non-normality
  - standard errors are computed using a sandwich-type estimator:

$$\begin{aligned} n\text{Cov}(\hat{\theta}) &= A^{-1}BA^{-1} \\ &= (\Delta' W \Delta)^{-1}(\Delta' W \Gamma W \Delta)(\Delta' W \Delta)^{-1} \end{aligned}$$

  - A is usually the expected information matrix (but not in Mplus)
  - references: Huber (1967), Browne (1984), Shapiro (1983), Bentler (1983), ...

- chi-square test statistic is robust to non-normality
  - test statistic is 'scaled' by a correction factor

$$T_{SB} = T_{ML}/c$$

  - the scaling factor c is computed by:

$$c = tr\left[U\Gamma\right]/\mathrm{df}$$

    where

$$U = (W - W\Delta(\Delta' W \Delta)^{-1}\Delta' W)$$

  - correction method described by Satorra & Bentler (1986, 1988, 1994)
- estimator MLM: for complete data only

  (DATA: LISTWISE=ON)

Estimator MLR

- Mplus 6 User's Guide page 533:

> MLR – maximum likelihood parameter estimates with standard errors and a chi-square test statistic (when applicable) that are robust to non-normality and non-independence of observations when used with TYPE=COMPLEX. The MLR standard errors are computed using a sandwich estimator. The MLR chi-square test statistic is asymptotically equivalent to the Yuan-Bentler T2* test statistic.

- parameter estimates are standard ML estimates

- standard errors are robust to non-normality
  - standard errors are computed using a (different) sandwich approach:

$$\begin{aligned} n\text{Cov}(\hat{\theta}) &= A^{-1} B A^{-1} \\ &= A_0^{-1} B_0 A_0^{-1} = C_0 \end{aligned}$$

where

$$A_0 = -\sum_{i=1}^{n} \frac{\partial \log L_i}{\partial \hat{\theta}\, \partial \hat{\theta}'} \quad \text{(observed information)}$$

and

$$B_0 = \sum_{i=1}^{n} \left(\frac{\partial \log L_i}{\partial \hat{\theta}} \right) \times \left(\frac{\partial \log L_i}{\partial \hat{\theta}} \right)'$$

  - for both complete and incomplete data
  - Huber (1967), Gourieroux, Monfort & Trognon (1984), Arminger & Schoenberg (1989)

- chi-square test statistic is robust to non-normality
  - test statistic is 'scaled' by a correction factor

$$T_{MLR} = T_{ML}/c$$

  - the scaling factor c is (usually) computed by

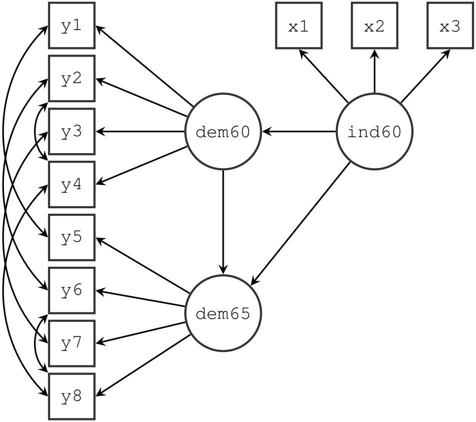
$$c = tr\,[M]$$

    where

$$M = C_1(A_1 - A_1\Delta(\Delta' A_1 \Delta)^{-1}\Delta' A_1)$$

  - A_1 and C_1 are computed under the unrestricted (H_1) model
  - correction method described by Yuan & Bentler (2000)
- information matrix (A) can be observed or expected
- for complete data, the MLR and MLM corrections are asymptotically equivalent

Example: Industrialization and Political Democracy (Bollen, 1989) N=75



Output

```
Title:      Example from Bollen (1989)
            Industrialization and
            Political Democracy
Data:       File  = democindus.txt;
            Type  = individual;
Variable:   Names = y1 y2 y3 y4 y5 y6
                    y7 y8 x1 x2 x3;
Analysis:   Estimator = ML;
           !Estimator = MLM;
           !Estimator = MLR;
           !Information = expected;
           !Information = observed;
Model:      ind60 by x1 x2 x3;
            dem60 by y1 y2 y3 y4;
            dem65 by y5 y6 y7 y8;

            dem60 on ind60;
            dem65 on ind60 dem60;

            y1 y2 y3 y4 y2 y6 pwith
            y5 y6 y7 y8 y4 y8;
```

Mplus 6.1 output: estimator = ML, information = observed

Output

```
Chi-Square Test of Model Fit

         Value                                 38.125
         Degrees of Freedom                        35
         P-Value                               0.3292


                                                     Two-Tailed
                  Estimate       S.E.  Est./S.E.    P-Value

 IND60    BY
    X1               1.000      0.000    999.000    999.000
    X2               2.180      0.139     15.685      0.000
    X3               1.819      0.152     11.949      0.000
 DEM60    BY
    Y1               1.000      0.000    999.000    999.000
    Y2               1.257      0.185      6.775      0.000
    Y3               1.058      0.148      7.131      0.000
    Y4               1.265      0.151      8.391      0.000
 DEM65    BY
    Y5               1.000      0.000    999.000    999.000
    Y6               1.186      0.171      6.920      0.000
    Y7               1.280      0.160      7.978      0.000
    Y8               1.266      0.163      7.756      0.000
```

Mplus 6.1 output: estimator = ML, information = expected

Output

```
Chi-Square Test of Model Fit

          Value                                   38.125
          Degrees of Freedom                          35
          P-Value                                 0.3292


                                                             Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 IND60     BY
    X1                 1.000      0.000    999.000    999.000
    X2                 2.180      0.139     15.742      0.000
    X3                 1.819      0.152     11.967      0.000
 DEM60     BY
    Y1                 1.000      0.000    999.000    999.000
    Y2                 1.257      0.182      6.889      0.000
    Y3                 1.058      0.151      6.987      0.000
    Y4                 1.265      0.145      8.722      0.000
 DEM65     BY
    Y5                 1.000      0.000    999.000    999.000
    Y6                 1.186      0.169      7.024      0.000
    Y7                 1.280      0.160      8.002      0.000
    Y8                 1.266      0.158      8.007      0.000
```

Mplus output: estimator = MLM, information = expected

Output

```
Chi-Square Test of Model Fit

         Value                              40.536*
         Degrees of Freedom                     35
         P-Value                            0.2393
         Scaling Correction Factor           0.941
           for MLM
                                                       Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 IND60    BY
    X1                 1.000      0.000    999.000    999.000
    X2                 2.180      0.126     17.251      0.000
    X3                 1.819      0.128     14.212      0.000
 DEM60    BY
    Y1                 1.000      0.000    999.000    999.000
    Y2                 1.257      0.137      9.193      0.000
    Y3                 1.058      0.133      7.971      0.000
    Y4                 1.265      0.119     10.585      0.000
 DEM65    BY
    Y5                 1.000      0.000    999.000    999.000
    Y6                 1.186      0.171      6.947      0.000
    Y7                 1.280      0.166      7.706      0.000
    Y8                 1.266      0.174      7.289      0.000
```

Mplus output: estimator = MLR, information = observed

Output

```
Chi-Square Test of Model Fit

         Value                              41.401*
         Degrees of Freedom                     35
         P-Value                            0.2114
         Scaling Correction Factor           0.921
           for MLR
                                                   Two-Tailed
                   Estimate       S.E.  Est./S.E.    P-Value

 IND60     BY
    X1                1.000      0.000    999.000    999.000
    X2                2.180      0.145     15.044      0.000
    X3                1.819      0.140     12.950      0.000
 DEM60     BY
    Y1                1.000      0.000    999.000    999.000
    Y2                1.257      0.150      8.392      0.000
    Y3                1.058      0.130      8.107      0.000
    Y4                1.265      0.146      8.661      0.000
 DEM65     BY
    Y5                1.000      0.000    999.000    999.000
    Y6                1.186      0.181      6.541      0.000
    Y7                1.280      0.173      7.415      0.000
    Y8                1.266      0.189      6.685      0.000
```

Mplus output: estimator = MLR, information = expected

Output

```
Chi-Square Test of Model Fit

        Value                                  40.936*
        Degrees of Freedom                         35
        P-Value                                0.2261
        Scaling Correction Factor               0.931
          for MLR
                                                         Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 IND60    BY
    X1                 1.000      0.000    999.000    999.000
    X2                 2.180      0.144     15.185      0.000
    X3                 1.819      0.139     13.070      0.000
 DEM60    BY
    Y1                 1.000      0.000    999.000    999.000
    Y2                 1.257      0.140      8.964      0.000
    Y3                 1.058      0.134      7.882      0.000
    Y4                 1.265      0.127      9.972      0.000
 DEM65    BY
    Y5                 1.000      0.000    999.000    999.000
    Y6                 1.186      0.171      6.926      0.000
    Y7                 1.280      0.166      7.694      0.000
    Y8                 1.266      0.171      7.399      0.000
```

What about other software: EQS

- main developer: Peter Bentler
- in EQS 6.1, you can request robust SE's and test statistics `(METHOD=ML,ROBUST)`
- you can switch between the observed and expected information matrix `(SE=OBS, or SE=FISHER)`
- if the data is complete, you get:
  - robust standard errors
  - a Satorra-Bentler-scaled chi-square statistic
  - should be similar to Mplus estimator MLM
- if the data is incomplete, you get:
  - robust standard errors
  - Yuan-Bentler-scaled chi-square statistic
  - should be similar to Mplus estimator MLR

1. Mplus estimator MLM is NOT identical to EQS

- both SE's and the value of the Satorra-Bentler scaled test statistic are computed differently
- formula standard errors:

$$n\text{Cov}(\hat{\theta}) = (\Delta' W \Delta)^{-1} (\Delta' W \Gamma W \Delta)(\Delta' W \Delta)^{-1}$$

- formula Satorra-Bentler scaling factor:

$$c = tr\left[U\Gamma\right]/\text{df}$$

where

$$U = (W - W\Delta(\Delta' W \Delta)^{-1}\Delta' W)$$

- however (if no meanstructure):

$$W = \frac{1}{2} D'(\hat{\Sigma}^{-1} \otimes \hat{\Sigma}^{-1}) D \quad \text{(EQS)}$$

$$W = \frac{1}{2} D'(S^{-1} \otimes S^{-1}) D \quad \text{(MPLUS)}$$

Satorra-Bentler: Mplus variant

Output

```
Lavaan (0.4-4) converged normally after 95 iterations

  Estimator                                         ML       Robust
  Minimum Function Chi-square                   38.125       40.536
  Degrees of freedom                                35           35
  P-value                                        0.329        0.239
  Scaling correction factor                                   0.941
    for the Satorra-Bentler correction (Mplus variant)

                   Estimate  Std.err  Z-value  P(>|z|)
Latent variables:
  ind60 =~
    x1                1.000
    x2                2.180    0.126   17.251    0.000
    x3                1.819    0.128   14.212    0.000
  dem60 =~
    y1                1.000
    y2                1.257    0.137    9.193    0.000
    y3                1.058    0.133    7.971    0.000
    y4                1.265    0.119   10.585    0.000
  dem65 =~
    y5                1.000
    y6                1.186    0.171    6.947    0.000
    y7                1.280    0.166    7.706    0.000
    y8                1.266    0.174    7.289    0.000
```

Satorra-Bentler: EQS variant

Output

```
Lavaan (0.4-4) converged normally after 91 iterations

  Estimator                                         ML       Robust
  Minimum Function Chi-square                   37.617       40.512
  Degrees of freedom                                35           35
  P-value                                        0.350        0.240
  Scaling correction factor                                   0.929
    for the Satorra-Bentler correction

                   Estimate  Std.err  Z-value  P(>|z|)
Latent variables:
  ind60 =~
    x1                1.000
    x2                2.180    0.143   15.287    0.000
    x3                1.819    0.138   13.158    0.000
  dem60 =~
    y1                1.000
    y2                1.257    0.139    9.024    0.000
    y3                1.058    0.133    7.936    0.000
    y4                1.265    0.126   10.039    0.000
  dem65 =~
    y5                1.000
    y6                1.186    0.170    6.973    0.000
    y7                1.280    0.165    7.746    0.000
    y8                1.266    0.170    7.448    0.000
```

Satorra-Bentler in Mplus: use MLR with expected information

Output

```
Lavaan (0.4-4) converged normally after 95 iterations

  Estimator                                         ML      Robust
  Minimum Function Chi-square                   38.125      40.936
  Degrees of freedom                                35          35
  P-value                                        0.329       0.226
  Scaling correction factor                                  0.931
    for the MLR correction

                   Estimate  Std.err  Z-value  P(>|z|)
Latent variables:
  ind60 =~
    x1                1.000
    x2                2.180    0.144   15.185    0.000
    x3                1.819    0.139   13.070    0.000
  dem60 =~
    y1                1.000
    y2                1.257    0.140    8.964    0.000
    y3                1.058    0.134    7.882    0.000
    y4                1.265    0.127    9.972    0.000
  dem65 =~
    y5                1.000
    y6                1.186    0.171    6.926    0.000
    y7                1.280    0.166    7.694    0.000
    y8                1.266    0.171    7.399    0.000
```

2. Mplus estimator MLR is NOT identical to EQS

- SE's are ok
- the value of the Yuan-Bentler scaled test statistic is computed differently
- formula Yuan-Bentler scaling factor:

$$c = tr\left[M\right]$$

where

$$M = C_1(A_1 - A_1\Delta(\Delta' A_1 \Delta)^{-1}\Delta' A_1)$$

- however, Mplus uses:

$$\text{tr}(M) = tr(B_1 A_1^{-1}) - tr(B_0 A_0^{-1})$$

- but: asymptotically equivalent

- Mplus 6 User's Guide page 533:

> MLR – maximum likelihood parameter estimates with standard errors and a chi-square test statistic (when applicable) that are robust to non-normality and non-independence of observations when used with TYPE=COMPLEX. The MLR standard errors are computed using a sandwich estimator. The MLR chi-square test statistic is asymptotically equivalent to the Yuan-Bentler T2* test statistic.

- Mplus 3 User's Guide page 401:

> MLR – maximum likelihood parameter estimates with standard errors and a chi-square test statistic (when applicable) that are robust to non-normality and non-independence of observations when used with TYPE=COMPLEX. The MLR standard errors are computed using a sandwich estimator. The MLR chi-square test statistic is also referred to as the Yuan-Bentler T2* test statistic.

Conclusions

- robust estimators in Mplus are similar, but not identical to EQS
- Mplus is not ‘wrong’, but uses different formulas
- in small samples, the differences can be substantial
- Mplus Users should be aware of the differences
- statisticians should study the differences
- we need open-source software (hint: `http://lavaan.org`)

Thank you for your attention