# Bayesian Analysis of Multiple Indicator Growth Modeling using Random Measurement Parameters Varying Across Time and Person

Tihomir Asparouhov & Bengt Muthén

Mplus
www.statmodel.com

# Overview

- New multilevel modeling features introduced in Mplus 7 based on Bayesian estimation
    - Random factor loadings for Two-Level SEM
    - Cross-Classified SEM
    - Random factor loadings for Cross-Classified SEM
    - Random effect interaction for Cross-Classified SEM
    - Model comparison using DIC, testing for random effect zero variance
- Applications to longitudinal analysis

# Standard Twolevel SEM

$Y_{ij}$ is a multivariate response vector for individual $i$ in cluster $j$.

$$Y_{ij} = Y_{1ij} + Y_{2j}$$

$$Y_{1ij} = \Lambda_1 \eta_{ij} + \varepsilon_{ij} \tag{1a}$$

$$\eta_{ij} = B_1 \eta_{ij} + \Gamma_{1j} x_{ij} + \xi_{ij} \tag{1b}$$

$$Y_{2j} = \nu_2 + \Lambda_2 \eta_j + \varepsilon_j \tag{2a}$$

$$\eta_j = \alpha_2 + B_2 \eta_j + \Gamma_2 x_j + \xi_j \tag{2b}$$

- ML estimation
- Random intercepts: $Y_{2j}$, $\varepsilon_j$, $\xi_j$
- Random slopes $\Gamma_{1j}$ for observed covariate are between level latent variables and are a part of the vector $\eta_j$
- Factor loadings are not random, the same across clusters.

# Twolevel SEM With Random Factor Loadings

$Y_{ij}$ is a multivariate response vector for individual $i$ in cluster $j$.

$$Y_{ij} = Y_{1ij} + Y_{2j}$$

$$Y_{1ij} = \Lambda_{1j}\eta_{ij} + \varepsilon_{ij} \tag{1a}$$

$$\eta_{ij} = \eta_{1ij} + \eta_{2j}$$

$$\eta_{1ij} = B_{1j}\eta_{1ij} + \Gamma_{1j}\, x_{ij} + \xi_{ij} \tag{1b}$$

$$Y_{2j} = \nu_2 + \Lambda_2\eta_j + \varepsilon_j \tag{2a}$$

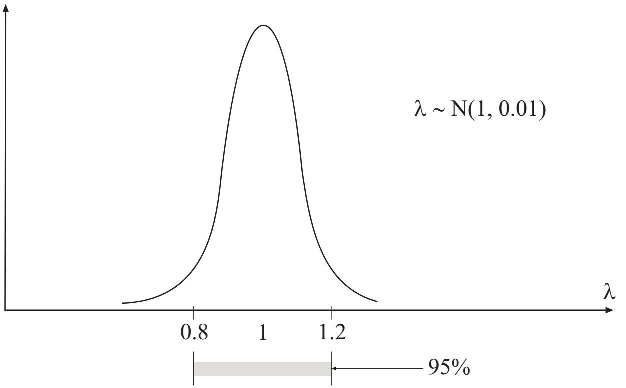
$$\eta_j = \alpha_2 + B_2\eta_j + \Gamma_2 x_j + \xi_j \tag{2b}$$

- Bayes estimation
- Random intercepts and slopes
- Random factor loadings $\Lambda_{1j}$, random structural parameters $B_{1j}$ and $\eta_{2j}$ are between level latent variables and are a part of the vector $\eta_j$
- Allows group specific SEM fit for more accurate and detailed measurement model.

# Two-Level Analysis with Random Loadings: A New Conceptualization of Measurement Invariance

Each measurement parameter varies across groups/clusters, but groups/clusters have a common mean and variance. E.g.

$$\lambda_j \sim N(\mu_\lambda, \sigma^2_\lambda). \quad (1)$$



# Random Factor Loadings: Example 1

$$Y_{ij} = \nu + \lambda_j \eta_{ij} + \varepsilon_j + \varepsilon_{ij} \quad (2)$$

$$\eta_{ij} = \xi_j + \xi_{ij} \quad (3)$$

$$\lambda_j = \lambda + \lambda_{0,j} \quad (4)$$

- $\varepsilon_j$ group specific deviation for the intercept $\nu$
- $\lambda_{0,j}$ group specific deviation for the loadings $\lambda$
- $\xi_j$ group specific deviation for the factor mean
- IRT-parametrization, Fox 2010, the same factor loading applies to the within and between part of the factor

# Random Factor Loadings: Example 2

$$Y_{ij} = \nu + \lambda_j \eta_{ij} + \varepsilon_j + \varepsilon_{ij} \tag{5}$$

$$\eta_{ij} = \xi_j + \xi_{ij} \tag{6}$$

$$\lambda_j = \lambda + \lambda \eta_{\psi,j} + \lambda_{0,j} \tag{7}$$

- Factor analysis model for the random loadings extracts the common variation in the random loadings and can be used to model group specific factor variance (approximately)
- Approximately factor variance is $(1 + \eta_{\psi,j})^2$
- Factors now have cluster specific mean and variance

# Random Factor Loadings: Example 3

$$Y_{ij} = \nu + \lambda_j \eta_{ij} + \lambda_2 \eta_j + \varepsilon_j + \varepsilon_{ij} \tag{8}$$

$$\lambda_j = \lambda + \lambda_{0,j} \tag{9}$$

- Separate between level non-random loadings $\lambda_2$ for the between part of the factor $\eta_j$, random within level loadings and non-random between level loadings
- Useful when there are multiple factors on the within level but only one on the between level

All of the above models can be compared using DIC

# Alignment: Alternative to Random Loadings SEM

Asparouhov T. & Muthén, B. (2014). Multiple-group factor analysis alignment. Structural Equation Modeling: A Multidisciplinary Journal, 21, 495-508

$$Y_{ij} = \nu_j + \lambda_j \eta_{ij} + \varepsilon_{ij} \quad (10)$$

$$\eta_{ij} \sim N(\alpha_j, \theta_j) \quad (11)$$

- $\nu_j$, $\alpha_j$, $\theta_j$ and $\lambda_j$ are non-random parameters
- Unidentified model - identified through minimizing CLF alignment function to reduce non-invariance, similar to EFA ability to identify all factor loadings through rotation criteria.
- Substantive difference between Alignment and Random Loadings SEM: Alignment minimizes the number of non-invariant parameters, Random Loadings SEM minimizes overall parameter variability across groups/clusters.
- Can be used with many groups - alternative to multilevel models
- Can be estimated with ML or Bayes

# Cross-Classified SEM

- $Y_{ijk}$ is a multivariate observation for person $i$ belonging to level 2 cluster $j$ and level 3 cluster $k$.
- Level 2 clusters are not nested within level 3 clusters (levels 2a and 2b in Mplus language)
- General Cross-Classified SEM

$$Y_{ijk} = Y_{1ijk} + Y_{2j} + Y_{3k}$$

Three sets of structural equations, one on each level

$$Y_{1ijk} = \nu + \Lambda_1 \eta_{ijk} + \varepsilon_{ijk} \tag{1a}$$

$$\eta_{ijk} = \alpha + B_1 \eta_{ijk} + \Gamma_1 x_{ijk} + \xi_{ijk} \tag{1b}$$

$$Y_{2j} = \Lambda_2 \eta_j + \varepsilon_j \tag{2a}$$

$$\eta_j = B_2 \eta_j + \Gamma_2 x_j + \xi_j \tag{2b}$$

$$Y_{3k} = \Lambda_3 \eta_k + \varepsilon_k \tag{3a}$$

$$\eta_k = B_3 \eta_k + \Gamma_3 x_k + \xi_k \tag{3b}$$

# Cross-Classified SEM with Random Slopes, Loadings and Structural Parameters

$$Y_{ijk} = Y_{1ijk} + Y_{2j} + Y_{3k}$$

$$Y_{1ijk} = \nu + \Lambda_{1jk}\eta_{ijk} + \varepsilon_{ijk} \tag{1a}$$

$$\eta_{ijk} = \eta_{1ijk} + \eta_{2j} + \eta_{3k}$$

$$\eta_{1ijk} = \alpha + B_{1jk}\eta_{1ijk} + \Gamma_{1jk}x_{ijk} + \xi_{ijk} \tag{1b}$$

$$Y_{2j} = \Lambda_2\eta_j + \varepsilon_j \tag{2a}$$

$$\eta_j = B_2\eta_j + \Gamma_2 x_j + \xi_j \tag{2b}$$

$$Y_{3k} = \Lambda_3\eta_k + \varepsilon_k \tag{3a}$$

$$\eta_k = B_3\eta_k + \Gamma_3 x_k + \xi_k \tag{3b}$$

$$\theta_{1jk} = \theta + \theta_j + \theta_k$$

where $\theta_{1jk}$ is any of the parameters in $\Lambda_{1jk}$, $B_{1jk}$, $\Gamma_{1jk}$. $\theta$ is a fixed parameter while $\theta_j$ and $\theta_k$ are zero mean cluster specific deviations one for each of the two cross sections.

# Cross-Classified SEM with Interactions

$$Y_{ijk} = Y_{1ijk} + Y_{2j} + Y_{3k} + \eta_{2j} \cdot \eta_{3k}$$

where $\eta_{2j}$ and $\eta_{3k}$ are arbitrary latent variables defined on the two different nesting levels. Thus a random slope can be used for a latent predictor from a different clustering level. This completes the interaction possibilities (L1xL2), (L1xL3), (L2xL3).

This feature is useful when the items are considered random items and the the two cross nestings are [individual] x [item] (Generalizability theory)

# Model Comparison and Testing

1. Model parameters: can be tested using Bayes credibility intervals
2. Non-model parameter: such as $\theta_1$-$\theta_2$ can be tested from forming the posterior distribution for these parameters and the Bayes credibility intervals (MODEL CONSTRAINTS using NEW parameters in Mplus)
3. Testing for zero variance of random effects: Verhagen & Fox, (2012) - TECH16 in Mplus.
4. DIC for model comparison
5. PPP (posterior predictive p-values) test for model fit for models without random slopes and loadings: test of fit for model implied covariance on each level v.s. unrestricted covariance on each level

- Spiegelhalter, D.J., Best, N.G., Carlin, B.P. & van der Linde, A. (2002). Bayesian measures of model complexity and fit. JRSS
- Compute the deviance conditional on all parameters and random effects (slopes and loadings) $\theta$

$$D(\theta) = -2\log(p(y|\theta))$$

- Compute $p_D$ the effective number of parameters.

$$p_D = \bar{D} - D(\bar{\theta})$$

- Compute DIC

$$DIC = p_D + \bar{D}$$

- $p_D$ is the only penalty for model complexity

# DIC Example 1

Generate cross-classified data 5 indicators, 1 factor, 100 clusters at the two levels. Estimate the true crossed model (M1) and a two-level model ignoring one of the cluster variables (M2).

Table: DIC results

| Model | $p_D$ | DIC |
|---|---|---|
| $M_1$ | 996.6 | 161084 |
| $M_2$ | 499.7 | 196113 |

$M_1$ has 1000 random effects, $M_2$ has 500. DIC makes the correct conclusion (by a wide margin) that the more advanced model $M_1$ is needed.

Now we generate two-level data and estimate the same models. Now $M_2$ is the true model.

Table: DIC results

| Model | $p_D$ | DIC |
|---|---|---|
| $M_1$ | 558.1 | 160380 |
| $M_2$ | 504.2 | 160373 |

$M_1$ has 1000 random effects but half of them are with near zero variance. Small difference in number of parameters and small difference in DIC. DIC makes the correct conclusion (by a small margin) that the less advanced model $M_2$ is the correct model.

# Categorical Data

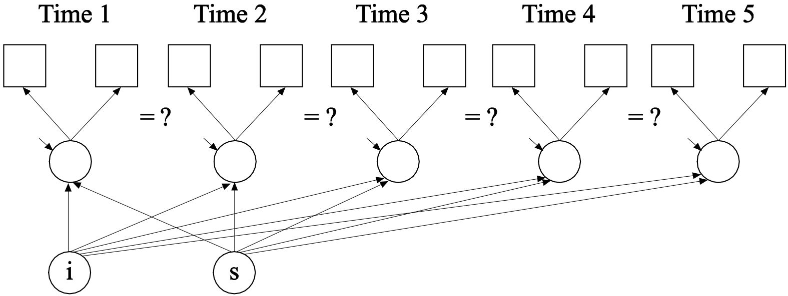
- All of the above models can be estimated with categorical data.
- The model is based on the standard probit link model, cut underlying continuous variable.
- DIC is currently not implemented, not possible in the general model but for some models it is possible. Even conditional on all random effects the likelihood is not explicit and requires numerical integration over within level latent variables.

# Applications to Longitudinal Analysis

Modeling choices:

- An old dilemma - wide v.s. long model
- Two new solutions - one wide and one long

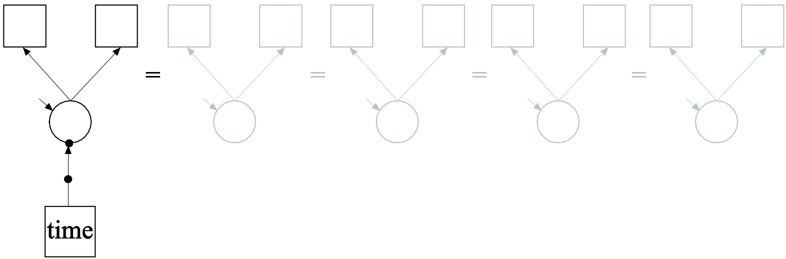
# Categorical Items, Wide Format, Single-Level Approach




Single-level analysis with $p \times T = 2 \times 5 = 10$ variables, $T = 5$ factors.

- ML hard and impossible as T increases (numerical integration)
- WLSMV possible but hard when $p \times T$ increases and biased unless attrition is MCAR or multiple imputation is done first
- Bayes possible
- Searching for partial measurement invariance is cumbersome

# Categorical Items, Long Format, Two-Level Approach



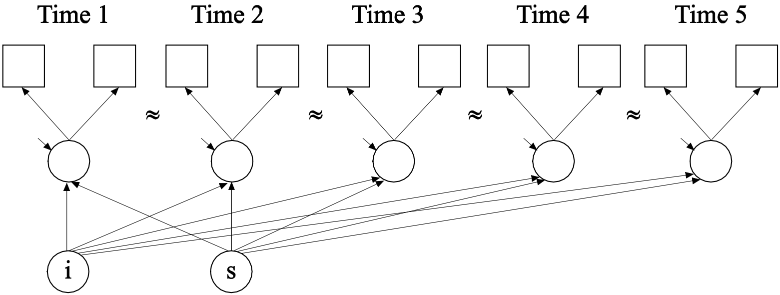

Two-level analysis with $p = 2$ variables, 1 within-factor, 2-between factors, **assuming full measurement invariance across time**.

- ML feasible
- WLSMV feasible with random intercept only (2-level WLSMV)
- Bayes feasible

# Measurement Invariance Across Time

- Both old approaches have problems
  - Wide, single-level approach easily gets significant non-invariance and needs many modifications
  - Long, two-level approach has to assume invariance
- New solution no. 1, suitable for small to medium number of time points
  - A new wide, single-level approach where time is a fixed mode
- New solution no. 2, suitable for medium to large number of time points
  - A new long, cross-classified approach where time is a random mode
  - No limit on the number of time points

# New Solution No. 1: Wide Format, Single-Level Approach

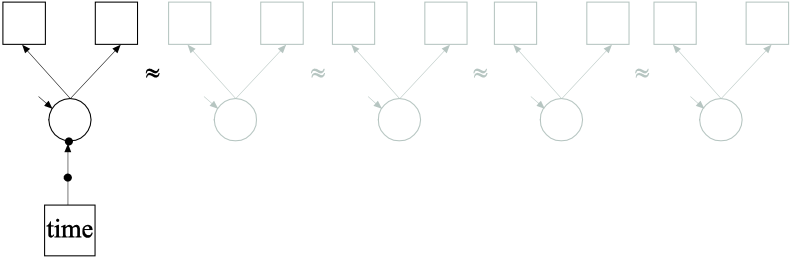



Single-level analysis with $p \times T = 2 \times 5 = 10$ variables, $T = 5$ factors.

- Bayes ("BSEM") using approximate measurement invariance, still identifying factor mean and variance differences across time
- Muthén, B. & Asparouhov, T. (2013). BSEM measurement invariance analysis.

# New Solution No. 2: Long Format, Cross-Classified Approach

- New solution no. 2, time is a random mode
- A new long, cross-classified approach
    - Best of both worlds: Keeping the limited number of variables of the two-level approach without having to assume invariance




Two-level analysis with $p = 2$ variables.

- Bayes cross-classified approach with random measurement parameters and random factor means and variances using Type=Crossclassified: Clusters are time and person
- Asparouhov & Muthén (2012). General random effect latent variable modeling: Random subjects, items, contexts, and par's

# Example: Aggressive-Disruptive Behavior In The Classroom

- Randomized field experiment in Baltimore public schools (Ialongo et al., 1999)
- Teacher-rated measurement instrument capturing aggressive-disruptive behavior among students
- The instrument consists of 9 items scored as 0 (almost never) through 6 (almost always)
- A total of 1174 students are observed in 41 classrooms from Fall of Grade 1 through Grade 7 for a total of 8 time points
- The multilevel (classroom) nature of the data is ignored in the current analyses
- The item distribution is very skewed with a high percentage in the Almost Never category. The items are therefore dichotomized into Almost Never versus the other categories combined
- We analyze the data on the original scale as continuous variables and also the dichotomized scale as categorical

# Aggressive-Disruptive Behavior In The Classroom: ML Versus BSEM for Eight Time Points

- Traditional ML analysis (wide model)
  - 8 dimensions of integration
  - Computing time: 25:44 with Integration = Montecarlo(5000)
  - Increasing the number of time points makes ML impossible
- BSEM analysis with approximate measurement invariance across time (wide model)
  - 156 parameters
  - Computing time: 4:01
  - Increasing the number of time points has relatively less impact

# Displaying Non-Invariant Items using BSEM: Time Points With Significant Differences Compared To The Mean (Prior Variance for Measurement Differences = 0.01)

| Item | Loading | Threshold |
|---|---|---|
| stub | 3 | 1, 2, 3, 6, 8 |
| bkrule | - | 5, 8 |
| harmo | 1, 8 | 2, 8 |
| bkthin | 1, 2, 3, 7, 8 | 2, 8 |
| yell | 2, 3, 6 | - |
| takep | 1, 2, 5 | 1, 2, 5 |
| fight | 1, 5 | 1, 4 |
| lies | - | - |
| tease | - | 1, 4, 8 |

# Cross-Classified Analysis Of Longitudinal Data

- Observations nested within time and subject
- A large number of time points can be handled via Bayesian analysis
- A relatively small number of subjects is needed

# Cross-Classified Analysis: Monte Carlo Simulation Generating The Data For Ex 9.27

```
TITLE:        this is an example of longitudinal modeling using a
              cross-classified data approach where observations are
              nested within the cross-classification of time and subjects
MONTECARLO:
              NAMES = y1-y3;
              NOBSERVATIONS = 7500;
              NREPS = 1;
              CSIZES = 75[100(1)];! 75 subjects, 100 time points
              NCSIZE = 1[1];
              WITHIN = (level2a) y1-y3;
              SAVE = ex9.27.dat;
ANALYSIS:
              TYPE = CROSSCLASSIFIED RANDOM;
              ESTIMATOR = BAYES;
              PROCESSORS = 2;
```

# Aggressive-Disruptive Behavior Example: Model 3 Setup

MODEL:

```
%WITHIN%
s1-s9 | f BY y1-y9;
f@1;
s | f ON time; ! slope growth factor s
%BETWEEN time% ! time variation
y1-y9; ! random intercepts
f@0; [f@0];
s@0; [s@0];
s1-s9*1; [s1-s9*1]; ! random slopes
%BETWEEN id% ! subject variation
y1-y9; ! random intercepts
f*1; [f@0]; ! intercept growth factor
s*1; [s*0]; ! slope growth factor
s1-s9@0; [s1-s9@0];
```

# Aggressive-Disruptive Behavior Example Continued: Model 3 Results For Continuous Analysis

| | Estimate | Posterior S.D. | One-Tailed P-Value | 95% C.I. Lower 2.5% | Upper 2.5 % |
|---|---|---|---|---|---|
| Within Level | | | | | |
| Residual Variances | | | | | |
| Y1 | 1.073 | 0.022 | 0.000 | 1.029 | 1.119 |
| Y9 | 0.630 | 0.014 | 0.000 | 0.604 | 0.658 |
| F | 1.000 | 0.000 | 0.000 | 1.000 | 1.000 |
| Between TIME Level | | | | | |
| Means | | | | | |
| Y1 | 1.632 | 0.120 | 0.000 | 1.377 | 1.885 |
| Y9 | 1.232 | 0.096 | 0.000 | 1.044 | 1.420 |
| S1 | 0.679 | 0.023 | 0.000 | 0.640 | 0.723 |
| S9 | 0.705 | 0.043 | 0.000 | 0.628 | 0.797 |
| Variances | | | | | |
| Y1 | 0.080 | 0.138 | 0.000 | 0.025 | 0.372 |
| Y9 | 0.047 | 0.109 | 0.000 | 0.017 | 0.266 |
| S1 | 0.002 | 0.004 | 0.000 | 0.000 | 0.013 |
| S9 | 0.010 | 0.079 | 0.000 | 0.003 | 0.052 |
| Between ID Level | | | | | |
| Variances | | | | | |
| Y1 | 0.146 | 0.016 | 0.000 | 0.118 | 0.180 |
| Y9 | 0.052 | 0.009 | 0.000 | 0.035 | 0.068 |
| F | 1.316 | 0.080 | 0.000 | 1.172 | 1.486 |
| S | 0.026 | 0.003 | 0.000 | 0.020 | 0.032 |

# Aggressive-Disruptive Behavior Example Continued: Model Comparison using DIC

Compare

- The cross-classified model (M1)
- The two-level model assuming loading and intercept invariance across time (M2)

Table: DIC results

| Model | $p_D$ | DIC |
|---|---|---|
| $M_1$ | 4132 | 150704 |
| $M_2$ | 3854 | 152125 |

Lower DIC indicates that the cross-classified model (M1) is better: measurement model changes over time / age of students.

# Aggressive-Disruptive Behavior Example Conclusions

- Unlike ML and WLS multivariate modeling, for the time intensive Bayes cross-classified SEM, the more time points there are the more stable and easy to estimate the model is
- Bayesian methods solve problems not feasible with ML or WLS
- Time intensive data naturally fits in the cross-classified modeling framework
- Asparouhov and Muthén (2012). General Random Effect Latent Variable Modeling: Random Subjects, Items, Contexts, and Parameters

# Further Applications of Bayes with Random Parameters: 1. Intensive Longitudinal Data

- Time intensive data: More longitudinal data are collected where very frequent observations are made using new tools for data collection. Walls & Schafer (2006)
- Typically multivariate models are developed but if the number of time points is large these models will fail due to too many variables and parameters involved
- Factor analysis models will be unstable over time. Is it lack of measurement invariance or insufficient model?
- Random loading and intercept models can take care of measurement and intercept invariance. A problem becomes an advantage.
- Random loading and intercept models produce more accurate estimates for the loadings and factors by borrowing information over time
- Random loading and intercept models produce more parsimonious model

# Further Applications of Bayes with Random Parameters: 2. Comparison of Many Groups

Groups seen as random clusters

- De Jong, Steenkamp & Fox (2007). Relaxing measurement invariance in cross-national consumer research using a hierarchical IRT model. Journal of Consumer Research, 34, 260-278.
- Fox (2010). Bayesian Item Response Modeling. Springer
- Fox & Verhagen (2011). Random item effects modeling for cross-national survey data. In E. Davidov & P. Schmidt, and J. Billiet (Eds.), Cross-cultural Analysis: Methods and Applications
- Asparouhov & Muthén (2012). General random effect latent variable modeling: Random subjects, items, contexts, and parameters
- Bayesian estimation needed because random loadings with ML give rise to numerical integration with many dimensions

# How to Learn More About Bayesian Analysis in Mplus: www.statmodel.com

- Topic 9 handout and video from the 6/1/11 Mplus session at Johns Hopkins
- Part 1 - Part 3 handouts and video from the August 2012 Mplus Version 7 training session at Utrecht University