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

Bengt Muthén & Tihomir Asparouhov

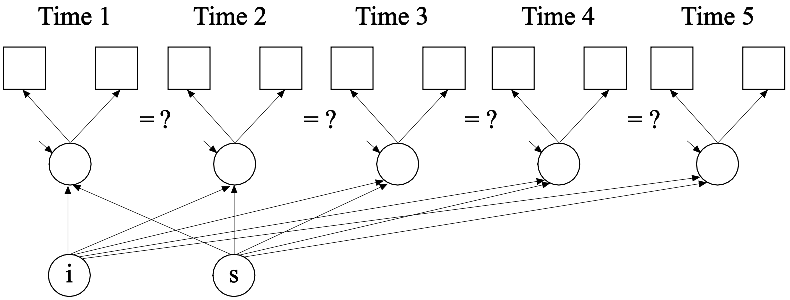
Mplus
www.statmodel.com

Presentation at the UConn M3 meeting, May 20, 2014

Table of Contents I

- An old dilemma
- Two new solutions

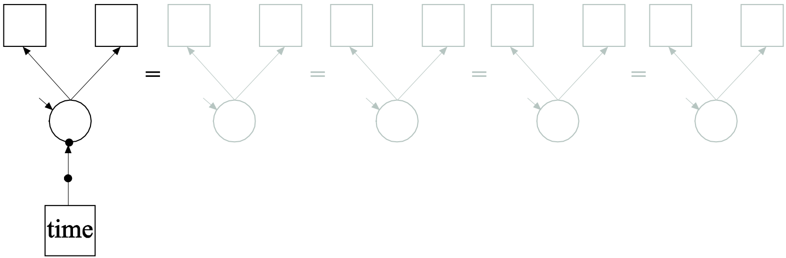
Categorical Items, Wide Format, Single-Level Approach



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



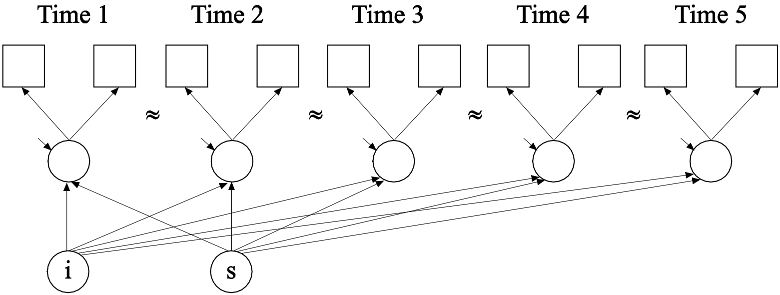
Two-level analysis with $p = 2$ variables, 1 within-factor, 2-between factors, **assuming full measurement invariance across time**.

- ML feasible
- WLSMV feasible (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, two-level approach where time is a random mode
  - No limit on the number of time points

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

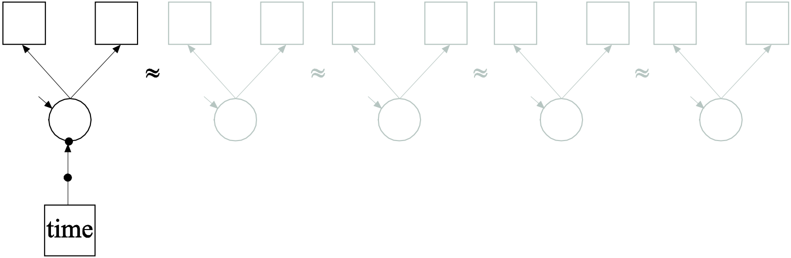


Single-level analysis with $p \times T = 2 \times 5 = 10$ variables, $T = 5$ factors.

- 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, Two-Level Approach

- New solution no. 2, time is a random mode
- A new long, two-level 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 twolevel random 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
  - 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
  - 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 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

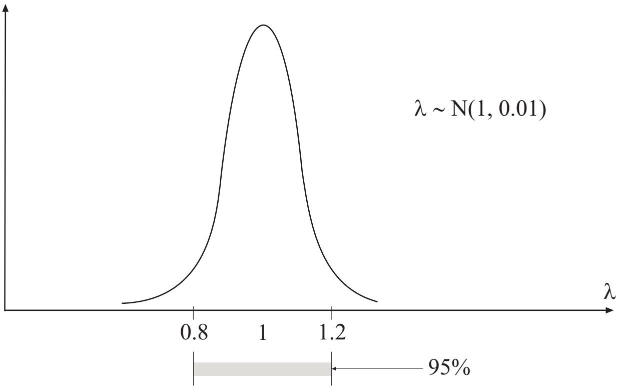
Groups seen as random clusters

- 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

Two-Level Analysis with Random Item Parameters: A New Conceptualization of Measurement Invariance

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

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



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