# Late-Breaking News: Some Exciting New Methods

Bengt Muthén & Tihomir Asparouhov

Mplus
www.statmodel.com

bmuthen@statmodel.com

Keynote Address at the Modern Modeling Methods Conference,
University of Connecticut,
May 22, 2013

# Overview

- Overview not provided in order to surprise you the better - summary at the end instead
- But all new methods to be presented are available in Mplus Version 7.1 to be released shortly
- Tomorrow's workshop shows how to use these new methods

# Some Provocative Statements

- Factor analysis gives the wrong factor correlations
  - EFA correlations too small
  - CFA correlations too large
- Non-identified models can be estimated, interpretable, and useful
- Measurement invariance is ignored in two-level factor analysis
- Random intercept, random slope models fit the data well, but lose in interpretability

# When Last We Spoke

Remember ESEM and BSEM?

- ESEM: Asparouhov & Muthén (2009). Exploratory structural equation modeling. Structural Equation Modeling, 16, 397-438.
- BSEM: Muthén & Asparouhov (2012). Bayesian SEM: A more flexible representation of substantive theory. Psychological Methods, 17, 313-335.

A frequentist and a Bayesian approach to approximate fit, in that case applied to factor analysis

The general strategies behind these papers will be further developed in different ways in today's talk.

# Topics

General features:

- Comparisons of many groups
- Measurement invariance & factor mean and variance estimation

Application areas:

- Cross-cultural studies (International Social Survey Program, European Social Survey)
- Health care ratings for different hospitals (Malcom Baldrige National Quality Award criteria)
- Achievement comparisons across countries (PISA, TIMSS, PIRLS)
- School comparisons (LSAY, ECLS)
- Teacher ratings of student behavior in classrooms

# Examples: 4 Data Sets

1. 34 countries (n=45,546): Cross-cultural study of nationalism and patriotism
2. 67 hospitals (n=7,168): Quality management
3. 40 countries (n=9,787): Math achievement
4. 39 classrooms (n=1,054): Aggressive-disruptive behavior

Let's do some analyses together!

# Analysis Choices for Multiple Groups/Clusters: Fixed vs Random Effect Factor Analysis (IRT)

- Fixed mode: Multiple-group analysis
  - Inference to the groups in the sample
  - Usually a relatively small number of groups
- Random mode: Two-level factor analysis
  - Inference to a population from which the groups/clusters have been sampled
  - Usually a relatively large number of groups/clusters

\- How could there possibly be something new to say in these areas??

# Refresher on Multiple-Group Factor Analysis: 3 Different Degrees of Measurement Invariance

1. CONFIGURAL (invariant factor loading pattern)
2. METRIC (invariant factor loadings; ”weak factorial invariance”)
   - Needed in order to compare factor variances across groups
3. SCALAR (invariant factor loadings and intercepts/thresholds; ”strong factorial invariance”)
   - Needed in order to compare factor means across groups

(These are automatically specified in Mplus Version 7.1 by 3 new options in the ANALYSIS command:
MODEL=CONFIGURAL METRIC SCALAR;)

# Refresher on Multiple-Group Factor Analysis: Formulas for Individual $i$ and Group $j$

- Configural:

$$y_{ij} = \nu_j + \lambda_j f_{ij} + \varepsilon_{ij},$$
$$E(f_j) = \alpha_j = 0, V(f_j) = \psi_j = 1.$$

- Metric:

$$y_{ij} = \nu_j + \lambda f_{ij} + \varepsilon_{ij},$$
$$E(f_j) = \alpha_j = 0, V(f_j) = \psi_j.$$

- Scalar:

$$y_{ij} = \nu + \lambda f_{ij} + \varepsilon_{ij},$$
$$E(f_j) = \alpha_j, V(f_j) = \psi_j.$$

# Specification Searches for Measurement Invariance

Measurement invariance ("item bias", "DIF") has traditionally been concerned with comparing a small number of groups such as with gender or ethnicity.

Likelihood-ratio chi-square testing of one item at a time:

- Bottom-up: Start with no invariance (configural case), imposing invariance one item at a time
- Top-down: Start with full invariance (scalar case), freeing invariance one item at a time, e.g. using modification indices

Neither approach is scalable - both are very cumbersome when there are many groups, such as 50 countries ($50 \times 49/2 = 1225$ pairwise comparisons for each item). The correct model may well be far from either of the two starting points, which may lead to the wrong model.

# Example 1: Cross-Cultural Data on Nationalism and Patriotism

Davidov (2009). Measurement equivalence of nationalism and constructive patriotism in the ISSP: 34 countries in a comparative perspective. Political Analysis,17, 64-82.

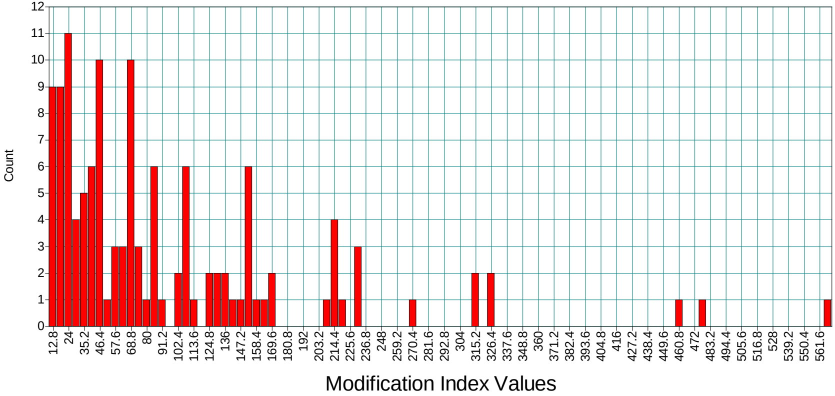
- Data from the International Social Survey Program (ISSP) 2003 National Identity Module
- 34 countries, n=45,546
- 5 measurements of nationalism and patriotism
- Expected 2-factor structure

# Nationalism and Patriotism Data: Item Wording

- Nationalism factor:
  - V21: The world would be a better place if people from other countries were more like in [own country]
  - V22: Generally speaking, [own country] is better than most other countries
- Constructive Patriotism factor:
  - V26: How proud are you of [respondent's country] in the way democracy works?
  - V29: How proud are you of [respondent's country] in its social security system?
  - V35: How proud are you of [respondent's country] in its fair and equal treatment of all groups in society?

# Nationalism and Patriotism Data: Multiple-Group ML CFA with Scalar Invariance Across All 34 Countries

Modification indices show a multitude of large values for the constrained measurement parameters (intercept modindices shown):



# Nationalism and Patriotism Data: Multiple-Group Analysis with Scalar Invariance Across All 34 Countries, Cont'd

- Multiple-group ML CFA:
  - Due to the many large modification indices and the many groups seen in the previous figure, model modification based on these indices will involve many steps with a big risk of model mis-specification
  - The author gave up on scalar invariance and was not able to compare country means on the factors
- Multiple-group ML ESEM:
  - More relaxed model (EFA-based) allowing cross-loadings
  - The scalar invariance model has $\chi^2(496) = 13,893$ with many large modification indices
- Allowing cross-loadings using BSEM also fits poorly

- A new method is needed!!!

# Two New Multiple-Group Factor Analysis Methods (Fixed Mode)

Both methods use completely non-identified models!

Both methods use the idea of approximate measurement invariance in line with the philosophy of Box (1955): "Statistical criteria should (1) be sensitive to change in the specific factors tested, (2) be insensitive to changes, of a magnitude likely to occur in practice, in extraneous factors."

1. Frequentist alignment optimization
   - Asparouhov & Muthén (2013). Web note 18
2. Bayesian multiple-group analysis (multiple-group BSEM)
   - Muthén & Asparouhov (2013). BSEM measurement invariance analysis. Web note 17

# Multiple-Group Alignment Optimization

1. Estimate the configural model (loadings and intercepts free across groups, factor means fixed @0, factor variances fixed @1)
2. Alignment optimization:
   - Free the factor means $\alpha_j$ and variances $\psi_j$, noting that for every set of factor means and variances the same fit as the configural model is obtained with loadings $\lambda_j$ and intercepts $\nu_j$ changed as:

$$\lambda_j = \lambda_{j,configural} / \sqrt{\psi_j},$$
$$\nu_j = \nu_{j,configural} - \alpha_j \, \lambda_{j,configural} / \sqrt{\psi_j}.$$

   - Choose $\alpha_j$ and $\psi_j$ to minimize the total amount of non-invariance using a simplicity function

$$F = \sum_p \sum_{j_1 < j_2} w_{j_1,j_2} f(\lambda_{pj_1} - \lambda_{pj_2}) + \sum_p \sum_{j_1 < j_2} w_{j_1,j_2} f(\nu_{pj_1} - \nu_{pj_2}),$$

   for every pair of groups and every intercept and loading using a component loss function (CLF) $f$ from EFA rotations (Jennrich, 2006) such as

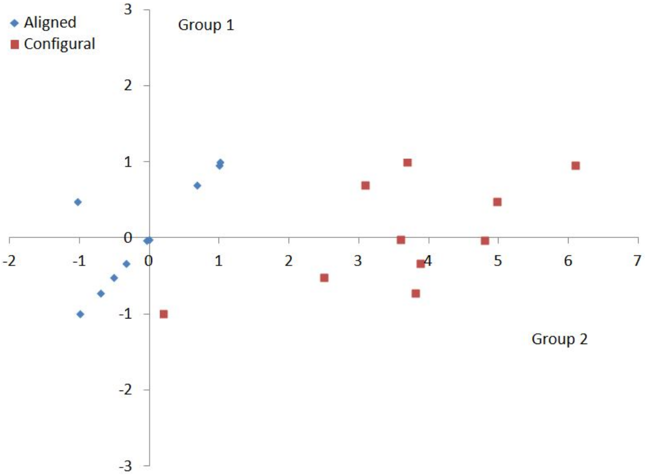
$$f(x) = \sqrt{\sqrt{x^2 + \varepsilon}}$$

   where $\varepsilon$ is a small number such 0.0001.

# Alignment Optimization, Continued

- The simplicity function $F$ is optimized at a few large non-invariant parameters and many invariant parameters rather than many medium-sized non-invariant parameters (compare with EFA rotations using functions that aim for either large or small loadings, not mid-sized loadings)
- In this way, a non-identified model where factor means and factor variances are added to the configural model is made identified by adding a simplicity requirement
- In line with having different EFA rotations, different variations of simplicity functions can be chosen such as $f(x) = \sqrt{\sqrt{x^2 + \varepsilon}}$ or $f(x) = \sqrt{\sqrt{\sqrt{x^2 + \varepsilon}}}$
- Simulation studies show that the alignment method works very well
- For well-known examples with few groups and few non-invariances, the results agree with the alignment method

# Alignment Optimization Visually for Two Groups



# How Do We use the Alignment Results?

In addition to the estimated aligned model, the alignment procedure gives

- Measurement invariance test results produced by an algorithm that determines the largest set of parameters that has no significant difference between the parameters
- Factor mean ordering among groups and significant differences produced by z-tests

# Back to the Example of Nationalism-Patriotism Multiple-Group Factor Analysis

Recall that CFA, ESEM, and BSEM with cross-loadings had all failed in that too many instances of scalar measurement non-invariance were found: Factor means could not be compared across groups.

The problem is that these methods start with the scalar model of full invariance which is too far from the true model which has some large non-invariances and many ignorable non-invariances.

The alignment method resolves this problem, making the factor means and variances comparable across groups and reducing the number of significant non-invariances.

# Nationalism and Patriotism Example: Alignment Results

Approximate Measurement (Non-) Invariance by Group

**Intercepts** for Nationalism indicators (V21, V22) and Patriotism indicators (V26, V29, V35)

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| V21 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 |
| | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | | |
| V22 | (1) | 2 | 3 | (4) | 5 | (6) | 7 | 8 | (9) | 10 | 11 | 12 |
| | 13 | 14 | (15) | (16) | 17 | 18 | (19) | (20) | 21 | (22) | (23) | 24 |
| | (25) | 26 | 27 | 28 | (29) | 30 | 31 | (32) | 33 | 34 | | |
| V26 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 |
| | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | | |
| V29 | (1) | 2 | 3 | (4) | (5) | 6 | 7 | (8) | (9) | 10 | 11 | 12 |
| | (13) | 14 | 15 | 16 | (17) | 18 | (19) | (20) | (21) | (22) | (23) | (24) |
| | (25) | 26 | 27 | 28 | 29 | (30) | 31 | 32 | 33 | (34) | | |
| V35 | (1) | (2) | 3 | (4) | 5 | 6 | 7 | (8) | (9) | (10) | 11 | 12 |
| | 13 | 14 | 15 | 16 | 17 | 18 | (19) | (20) | 21 | (22) | 23 | (24) |
| | 25 | 26 | (27) | (28) | (29) | (30) | 31 | 32 | (33) | 34 | | |

# Nationalism and Patriotism Example: Alignment Results

**Loadings** for NATIONALISM factor

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| V21 | 1 | (2) | (3) | 4 | 5 | 6 | 7 | (8) | (9) | (10) | 11 | 12 |
| | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | (23) | (24) |
| | (25) | 26 | 27 | 28 | 29 | (30) | 31 | 32 | 33 | 34 | | |
| V22 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 |
| | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | | |

Loadings for PATRIOTISM factor

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| V26 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | (21) | (22) | 23 | 24 |
| | 25 | 26 | 27 | (28) | 29 | 30 | 31 | 32 | 33 | 34 | | |
| V29 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| | 13 | 14 | 15 | 16 | 17 | 18 | (19) | 20 | 21 | 22 | 23 | (24) |
| | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | | |
| V35 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 |
| | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | | |

# Nationalism and Patriotism Example: Factor Mean Comparisons (5% Significance Level)

Results for NATIONALISM factor

| Ranking | Group | Value | Groups with significantly smaller factor mean |
|---|---|---|---|
| 1 | 22 | 0.067 | 2 19 11 12 9 24 23 10 15 20 33 14 32 29 13 7 6 8 16 4 21 1 26 27 34 30 31 3 25 5 |
| 2 | 28 | 0.000 | 19 11 12 9 24 23 15 20 33 14 32 29 13 7 6 8 16 4 21 1 26 27 34 30 31 3 25 5 18 17 |
| 3 | 2 | -0.284 | 6 16 4 21 1 26 27 34 31 3 25 5 18 17 |
| 4 | 19 | -0.333 | 32 13 7 6 16 4 21 1 26 27 34 31 3 25 5 18 17 |
| 5 | 11 | -0.344 | 33 32 13 7 6 16 4 21 1 26 27 34 31 3 25 5 18 17 |
| 6 | 12 | -0.352 | 13 7 6 16 4 21 1 26 27 34 31 3 25 5 18 17 |
| 7 | 9 | -0.357 | 7 6 16 4 21 1 26 27 34 31 3 25 5 18 17 |
| 8 | 24 | -0.379 | 6 16 4 21 1 26 27 34 31 3 25 5 18 17 |
| 9 | 23 | -0.388 | 13 7 6 16 4 21 1 26 27 34 31 3 25 5 18 17 |
| 10 | 10 | -0.395 | 16 4 21 1 26 27 34 31 3 25 5 18 17 |
| 11 | 15 | -0.396 | 13 7 6 16 4 21 1 26 27 34 31 3 25 5 18 17 |
| 12 | 20 | -0.413 | 13 7 6 16 4 21 1 26 27 34 31 3 25 5 18 17 |

# Switching to Random Mode: What Can Two-Level Factor Analysis Tell Us About Invariance?

Refresher on Two-Level Factor Analysis - 3 Major Types of Models:

1. Random intercepts: Different Within and Between factor structures (from factor analysis tradition)
2. Non-random intercepts: Same Within and Between factor structures and Between residual variances = 0 (used in IRT)
3. Random intercepts & random loadings (Bayesian analysis)

# Two-Level Factor Analysis: Different Within and Between Factor Structures

Recall random effect ANOVA for individual $i$ in cluster $j$,

$$y_{ij} = \nu + y_{B_j} + y_{W_{ij}}.$$

Two-level factor analysis generalizes this to

$$y_{ij} = \nu + \lambda_B f_{B_j} + \epsilon_{B_j} \; + \; \lambda_W f_{W_{ij}} + \epsilon_{W_{ij}}$$

with covariance structure $V(y_{ij}) = \Sigma_B + \Sigma_W$, where

$$\Sigma_B = \Lambda_B \, \Psi_B \, \Lambda_B' + \Theta_B,$$
$$\Sigma_W = \Lambda_W \, \Psi_W \, \Lambda_W' + \Theta_W.$$

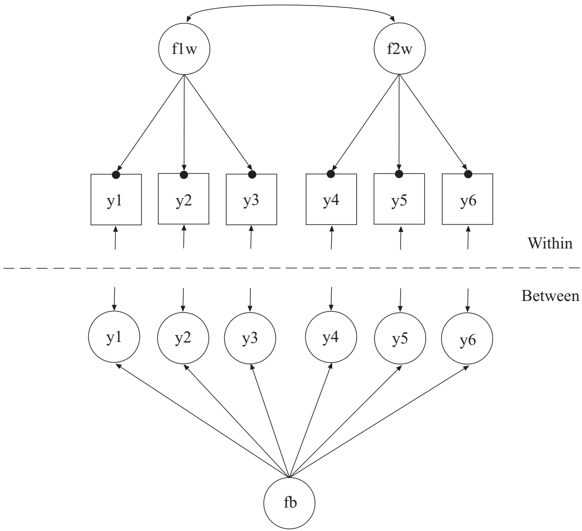
# Random Intercept Two-Level Factor Analysis: Different Within and Between Factor Structures

The two-level factor analysis model

$$y_{ij} = \nu + \lambda_B \, f_{B_j} + \varepsilon_{B_j} \; + \; \lambda_W \, f_{W_{ij}} + \varepsilon_{W_{ij}}$$

can be viewed as a random intercept model:

$$\begin{aligned} Level\,1: \; y_{ij} &= \nu_j + \lambda_W \, f_{W_{ij}} + \varepsilon_{W_{ij}}, \\ Level\,2: \; \nu_j &= \nu + \lambda_B \, f_{B_j} + \varepsilon_{B_j}. \end{aligned}$$


f1w
f2w
y1
y2
y3
y4
y5
y6
Within
Between
y1
y2
y3
y4
y5
y6
fb

# Connections Between Random Intercept Two-Level Factor Analysis, Conventional Two-Level IRT, and Measurement Invariance

- Random intercept two-level factor analysis:

$$Level\ 1: \ y_{ij} = \nu_j + \lambda_W f_{W_{ij}} + \varepsilon_{W_{ij}},$$
$$Level\ 2: \ \nu_j = \nu + \lambda_B f_{B_j} + \varepsilon_{B_j},$$

- Conventional two-level IRT:
  If $\lambda_W = \lambda_B = \lambda$ and $V(\varepsilon_{B_j}) = 0$, then the above equations become

$$y_{ij} = \nu + \lambda f_{ij} + \varepsilon_{ij},$$
$$f_{ij} = f_{B_j} + f_{W_{ij}},$$

- The IRT model implies that we have measurement invariance across the clusters for both the intercepts and the loadings

# Testing Measurement Invariance with Random Intercept Two-Level Factor Analysis

- Jak et al. (2013a). A test for cluster bias: Detecting violations of measurement invariance across clusters in multilevel data. SEM journal, April-June issue.
- Jak et al. (2013b). Measurement bias in multilevel data. To appear in SEM.

# Example 2: Hospital Data Example

Shortell et al. (1995). Assessing the impact of continuous quality improvement/total quality management: concept versus implementation. Health Services Research, 30, 377-401.

- Survey of 67 hospitals, $n = 7168$ employee respondents, approximately 100/hospital
- 6 dimensions of an overall ”quality improvement implementation” based on the Malcom Baldrige National Quality Award criteria
- Focus on 6 items measuring a quality management dimension

# Hospital as Random Mode: Regular Random Intercept, Two-Level Factor Analysis using Jak's Approach

- Testing $\Lambda_B = \Lambda_W$, $\Theta_B = 0$: $\chi^2(20) = 206.33$, p-value $= 0.000$.
- Modification indices for Between Level point to $\Theta_B$ for QM53:

| | M.I. | E.P.C. | Std E.P.C. | StdYX E.P.C. |
|---|---|---|---|---|
| BY Statements | | | | |
| QMB BY QM53 | 41.191 | 0.307 | 0.075 | 0.619 |
| QMB BY QM56 | 23.359 | -0.213 | -0.052 | -0.343 |
| QMB BY QM57 | 10.394 | 0.125 | 0.031 | 0.187 |
| Residual Variances | | | | |
| QM53 | **248.063** | 0.021 | 0.021 | 1.402 |
| QM54 | 41.552 | 0.006 | 0.006 | 0.257 |
| QM55 | 57.373 | 0.008 | 0.008 | 0.369 |
| QM57 | 15.049 | 0.003 | 0.003 | 0.121 |
| QM58 | 14.616 | 0.004 | 0.004 | 0.185 |

# Hospital Data: Quality Management Items

- **QM53: The hospital regularly checks equipment and supplies to make sure they meet quality requirements**
- QM54: The quality assurance staff effectively coordinate their efforts with others to improve the quality of services the hospital provides.
- QM55: Hospital employees have a good understanding of how to improve the quality of services
- QM56: Data from suppliers are used when developing the hospital's plan to improve quality
- QM57: The hospital has effective policies for improving the quality of services
- QM58: The hospital works closely with suppliers to improve the quality of their products and services

# Hospital as Fixed Mode: Alignment Optimization with Approximate Intercept (Non-) Invariance by Group

| | |
|---|---|
| QM53 | 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 **(50)** 51 **(52)** 53 54 55 56 57 58 59 60 61 **(62)** 63 64 65 66 67 |
| QM54 | 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 |
| QM55 | 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 |
| QM56 | 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 |
| QM57 | 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 |
| QM58 | 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 |

# Random Intercepts, Random Loadings Two-Level Factor Analysis (IRT)

Proposed for Bayesian IRT (de Jong, Fox, Asparouhov-Muthén). For an item $y_{ij}$ observed for individual $i$ in cluster $j$ and measuring the factor $\theta_{ij}$,

$$
\begin{aligned}
Level\ 1: \ y_{ij} &= \nu_j + \lambda_j\, \theta_{ij} + \varepsilon_{ij}, \\
\theta_{ij} &= \theta_{B_j} + \theta_{W_{ij}}, \\
Level\ 2: \ \nu_j &= \nu + \delta_{1j}, \\
Level\ 2: \ \lambda_j &= \lambda + \delta_{2j},
\end{aligned}
$$

The many random loadings require Bayesian analysis. Factor variance variation across clusters can be modeled to not confound this with loading non-invariance. Implemented in Mplus Version 7.

- Fox, J.-P., and A. J. 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
- Fox (2010). Bayesian Item Response Modeling. Springer
- Program for International Student Assessment (PISA 2003)
- 9,769 students across 40 countries
- 8 binary math items

# Random Loadings In IRT using IRT Formulas

- $Y_{ijk}$ - outcome for student i, in country j and item k

$$P(Y_{ijk} = 1) = \Phi(a_{jk}\theta_{ij} + b_{jk})$$

$$a_{jk} \sim N(a_k, \sigma_{a,k}), b_{jk} \sim N(b_k, \sigma_{b,k})$$

  Both discrimination ($a$) and difficulty ($b$) vary across country
- The $\theta$ ability factor is decomposed as

$$\theta_{ij} = \theta_j + \varepsilon_{ij}$$

$$\theta_j \sim N(0, v), \varepsilon_{ij} \sim N(0, v_j), \sqrt{v_j} \sim N(1, \sigma)$$

- The mean and variance of the ability vary across country
- For identification purposes the mean of $\sqrt{v_j}$ is fixed to 1, this replaces the traditional identification condition that $v_j = 1$
- Model preserves common measurement scale while accommodating measurement non-invariance as long as the variation in the loadings is not big

Fox & Verhagen (2011):

- The factor is on a comparable scale across countries despite random measurement non-invariance
- Measurement invariance testing performed by considering item parameter variance across countries (Bayes Factors)
- ”It is concluded that only the difficulty of item 8 is invariant”

Intercepts

| | |
|---|---|
| Y1 | 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 **(18)** 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 |
| Y2 | 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 **(38)** 39 40 |
| Y3 | 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 **(34)** 35 36 37 38 39 40 |
| Y4 | 1 2 3 4 5 6 7 8 9 10 11 **(12)** 13 14 15 16 17 18 19 20 21 22 23 24 25 26 **(27)** 28 29 30 31 32 33 34 35 36 37 38 39 40 |
| Y5 | 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 |
| Y6 | 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 **(18)** 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 |
| Y7 | 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 |
| Y8 | 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 |

# Alignment vs Two-Level Factor Analysis (Fixed vs Random)

Alignment advantages:

- Convenient, one-step analysis
- Points to which groups/clusters contribute to non-invariance
- Is not limited to just $> 30$ clusters, but works well with any number of groups/clusters (say $< 100$, or say $< 3,000$ configural parameters)
- Gives an ordering of the factor means without having to estimate factor scores for each group/cluster
- Allows factor variance variation across groups/clusters without involving random slopes
- Does not assume normally-distributed non-invariance

Two-level advantages:

- Easy to handle a huge number of groups/clusters
- Easy to relate measurement non-invariance to variables on the group/cluster level (Jak et al., 2013b), explaining part of the item parameter variance

# Alignment with Small Group Sizes

Monte Carlo simulations of alignment optimization show good parameter recovery and coverage also for group sizes as small as 30.

Useful for classroom-based research.

# Example 4: Analysis of Aggressive-Disruptive Behavior in Baltimore Classrooms

- Teacher-rated aggressive-disruptive behavior of first-grade students in Baltimore public schools (TOCA)
- 1054 students in 39 classrooms (Cohort 1)
- Behavior rated on a 6-point scale from Almost Never to Almost Always. 9 items considered here:
  - Stubborn
  - Breaks rules
  - Harms others and property
  - Breaks things
  - Yells at others
  - Takes others' property
  - Fights
  - Lies
  - Teases classmates

# The Measurement Process in the Baltimore Study

- The students in each classroom are rated by the same single teacher
- The teachers use a standardized interview protocol that requires 3 days of training
- But, do the teachers actually use the rating scale the same way?
    - For instance, does ”teasing classmates” mean the same thing for different teachers?
- Finding the answer: Treat classrooms/teachers as groups and do multiple-group factor analysis using alignment
- Difficulty specific to this type of application: Small group sizes

# Classroom Sizes in the Baltimore Study

| Size (s) | Classroom ID with Size s |
|---|---|
| 15 | 32 |
| 20 | 36 |
| 21 | 35 31 |
| 22 | 33 34 8 23 |
| 23 | 10 |
| 24 | 7 2 41 |
| 25 | 9 29 21 |
| 26 | 5 22 14 26 |
| 27 | 4 24 25 |
| 29 | 3 15 6 1 |
| 30 | 28 18 19 |
| 31 | 13 16 39 40 20 |
| 33 | 11 17 |
| 34 | 38 |
| 36 | 37 |
| 37 | 12 |
| Average classroom size | 27.026 |

Intercepts

| | |
|---|---|
| STUB1F | 1 2 3 **(4)** 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 |
| BKRULE1F | **(1)** 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 **(18)** 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 |
| HARMO1F | 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 |
| BKTHIN1F | 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 |
| YELL1F | 1 2 3 4 5 6 7 8 9 10 11 **(12)** 13 14 15 16 17 18 19 20 21 22 23 **(24)** 25 26 27 28 29 30 31 32 33 34 **(35)** 36 37 38 39 |
| TAKEP1F | 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 **(21)** 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 |
| FIGHT1F | 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 |
| LIES1F | 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 **(30)** **(31)** 32 33 34 35 36 37 38 39 |
| TEASE1F | 1 2 3 4 5 6 7 8 9 10 11 12 **(13)** 14 15 16 17 **(18)** 19 20 21 22 23 24 25 **(26)** 27 28 **(29)** 30 31 32 33 **(34)** 35 36 37 38 39 |

# It is Simple to Set Up an Alignment Analysis: Nationalism & Patriotism in 34 Countries ($n = 45,546$)

```
DATA:        FILE = issp.txt;
VARIABLE:    NAMES = country v21 v22 v26 v29 v35;
             USEVARIABLES = v21-v35;
             CLASSES = c(34);
             KNOWNCLASS = c(country);
ANALYSIS:    TYPE = MIXTURE;
             ESTIMATOR = ML;
             ALIGNMENT = FREE;
MODEL:       %OVERALL%
             nat BY v21-v22;
             pat BY v26-v35;
```

Takes 34 seconds.

# Fit of the Model in Each Group/Cluster

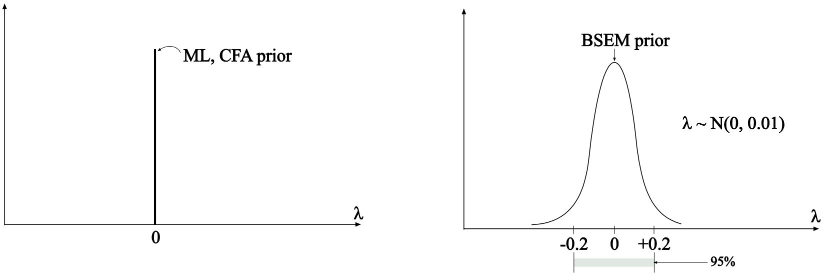
Using fixed mode analysis, the configural model often does not fit the data in every group. - How serious is that?

Box & Draper (1987): ”essentially, all models are wrong, but some are useful”.

- Fixed mode: Alignment model fit same as configural model fit
  - Measurement invariance analysis is questionable if the configural model does not fit in each group
  - Fit judged by ML $\chi^2$ or Bayes Posterior Predictive Checking
- Random mode: Two-level factor analysis does not automatically judge fit in each cluster

What does Bayes contribute?

- ML CFA is characterized by many zero factor loadings
- ML CFA implicitly uses a strong prior with an exact zero loading
- BSEM uses an approximate zero loading using a zero-mean, small-variance prior for the loading:




BSEM can be used to specify approximate zeros for

- Cross-loadings
- Residual correlations
- Direct effects from covariates
- Group and time differences in intercepts and loadings

# The Several Uses of BSEM

Using zero-mean, small-variance priors.

- Single group analysis (2012 Psych Methods article):
  - Cross-loadings
  - Residual covariances
  - Direct effects in MIMIC
- Multiple-group analysis:
  - Configural and scalar analysis with cross-loadings and/or residual covariances
  - Approximate measurement invariance (Web Note 17)
  - BSEM-based alignment optimization (Web Note 18):
    - Residual covariances
    - Approximate measurement invariance

# Bayes and BSEM Alignment for Fixed Mode Analysis

What does Bayes contribute to assessing model fit?

1. Configural model: Bayes with informative, zero-mean, small-variance priors for residual covariances can allow better configural fit - configural misfit in some groups is a common problem
2. Scalar model: Bayes with informative, zero-mean, small-variance priors for measurement parameter differences across groups (multiple-group BSEM) can allow better scalar fit
   - MG-BSEM as an alternative to alignment (finds non-invariance; needs alignment unless non-invariant parameters are freed)
   - MG-BSEM-based alignment (advantageous for small samples?)

Further Bayes advantage: Bayes alignment can produce plausible values for the subjects' factor score values to be used in further analyses

# Back to the Hospital Example: ML Invariance Testing of One-Factor Model for 67 hospitals, n=7,168

| Model | Number of Parameters | Chi-square | Degrees of Freedom | P-value |
|---|---|---|---|---|
| Configural | 1206 | 1672.639 | 603 | 0.0000 |
| Metric | 876 | 2226.655 | 933 | 0.0000 |
| Scalar | 546 | 3005.402 | 1263 | 0.0000 |

| Models Compared | Chi-square | Degrees of Freedom | P-value |
|---|---|---|---|
| Metric against Configural | 554.016 | 330 | 0.0000 |
| Scalar against Configural | 1332.763 | 660 | 0.0000 |
| Scalar against Metric | 778.747 | 330 | 0.0000 |

# Hospital Example: Bayes and BSEM Invariance Testing

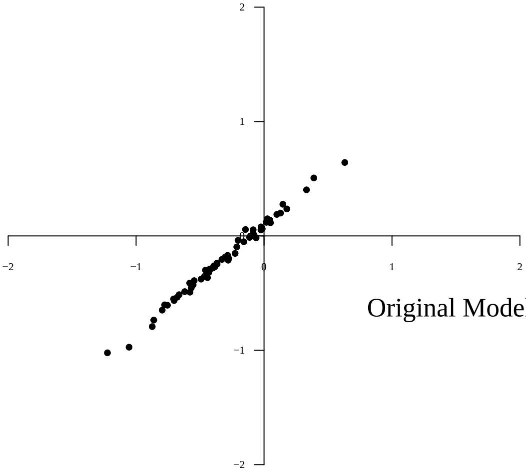
Posterior Predictive Checking: 95% CIs for the difference between observed and replicated $\chi^2$ values and Posterior Predictive p-values

1. Bayes Alignment using the ML model: [542, 918], p=0.000
2. Bayes Alignment allowing for residual covariances using zero-mean, small-variance IW priors: [66, 422], p=0.000
3. Bayes (BSEM-based) Alignment with approximate measurement invariance and allowing for residual covariances using zero-mean, small-variance IW priors: [-22, 306], p=0.078

For the number 3 model, only a few hospitals show significant residual covariances for only a few pairs of items

# Hospital Example: Plot of Factor Means Using Two Bayes Approaches

BSEM allowing residual covariances




Original Model

# Summary

- Multiple groups/clusters data can be represented by fixed or random mode models
  - Having many groups/clusters does not preclude fixed-mode, multiple-group analysis
- Fixed mode modeling can explore the data using non-identified models:
  - Alignment optimization
  - BSEM methods
- Random mode modeling:
  - Conventional two-level factor analysis reveals some limited forms of non-invariance (intercepts)
  - Random slope two-level factor analysis reveals more general forms of non-invariance

# Summary, Continued

- Fixed mode modeling using alignment optimization has many advantages over random mode modeling:
  - Convenient, one-step analysis
  - Points to which groups/clusters contribute to non-invariance
  - Is not limited to just $> 30$ clusters, but works well with any number of groups/clusters (say $< 100$, or say $< 3,000$ configural parameters)
  - Gives an ordering of the factor means without having to estimate factor scores for each group/cluster
  - Allows factor variance variation across groups/clusters without involving random slopes
  - Does not assume normally-distributed non-invariance

# To Conclude

The big news: Alignment optimization

- Does modeling with group-specific measurement intercepts, measurement loadings, factor means, and factor variances
- Aligns to minimal measurement non-invariance
- Uses EFA-like tools to identify non-identified parameters
- Is easy to do

The other news: The Alignment optimization companion technique - Multiple-group BSEM

# Thank You

- See you tomorrow morning at 8:30!