New Methods to Study Measurement Invariance in Many Groups

Bengt Muthén & Tihomir Asparouhov

Mplus
www.statmodel.com

bmuthen@statmodel.com

Talk at the European Survey Research Association (ESRA) Conference, University of Ljubljana, July 16, 2013

Topics

General features:

- Factor analysis model for a measurement instrument
- Comparisons of many groups
- Measurement invariance & factor mean and variance estimation
- Approximate measurement invariance

Application areas:

- Cross-cultural studies (International Social Survey Program, European Social Survey)
- Achievement comparisons across countries (PISA, TIMSS, PIRLS)
- School comparisons (LSAY, ECLS)
- Organizational research

Examples: 4 Data Sets

1. 26 countries (n=49,894): ESS cross-cultural study of tradition & conformity items
2. 40 countries (n=9,787): PISA math achievement
3. 53 countries (n=420,000): PISA items on teacher-student relationships
4. 67 hospitals (n=7,168): Health care ratings for different hospitals

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

Fixed Mode: 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: ESS Tradition-Conformity Items

Beierlein, Davidov, Schmidt, & Schwartz (2012). Testing the discriminant validity of Schwartz' portrait value questionnaire items - A replication and extension of Knoppen and Saris (2009). Survey Research Methods, 6, 25-36.

- European Social Survey comprising 26 countries and approximately 50,000 subjects (average country size 1,900)
- 4 items measuring the concepts of tradition and conformity
- 1-factor model

Excerpts from the ESS Human Values Scale

How similar to me is the person in the portrait?

Tradition (TR):	9. It is important for him to be humble and modest. He tries not to draw attention to himself (ipmodst).
	20. Tradition is important to him. He tries to follow the customs handed down by his religion or family (imptrad).
Conformity (CO):	7. He believes that people should do what they're told. He thinks people should follow rules at all times, even when no one is watching (ipfrule).
	16. It is important for him to always behave properly. He wants to avoid doing anything people would say is wrong (ipbhprp).

Note: High value means low Tradition/Conformity.

ESS Tradition-Conformity Items: Multiple-Group CFA with Scalar Invariance

- Poor fit: ML $\chi^2(202) = 8,654$, RMSEA $= 0.148$, CFI $= 0.677$
- Many modification indices > 10:
  33 for intercepts and loadings, 56 for residual covariances
- Freeing just a few parameters will not improve model fit sufficiently: 78 of the 89 modification indices are in the 10-40 range and none higher than 142

Conclusions:

- Multiple-group CFA fails due to too many necessary model modifications; the model search easily leads to the wrong model
  - The groups cannot be compared with respect to factor means
- A new method is needed: Alignment (Asparouhov-Muthén, 2013, Web Note 18)

Multiple-Group CFA 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 and variances and choose their values 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)
   - The simplicity function F is optimized at a few large non-invariant parameters and many approximately 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)

Alignment Optimization, Continued

- 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
- This model has the same fit as the configural model:
  - Free the factor means α_j and variances ψ_j, noting that for every set of factor means and variances the same fit as the configural model is obtained with loadings λ_j and intercepts ν_j changed as:

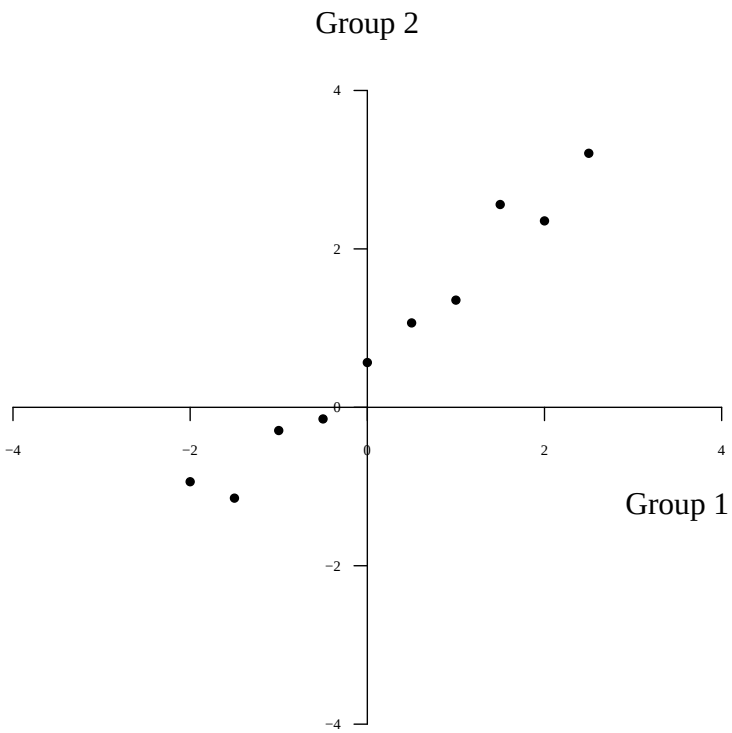
$$\lambda_j = \lambda_{j,configural} / \sqrt{\psi_j},$$
$$\nu_j = \nu_{j,configural} - \alpha_j \, \lambda_{j,configural} / \sqrt{\psi_j}.$$

- Simulation studies show that the alignment method works very well unless there is a majority of significant non-invariant parameters or small group sizes
- For well-known examples with few groups and few non-invariances, the results agree with the alignment method

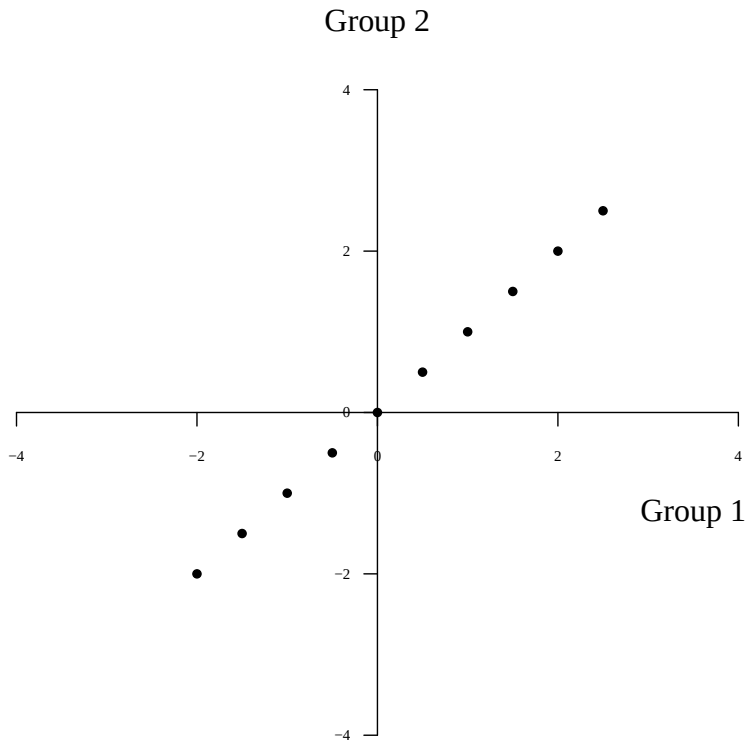
A Visual Answer to Why it is Called Alignment

Consider group-invariant intercepts for 10 items and 2 groups with factor means = 0, -1 and factor variances = 1, 2

Unaligned: Configural model (mean=0, variance=1 in both groups)



Aligned: Taking into account the group differences in means and variances



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

Input for Alignment Analysis of ESS Tradition-Conformity Items

```
DATA:       FILE = ess05Traco.dat;
VARIABLE:   NAMES = country essround ipfrule ipmodst ipbhprp imptrad;
            USEVARIABLES = ipmodst imptrad ipfrule ipbhprp;
            MISSING = ipfrule-imptrad (7-9);
            CLASSES = c(26);
            KNOWNCLASSES = c(country = 2 3 4 5 6 7 8 9 10 11 12
            13 14 15 16 17 18 21 22 23 24 25 26 27 28 30);
ANALYSIS:   TYPE = MIXTURE;
            ESTIMATOR = ML;
            ALIGNMENT = FREE;
MODEL:      %OVERALL%
            traco BY ipmodst-ipbhprp;
OUTPUT:     TECH1 TECH8 ALIGN;
PLOT:       TYPE = PLOT2;
```

Output Excerpts for ESS Tradition-Conformity Items

- STANDARD ERROR COMPARISON INDICATES THAT THE FREE ALIGNMENT MODEL MAY BE POORLY IDENTIFIED. USING THE FIXED ALIGNMENT OPTION MAY RESOLVE THIS PROBLEM.

Choose group with smallest factor mean to be the reference group (factor mean zero, factor variance 1) in a fixed alignment run:

```
ANALYSIS:    TYPE = MIXTURE;
             ESTIMATOR = ML;
             ALIGNMENT = FIXED(22);
```

ESS Tradition-Conformity Items: Approximate Measurement (Non-) Invariance for Intercepts

Groups/countries in parenthesis are non-invariant.

IPMODST	**(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)**
IMPTRAD	**(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)**
IPFRULE	**(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
IPBHPRP	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

ESS Tradition-Conformity Items: Approximate Measurement (Non-) Invariance for Loadings

IPMODST	**(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
IMPTRAD	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
IPFRULE	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
IPBHPRP	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

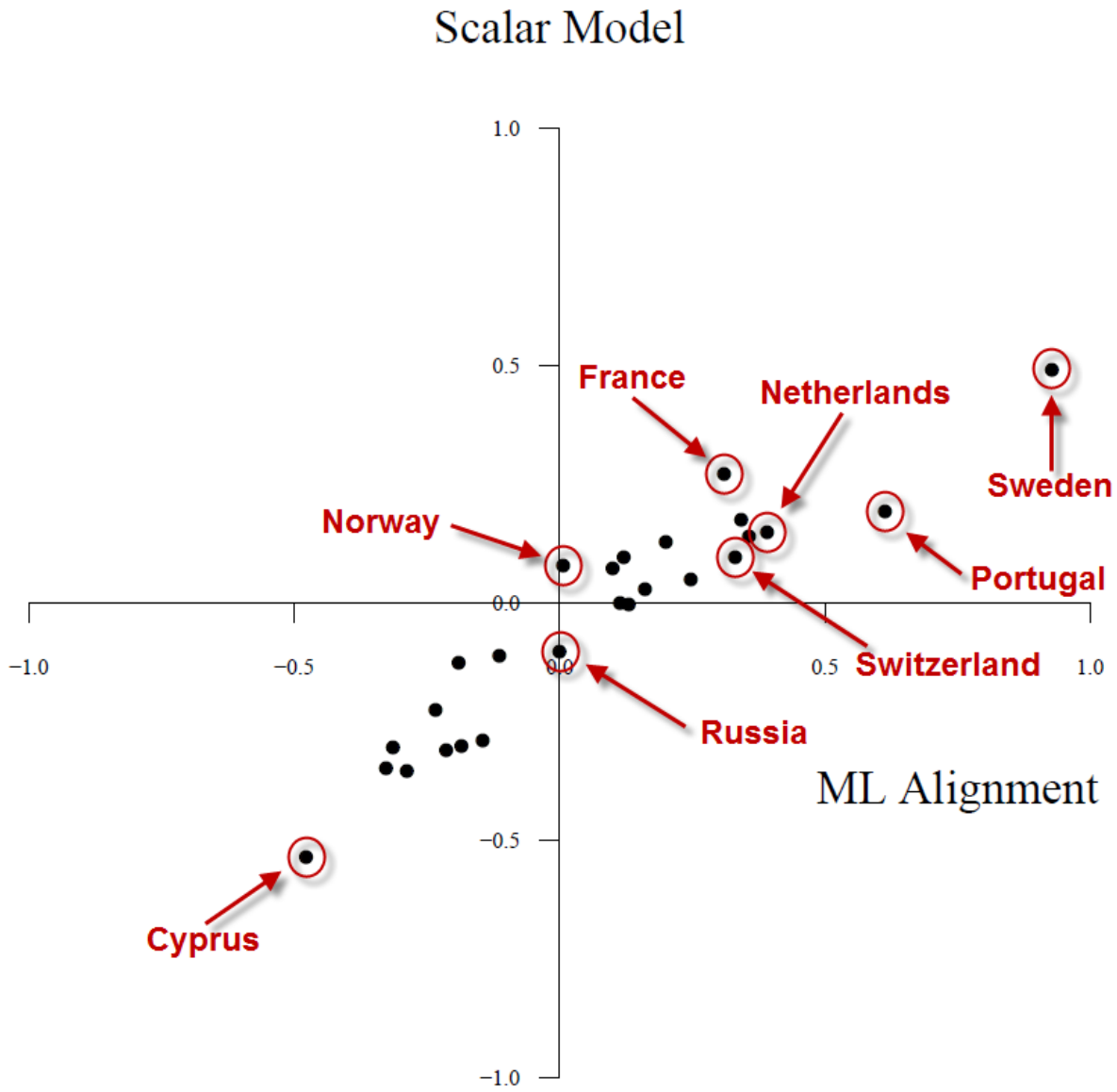
ESS Tradition-Conformity Items: Factor Mean Comparisons

Ranking	Group	Value	Groups with significantly smaller factor mean
1	23	0.928	21 18 6 10 3 11 26 7 5 16 8 1 12 19 22 14 20 25 15 17 9 2 13 24 4
2	21	0.613	18 6 10 3 11 26 7 5 16 8 1 12 19 22 14 20 25 15 17 9 2 13 24 4
3	18	0.391	26 7 5 16 8 1 12 19 22 14 20 25 15 17 9 2 13 24 4
4	6	0.357	26 7 5 16 8 1 12 19 22 14 20 25 15 17 9 2 13 24 4
5	10	0.342	7 5 16 8 1 12 19 22 14 20 25 15 17 9 2 13 24 4
6	3	0.331	7 5 16 8 1 12 19 22 14 20 25 15 17 9 2 13 24 4
7	11	0.310	5 16 8 1 12 19 22 14 20 25 15 17 9 2 13 24 4
8	26	0.247	16 8 1 12 19 22 14 20 25 15 17 9 2 13 24 4
9	7	0.200	12 19 22 14 20 25 15 17 9 2 13 24 4
10	5	0.161	19 22 14 20 25 15 17 9 2 13 24 4
11	16	0.130	19 22 14 20 25 15 17 9 2 13 24 4
12	8	0.121	19 22 14 20 25 15 17 9 2 13 24 4
13	1	0.114	19 22 14 20 25 15 17 9 2 13 24 4

ESS Tradition-Conformity Items: Factor Mean Comparison Continued

Ranking	Group	Value	Groups with significantly smaller factor mean
14	12	0.100	22 14 20 25 15 17 9 2 13 24 4
15	19	0.007	14 20 25 15 17 9 2 13 24 4
16	22	0.000	14 20 25 15 17 9 2 13 24 4
17	14	-0.114	17 9 2 13 24 4
18	20	-0.145	9 2 13 24 4
19	25	-0.185	2 13 24 4
20	15	-0.190	2 13 24 4
21	17	-0.214	13 24 4
22	9	-0.234	13 24 4
23	2	-0.288	4
24	13	-0.314	4
25	24	-0.327	4
26	4	-0.478	

Factor Mean Ordering: Scalar Model vs Alignment



How Do We Know That We Can Trust The Alignment Results? Monte Carlo Studies

- Simulations in Asparouhov-Muthén Web Note 18
- Simulations based on the estimated model:
  - Request SVALUES for real-data alignment run (parameter estimates arranged as starting values)
  - Do a Monte Carlo run with these parameter values as population values, choosing the sample size and check parameter bias, SE bias, and the coverage
  - Do a ”real-data” run on Monte-Carlo generated data from one or more replications to study the measurement invariance assessment

Input for Alignment Monte Carlo Study

- Copy SVALUES results from real-data run into Monte Carlo run
- Do a global change of the class label ”c” to ”g” (reverse unwanted changes: Montegarlo, Progessors, etc)
- Change f BY in OVERALL to give starting values

```
MONTECARLO:     NAMES = ipfrule ipmodst ipbhprp imptrad;
                NGROUPS= 26;
                NOBSERVATIONS = 26(2000);
                NREPS = 100;
                REPSAVE = ALL;
                SAVE = n2000f-22rep*.dat;
ANALYSIS:       TYPE = MIXTURE;
                ESTIMATOR = ML;
                ALIGNMENT = FIXED(22);
                PROCESSORS = 8;
MODEL POPULATION:
                %OVERALL%
                traco BY ipfrule-imptrad*1;
                [ g#1*-0.10053 ];
                etc
```

Monte Carlo Results for ESS Tradition-Conformity Items

Average group size in real data: 1900.

- Group size $n = 100 - 300$: Ok results, but not great
- Group size $n = 1000 - 2000$: Good results
- Group size $n = 10,000$: Great results

Example 2: PISA Teacher-Student Relationship Items

Cheung (2013). Presentation at the IACCP Regional Conference in LA (June 20-22) on cross-cultural research.

- PISA (Program for International Student Assessment) 2009 study, 53 countries, 420,000 15-year olds
- 5 items measuring teacher-student relationships
- 1-factor model

PISA Teacher-Student Relationship Items

Teacher-student Relationships Measure (PISA 2009 Student Questionnaire)
1=strongly disagree, 4=strongly agree

---

1. I get along with most of my teachers
2. Most of my teachers are interested in my well-being
3. Most of my teachers really listen to what I have to say
4. If I need extra help, I will receive it from my teachers
5. Most of my teachers treat me fairly

Approximate Measurement (Non-) Invariance - PISA

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 41 42 43 **(44) (45)** 46 47 **(48)** 49 **(50)** 51 **(52) (53) (54) (55)** 56
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) (41)** 42 **(43)** 44 45 **(46)** 47 48 **(49)** 50 **(51) (52) (53)** 54 55 56
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) (41)** 42 43 **(44)** 45 **(46)** 47 48 **(49) (50) (51) (52) (53) (54) (55) (56)**
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 **(41)** 42 **(43) (44)** 45 **(46) (47) (48) (49)** 50 **(51)** 52 **(53) (54) (55) (56)**
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)** 41 **(42) (43)** 44 **(45)** 46 47 48 **(49) (50) (51) (52)** 53 **(54) (55) (56)**

Approximate Measurement (Non-) Invariance - PISA

Loadings

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 41 42 43 44 45 46 47 48 **(49)** **(50)** 51 52 53 54 **(55)** 56
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 **(41)** 42 **(43)** 44 45 **(46)** 47 48 49 50 51 52 53 54 55 56
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 41 **(42)** 43 44 **(45)** 46 **(47)** 48 **(49)** 50 51 52 53 54 **(55)** 56
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 41 42 43 44 45 46 47 **(48)** **(49)** 50 51 52 **(53)** 54 55 56
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 41 42 43 44 45 46 **(47)** 48 49 50 51 52 53 54 **(55)** 56

PISA Teacher-Student Relationship Items: Monte Carlo Simulations

Monte Carlo simulations based on these data show failure in recovering the population values - too large degree of non-invariance.

Example 3: PISA Binary Math Items

- Items from the PISA (Program for International Student Assessment) survey of 2003
- A total of 9796 students from 40 countries
- Analyzed by Fox (2010). Bayesian Item Response Modeling
- A 40-group, one-factor model for eight mathematics test items
- 2-parameter probit IRT model that accommodates country measurement non-invariance for all difficulty (threshold) and discrimination (loading) parameters as well as country-specific factor means and variances

Input for PISA Alignment

```
DATA:       FILE = pisa2003.dat;
VARIABLE:   NAMES = cn y1-y8;
            CATEGORICAL = y1-y8; ! Requires Bayesian analysis
            USEVARIABLES = y1-y8;
            MISSING = y1-y8(9);
            CLASSES = c(40);
            KNOWNCLASS = c(cn = 1-40);
ANALYSIS:   TYPE = MIXTURE;
            ESTIMATOR = BAYES;
            PROCESSORS = 2;
            ALIGNMENT = FREE;
            THIN = 10;
            BITERATIONS = (5000);
MODEL:      %OVERALL%
            f BY y1-y8;
OUTPUT:     TECH1 TECH8 ALIGN;
PLOT:       TYPE = PLOT2;
```

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

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

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

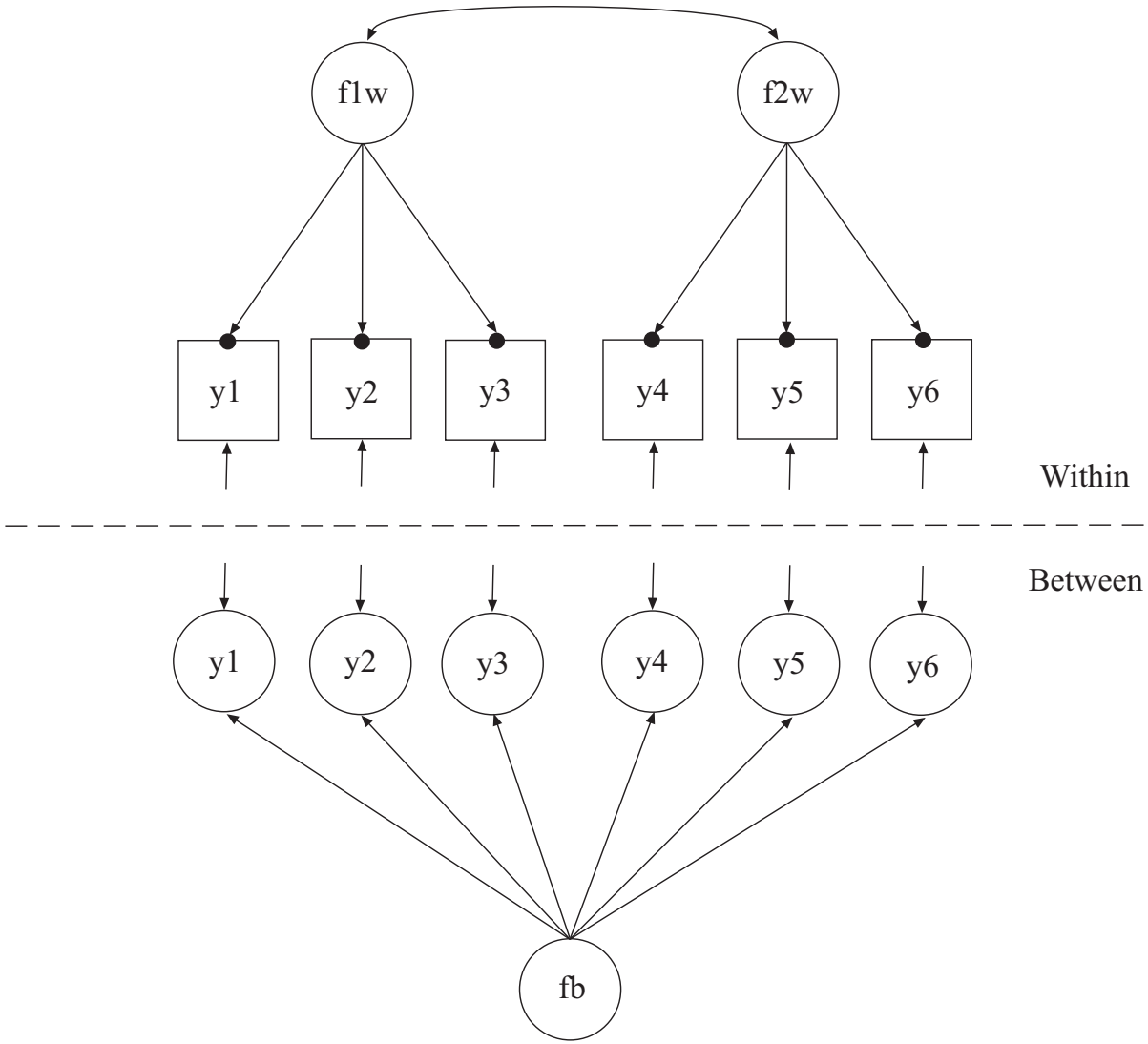
The two-level factor analysis model

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

can be viewed as a random intercept model:

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

Random Intercept Two-Level Factor Analysis in Figure Form



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

- Random intercept two-level factor analysis:

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

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

$$\begin{aligned} y_{ij} &= \nu + \lambda f_{ij} + \varepsilon_{ij}, \\ f_{ij} &= f_{B_j} + f_{W_{ij}}, \end{aligned}$$

- 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 4: 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 Θ_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

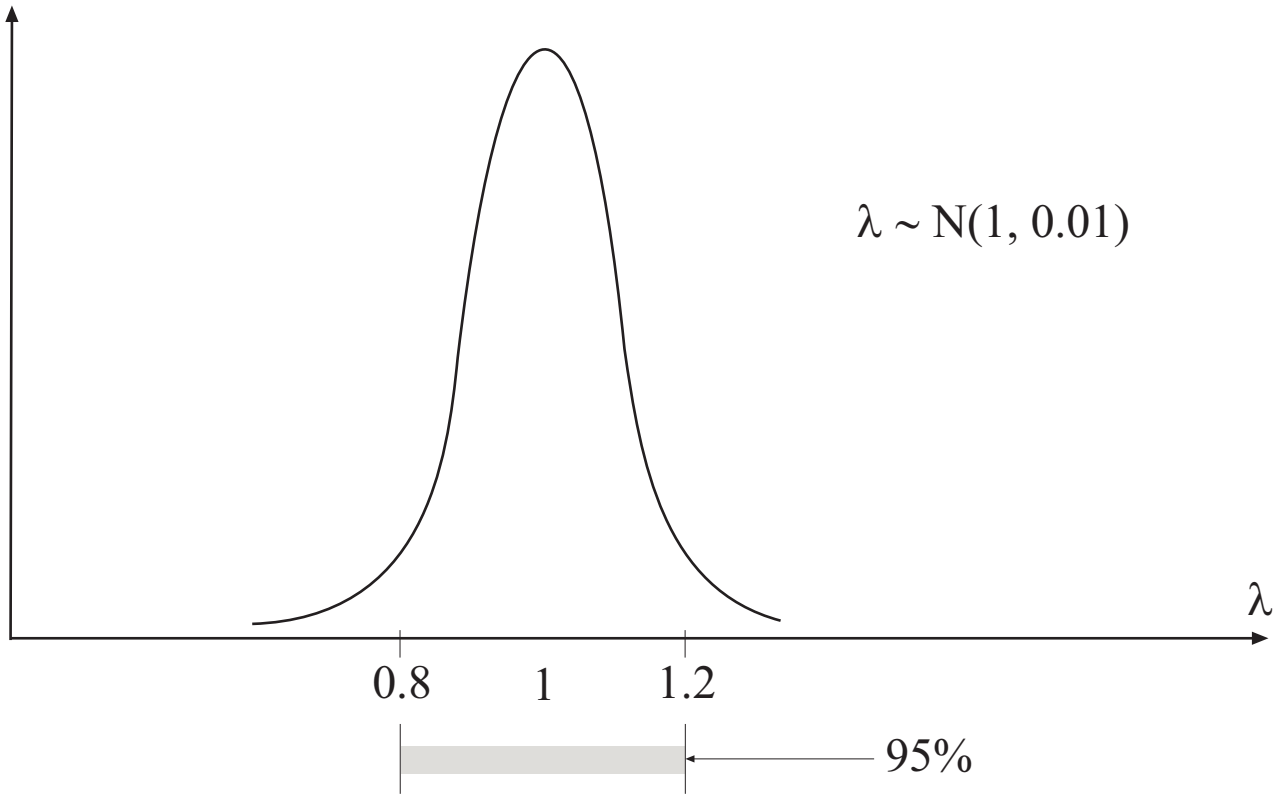
Two-Level Analysis with Random Item Parameters

- 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.

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



Random Item Parameters In IRT

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

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

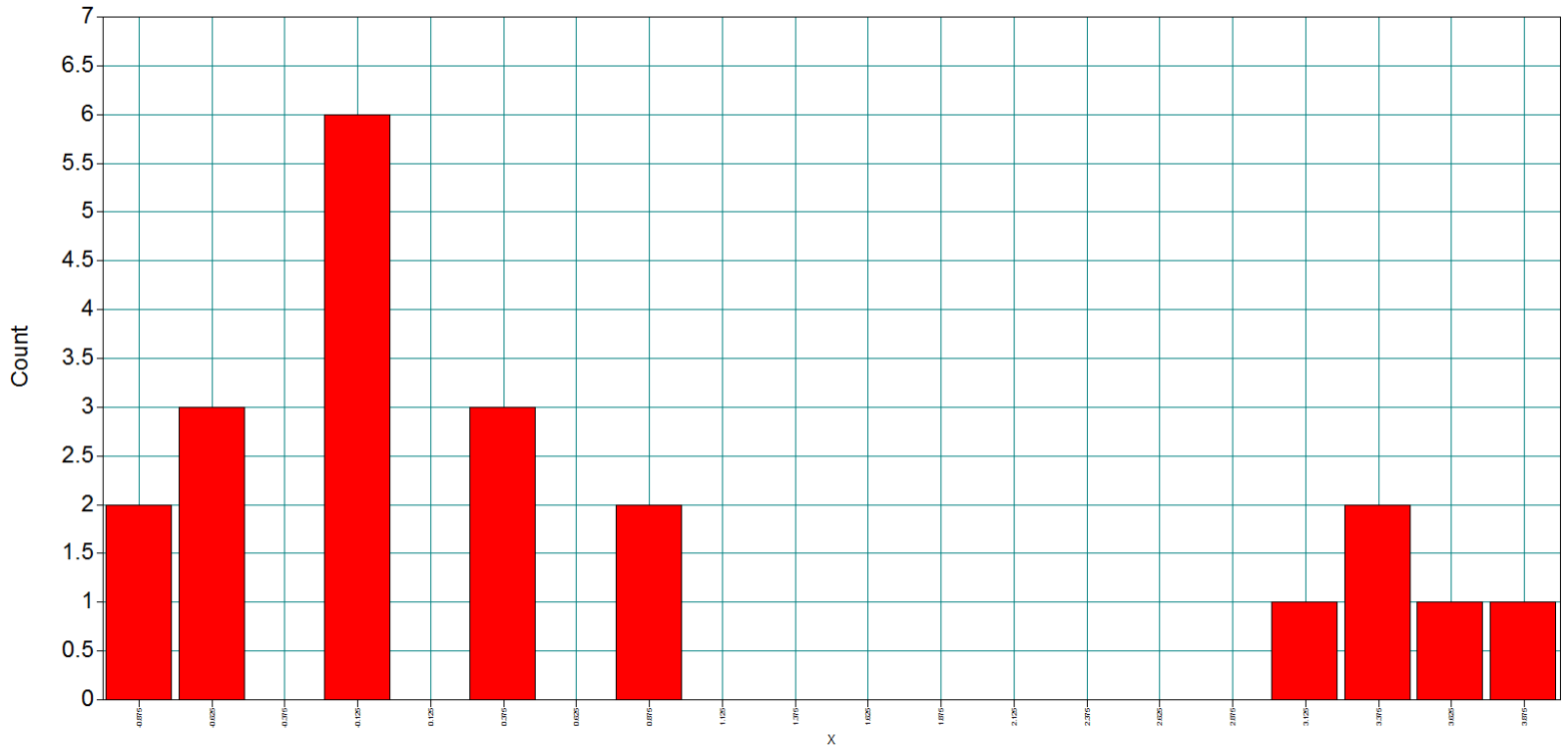
$$a_{jk} \sim N(a_k, \sigma_{a,k}), b_{jk} \sim N(b_k, \sigma_{b,k})$$

  This is a 2-parameter probit IRT model where both discrimination (a) and difficulty (b) vary across country
- The θ ability factor is decomposed as

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

- The mean and variance of the ability vary across country
- Model preserves common measurement scale while accommodating measurement non-invariance
- The ability for each country obtained by factor score estimation

Violations of the Measurement Non-Invariance Normality Assumption for the Item Parameter Distributions



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
- Handles small group/cluster sizes
- Easy to relate measurement non-invariance to variables on the group/cluster level (Jak et al., 2013b)

Fit of the Model in Each Group/Cluster: How Important is Model Misfit?

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 - and that is often the case
  - Fit judged by ML χ^2 or Bayes Posterior Predictive Checking
- Random mode: Two-level factor analysis does not automatically judge fit in each cluster

What does Bayes contribute?

The Several Uses of BSEM

Using zero-mean, small-variance priors for parameters not identifiable in ML.

- 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

BSEM Input for ESS Tradition-Conformity Items Adding Residual Covariances

```
ANALYSIS:          TYPE = MIXTURE;
                   ESTIMATOR = BAYES;
                   PROCESSORS = 2;
                   THIN = 10;
                   BITERATIONS = (1000);
                   ALIGNMENT = FIXED(4);
MODEL:             %OVERALL%
                   traco BY ipmodst-ipbhprp;
                   ipmodst-ipbhprp (p#_1-p#_4);
                   ipmodst-ipbhprp WITH ipmodst-ipbhprp;
MODEL PRIORS:      DO(1,26) p#_1~iw(50,50);
                   DO(1,26) p#_2~iw(50,50);
                   DO(1,26) p#_3~iw(50,50);
                   DO(1,26) p#_4~iw(50,50);
```

ESS Tradition-Conformity Items: ML, Bayes and BSEM Invariance Testing

- ML scalar model:
  $\chi^2(202) = 8,654$ (p=0.000), RMSEA $= 0.148$, CFI $= 0.677$
- ML alignment - fit of configural model:
  $\chi^2(52) = 317$ (p=0.000), RMSEA $= 0.052$, CFI $= 0.990$
- Bayes Posterior Predictive Checking: 95% CIs for the difference between observed and replicated χ^2 values and Posterior Predictive p-values:
  - Bayes Alignment - fit of configural model:
    [194, 324], PPP=0.000
  - Bayes Alignment allowing for residual covariances (BSEM):
    [-62, 99], PPP=0.458

For the last model, there are only a few significant residual covariances.

Approximate Measurement (Non-) Invariance for ESS Items

Intercepts

IPMODST	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
IMPTRAD	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
IPFRULE	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
IPBHPRP	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

Loadings

IPMODST	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
IMPTRAD	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
IPFRULE	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
IPBHPRP	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

Factor Mean Comparison: ESS Tradition-Conformity Items

Factor mean comparison at the 5 % significance level in descending order

Ranking	Group	Value	Groups with significantly smaller factor mean
1	23	1.827	26 7 5 8 16 1 12 19 22 14 20 15 9 25 17 2 13 24 4
2	21	1.228	22 14 20 15 9 25 17 2 13 24 4
3	18	1.217	16 1 22 14 20 15 9 25 17 2 13 24 4
4	10	1.178	5 16 1 12 22 14 20 15 9 25 17 2 13 24 4
5	6	1.172	16 1 22 14 20 15 9 25 17 2 13 24 4
6	3	1.123	22 14 20 15 9 25 17 2 13 24 4
7	11	1.118	22 14 20 15 9 25 17 2 13 24 4
8	26	1.030	22 14 20 15 9 25 17 2 13 24 4
9	7	0.955	20 15 9 25 17 2 13 24 4
10	5	0.913	14 20 15 9 25 17 2 13 24 4
11	8	0.900	14 20 15 9 25 17 2 13 24 4
12	16	0.868	20 15 9 25 17 2 13 24 4
13	1	0.855	20 15 25 17 2 13 24 4

Factor Mean Comparison: ESS Tradition-Conformity Items, Continued

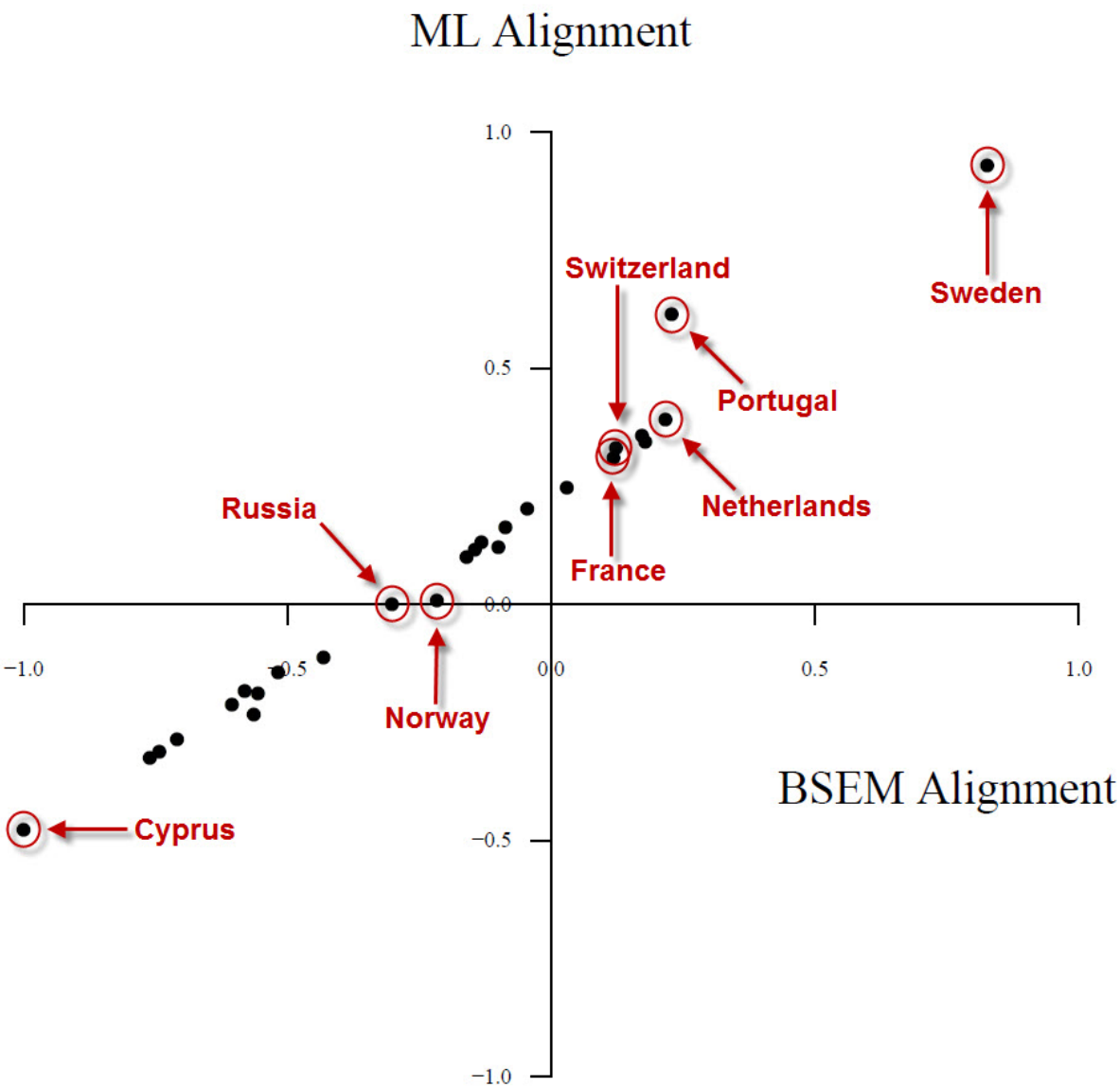
Factor mean comparison at the 5 % significance level in descending order

Ranking	Group	Value	Groups with significantly smaller factor mean
14	12	0.840	20 15 9 25 17 2 13 24 4
15	19	0.783	20 25 17 2 13 24 4
16	22	0.698	20 25 17 2 13 24 4
17	14	0.568	2 13 24 4
18	20	0.482	2 13 4
19	15	0.444	13 4
20	9	0.436	4
21	25	0.419	4
22	17	0.395	4
23	2	0.291	4
24	13	0.257	4
25	24	0.239	4
26	4	0.000	

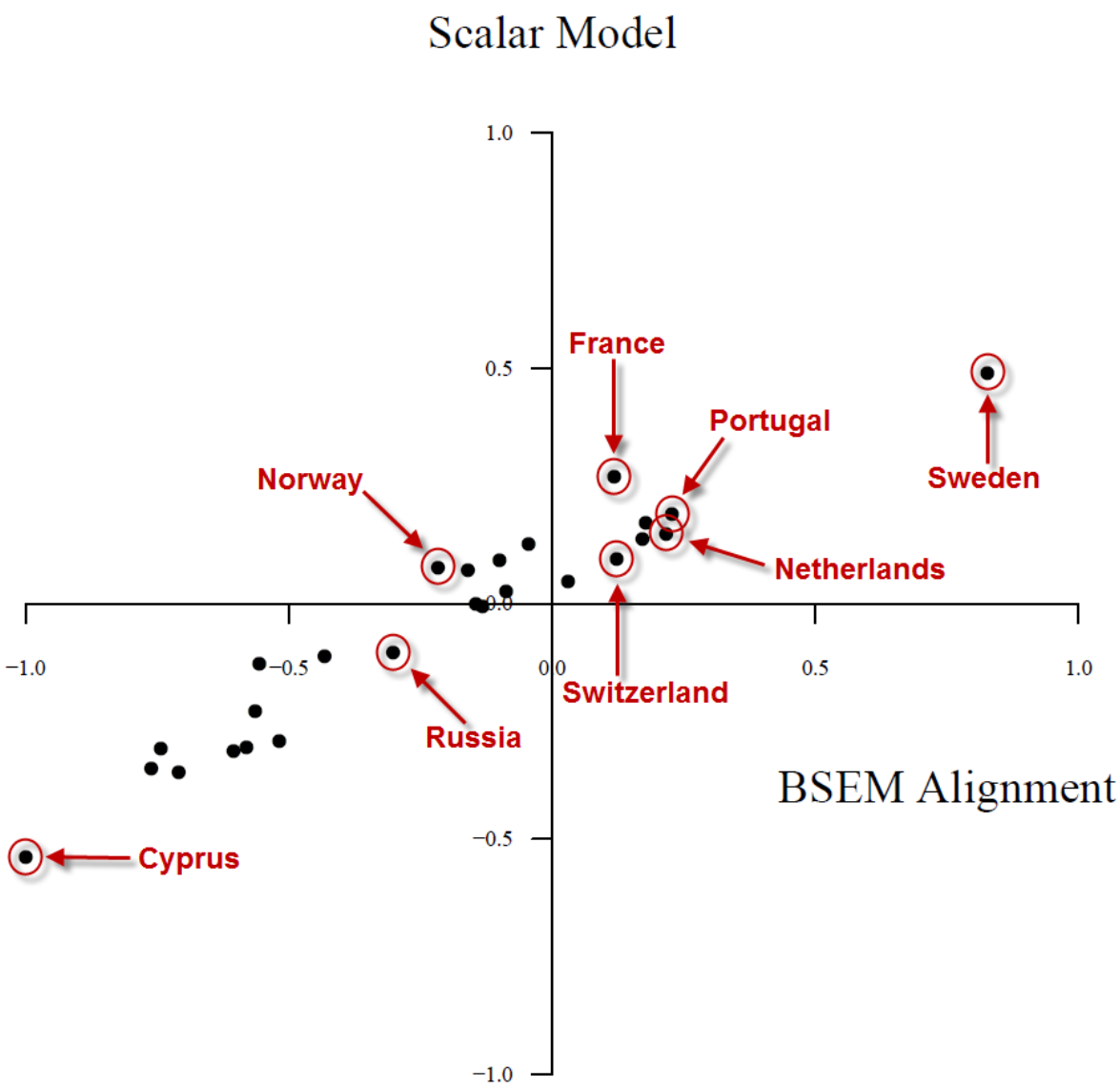
Summary of ESS Tradition-Conformity Items Alignment

- The configural model fits well only when using BSEM to allow for a small degree of residual covariance
- The residual covariances absorb the influence of minor factors, the strength of which vary across countries, while the major factor is sufficiently invariant
- Most countries have negligible residual covariances
- There are only 10 significant residual covariances and they are small
- Monte Carlo simulation of data generated by the model allowing residual covariances, while ignored in the ML analysis, points to the same non-invariance picture as seen in the initial real-data analysis:
  - Residual covariances can be mistaken for non-invariance in intercepts and loadings

Sensitivity Analysis: How Much Difference Does it Make to Allow BSEM Residual Covariances?



How Does the Scalar Model Compare to BSEM Alignment?



A Second Look at PISA Teacher-Student Relationship Items using BSEM with Residual Covariances

- Allowing residual covariances does not change the large degree of non-invariance
- The PISA teacher-student relationship items cannot be aligned in a trustworthy way

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