Bayesian Measurement Invariance Testing

Jean-Paul Fox*

University of Twente
* Department of Research Methodology, Measurement and Data Analysis
Faculty of Behavioural, Management, and Social Sciences
Enschede, Netherlands

Overview

Overview

1. Bayesian Measurement Invariance Testing
2. Conditional Modelling/Testing Approach
   - Random Item Parameters

Overview

1. Bayesian Measurement Invariance Testing
2. Conditional Modelling/Testing Approach
   - Random Item Parameters
3. Marginal Modelling/Testing Approach
   - Random Item Parameters

Overview

Measurement Invariance

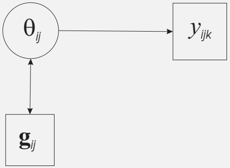
Conditional methods (Scheuneman, 1979; Lord, 1980; Mellenbergh, 1982)

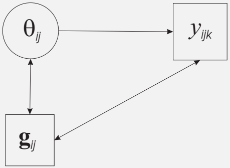
$$P\left(Y=1 \mid \theta, g\right) \quad = \quad P\left(Y=1 \mid \theta\right)$$

for all g, θ

Measurement Invariance

Graphical display of Measurement (In)Variance





Conditional DIF Detection

Distributional assumptions (Thissen et al, 1986; Reise et al, 1993)

1. Evaluate fit of item response model in each group

Conditional DIF Detection

Distributional assumptions (Thissen et al, 1986; Reise et al, 1993)

1. Evaluate fit of item response model in each group
2. Different population distributions, same item curves, simultaneous analysis

Conditional DIF Detection

Distributional assumptions (Thissen et al, 1986; Reise et al, 1993)

1. Evaluate fit of item response model in each group
2. Different population distributions, same item curves, simultaneous analysis
3. Full measurement invariance versus partial measurement invariance

Conditional DIF Detection

Distributional assumptions (Thissen et al, 1986; Reise et al, 1993)

1. Evaluate fit of item response model in each group
2. Different population distributions, same item curves, simultaneous analysis
3. Full measurement invariance versus partial measurement invariance
4. Inspection of item parameters for DIF detection

Bayesian MI Detection

Procedure

- Bayesian Modeling of MI violation(s)
- Obtain (MCMC) samples from posterior distributions
- Bayesian MI Tests (BF Testing)

Advantages

- Quantify evidence in favor of (null) hypothesis
- Avoid use of anchor items
- Simultaneously testing of multiple hypotheses
- Avoid identification issues

Overview

Random Item Parameter

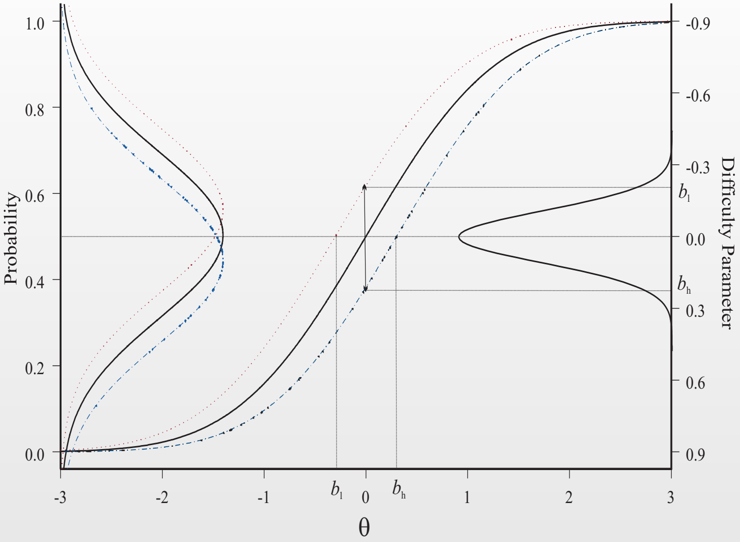


Figure:

Measurement invariance in cross-national surveys

Persons from each country with the same value of the underlying construct should have the same probability of endorsing the item.

$$P\left(Y=1 \mid \theta, g\right) \quad = \quad P\left(Y=1 \mid \theta\right) \text{ for all } g, \theta$$

Two parameter normal ogive IRT model:

$$P\left(Y=1 \mid \theta, a, b\right) \quad = \quad \Phi\left(a\theta - b\right)$$

1. Items should have the same difficulty/endorsement level in all countries
2. Items should discriminate equally in all countries

When measurement invariance is not present or when the variance is not modeled, individual test scores cannot be compared across countries

Issues with Many groups

1. Specifying the population distribution (number of parameters)

Issues with Many groups

1. Specifying the population distribution (number of parameters)
2. Finding anchor items

Issues with Many groups

1. Specifying the population distribution (number of parameters)
2. Finding anchor items
3. DIF in difficulty and discrimination parameters.
   (Tests for partial measurement invariance)

Random item effects approach

N persons in J groups answering to K binary $(0/1)$ items:

$$P\left(Y_{ijk} = 1 \mid \theta_{ij}, \tilde{a}_{kj}, \tilde{b}_{kj}\right) = \Phi\left(\tilde{a}_{kj}\theta_{ij} - \tilde{b}_{kj}\right)$$

The country specific item parameters are modeled as random effects parameters, assumed to be normally distributed around international item parameters:

$$\begin{aligned} \tilde{a}_{kj} \mid a_k, \sigma^2_{a_k} &\sim \mathcal{N}\left(a_k, \sigma^2_{a_k}\right) \\ \tilde{b}_{kj} \mid b_k, \sigma^2_{b_k} &\sim \mathcal{N}\left(b_k, \sigma^2_{b_k}\right) \end{aligned}$$

Random item effects approach

N persons in J groups answering to K binary $(0/1)$ items:

$$P\left(Y_{ijk} = 1 \mid \theta_{ij}, \tilde{a}_{kj}, \tilde{b}_{kj}\right) = \Phi\left(\tilde{a}_{kj}\theta_{ij} - \tilde{b}_{kj}\right)$$

The country specific item parameters are modeled as random effects parameters, assumed to be normally distributed around international item parameters:

$$\begin{aligned} \tilde{a}_{kj} \mid a_k, \sigma^2_{a_k} &\sim \mathcal{N}\left(a_k, \sigma^2_{a_k}\right) \\ \tilde{b}_{kj} \mid b_k, \sigma^2_{b_k} &\sim \mathcal{N}\left(b_k, \sigma^2_{b_k}\right) \end{aligned}$$

Person parameters are assumed to be normally distributed around group means with country specific variance:

$$\theta_{ij} \mid \mu_j, \sigma^2_{\theta_j} \sim \mathcal{N}\left(\mu_j, \sigma^2_{\theta_j}\right)$$

Mplus Model

EXAMPLE 9.26: IRT WITH RANDOM BINARY ITEMS USING CROSS-CLASSIFIED DATA

```
TITLE:      this is an example of IRT with random
            binary items using cross-classified
            data
DATA:       FILE = ex9.26.dat;
VARIABLE:   NAMES = u subject item;
            CATEGORICAL = u;
            CLUSTER = item subject;
ANALYSIS:   TYPE = CROSSCLASSIFIED RANDOM;
            ESTIMATOR = BAYES;
            PROCESSORS = 2;
MODEL:      %WITHIN%
            %BETWEEN subject%
            s | f BY u;
            f@1;
            u@0;
            %BETWEEN item%
            u; [u$1];
            s; [s];
OUTPUT:     TECH1 TECH8;
```

Figure:

Testing invariance: Random item effects approach

Hypotheses:

- Item parameter variances are zero
  H0: $\sigma^2_{a_k} = 0$, $\sigma^2_{b_k} = 0$
- Country specific item parameters equal to international item parameters
  H_0: for each country j, b_{kj}=b_k and a_{kj}=a_k

Advantages:

1. No anchor items needed
   Restrictions: for each j, $\sum_k \tilde{b}_{kj} = 0$ for each j, $\prod_k \tilde{a}_{kj} = 1$
2. All invariance assumptions can be tested simultaneously without conditioning on other invariance assumptions
3. Only most general model needs to be estimated for BF tests

Bayes factor

Encompassing prior approach

- $H_0 : \sigma^2 = 0$ on boundary of parameter space for variances
- Choose value δ that represents "no relevant variance"
  Evaluate H0: $\sigma^2 < \delta$
- Prior under M0 is a restriction of the encompassing prior under M1:

$$\mathrm{BF} = \frac{P(\sigma^2 < \delta | \mathbf{y}, M1)}{P(\sigma^2 < \delta | M1)}$$

Simulation study: 20 items, 40 groups each 200 members

H_0	BF>3	$P(H_0 \mid \mathbf{y})$	BF>3	$P(H_0 \mid \mathbf{y})$	BF>3	$P(H_0 \mid \mathbf{y})$
	$\delta < .0016$		$\delta < .0025$		$\delta < .0036$	
$\sigma^2_{a_k} = 0$, invariant item discrimination						
$.00$	0.92	0.92	0.99	0.97	0.99	0.97
$.04$	0.05	0.08	0.07	0.15	0.04	0.18
$.07$	0.01	0.02	0.01	0.08	0.01	0.11
$.10$	0.01	0.03	0.02	0.07	0.00	0.09
$.13$	0.01	0.03	0.02	0.07	0.00	0.09
$\sigma^2_{b_k} = 0$, invariant item difficulty						
$.00$	0.99	0.99	1.00	0.99	1.00	0.99
$.05$	0.00	0.00	0.00	0.00	0.00	0.00
$.10$	0.00	0.00	0.00	0.00	0.00	0.00
$.20$	0.00	0.00	0.00	0.00	0.00	0.00
$.30$	0.00	0.00	0.00	0.00	0.00	0.00

ESS attitude towards immigrants

	Greece	Norway	Poland	Sweden
Not allow immigrants poor countries	87 %	39%	44%	15%
Make worse country	67%	39%	27%	17%
Take away jobs	78%	19%	52%	12%

ESS attitude towards immigrants

	Greece	Norway	Poland	Sweden
Not allow immigrants poor countries	87 %	39%	44%	15%
Make worse country	67%	39%	27%	17%
Take away jobs	78%	19%	52%	12%

ESS attitude towards immigrants

	Greece	Norway	Poland	Sweden
Not allow immigrants poor countries	87 %	39%	44%	15%
Make worse country	67%	39%	27%	17%
Take away jobs	78%	19%	52%	12%

ESS attitude towards immigrants

	Greece	Norway	Poland	Sweden
means	1.01	-0.42	-0.36	-1.09
Not allow immigrants poor countries	87 %	39%	44%	15%
Make worse country	67%	39%	27%	17%
Take away jobs	78%	19%	52%	12%

ESS attitude towards immigrants

	Greece	Norway	Poland	Sweden
means	1.01	-0.42	-0.36	-1.09
Allow from poor countries	87 %	39%	44%	15%
b_{kj}	-0.69	-0.03	-0.11	0.29
Make worse country	67%	39%	27%	17%
b_{kj}	0.52	-0.06	0.47	0.17
Take away jobs	78%	19%	52%	12%
b_{kj}	-0.18	0.76	-0.35	0.67

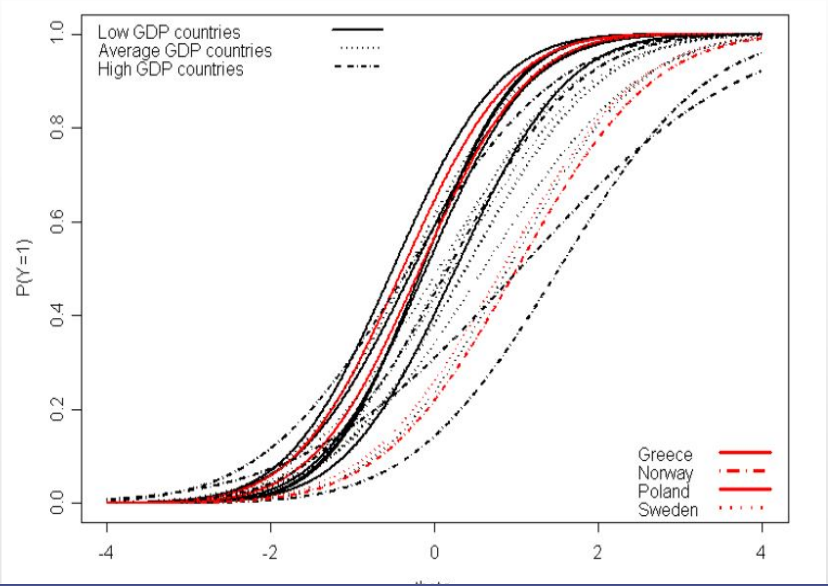
ESS covariates

Table: Covariate coefficients for a and b

	δ_{a_k}			δ_{b_k}		
	% immig	% unemp	GDP	% immig	% unemp	GDP
allow poor countries	-0.02	0.04	*0.15*	-0.03	0.04	0.03
allow same ethnicity	0.07	0.01	0.07	0.05	0.03	-0.02
worse country	-0.02	-0.01	0.03	0.00	0.05	0.01
bad for economy	*-0.11*	0.01	-0.04	0.10	-0.01	-0.02
undermine culture	0.02	-0.08	-0.01	0.00	0.03	0.00
take jobs away	0.01	-0.02	**-0.15**	**-0.18**	-0.06	**0.32**
take out more	0.06	0.01	-0.08	0.01	-0.06	*-0.17*
worse crime rate	-0.01	0.04	0.05	0.05	-0.03	-0.15

$Pr(\beta > 0) > .90$ $Pr(\beta > 0) > \mathbf{.95}$

ESS covariates



Overview

Marginal IRT Modeling

Conditional IRT Model

In the conditional model, responses are modeled conditional on the unknown person parameter

$$\begin{aligned} P\left(Y_{ik}=1 \mid \theta_i, b_k\right) &= \Phi\left(\theta_i - b_k\right) \\ \theta_i &\sim \mathcal{N}\left(\mu_\theta, \tau\right) \end{aligned}$$

Marginal IRT Model

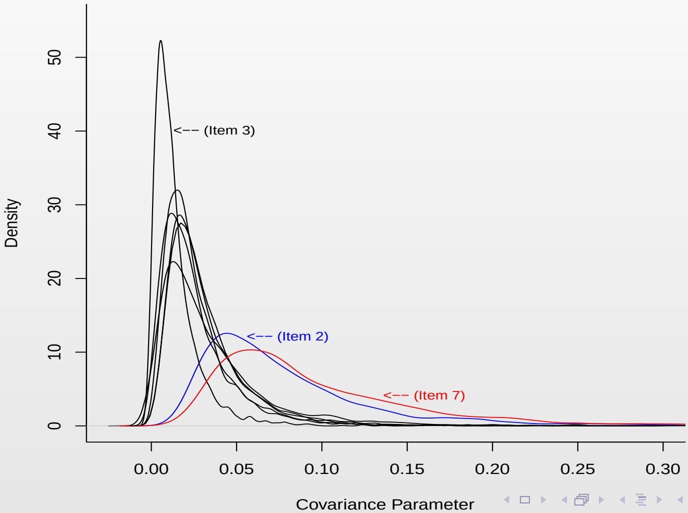
When integrating out the random person parameter, responses are assumed to be multivariate distributed

$$P\left(Y_{ik}=1 \mid b_k\right) = \Phi\left(\mu_\theta - b_k\right)$$

Application PISA 2003: 8 items, 40 countries, n=250

Item	Difficulty		Covariance		Balanced Prior	Reference Prior
Number j	Mean	SD	Mean	SD	$\sigma_{b_j} \leq 0, \sigma_{b_j} > 0$	$\sigma_{b_j} \leq 0, \sigma_{b_j} > 0$
1	-.77	.06	.03	.03	-20.44	-16.79
2	-.04	.10	.08	.06	-59.09	-51.21
3	-.20	.05	.01	.02	-7.66	-5.65
4	-.53	.06	.03	.02	-16.54	-13.35
5	-.13	.06	.03	.03	-20.73	-17.04
6	-1.65	.08	.04	.04	-23.95	-19.98
7	-.93	.11	.10	.07	-75.03	-65.46
8	-1.01	.06	.03	.03	-16.36	-13.23

Application PISA 2003: 8 items, 40 countries, n=250



Simulation Study Setup

Simulation 1

- N=1000
- J=20 (m=N/J=50)
- K=10
- XG=5000 (burn-in=1000)
- Number of replications = 40

Simulation 2

- N=500
- J=2 (m=N/J=250)
- K=10
- XG=5000 (burn-in=1000)
- Number of replications = 40

Simulation Study: Measurement Variance of J=20 Groups

item	τ	τ'	τ - τ'	BF_{10}
1	-0.020	0.004	-0.024	0.664
2	0.000	0.018	-0.018	2.477
3	0.025	0.038	-0.013	135.504
4	0.050	0.057	-0.007	$>$150
5	0.075	0.074	0.001	$>$150
6	0.100	0.103	-0.003	$>$150
7	0.125	0.124	0.001	$>$150
8	0.150	0.146	0.004	$>$150
9	0.175	0.163	0.012	$>$150
10	0.200	0.193	0.007	$>$150

Simulation Study: Measurement Variance of J=2 Groups

item	τ	τ'	τ - τ'	BF_{10}
1	-0.004	0.036	-0.040	1.660
2	0.000	0.037	-0.037	1.685
3	0.025	0.064	-0.039	5.686
4	0.050	0.067	-0.017	7.214
5	0.075	0.070	0.005	8.750
6	0.100	0.099	0.001	$>$150
7	0.125	0.089	0.036	57.628
8	0.150	0.097	0.053	$>$150
9	0.175	0.112	0.063	$>$150
10	0.200	0.111	0.089	$>$150

Discussion

Summary

- Bayesian Marginal MI Testing (no identification issues, simultaneous testing, no-step-wise procedures, no anchors,...)
- BF Tests (simple computations, flexible priors)

Future Research

- Cross-classified Structures
- Unbalanced designs are more complex
- Extensions to more complex (multilevel) models
- Extensions to deal with other data types