Artur Pokropek
artur.pokropek@gmail.com

Assessing Item Measurement Invariance in Large Scale Cross-Country Surveys. Monte Carlo Simulation Study

Mini Conference Structural Equation Modeling, Zurich, on 7-8 April 2016

Background

Particularly in large-scale cross-cultural studies the lack of convergence of information provided by the common fit statistics, combined with the ***absence of adequate Monte Carlo studies*** *and experience with fit statistics in similar cases, can create problems in choosing the most adequate model* (van de Vijver 2011: 28)

Simulations are needed to provide a more informative recommendation for how applied researchers should handle partial measurement equivalence when full equivalence is not given (...) ***Further studies using survey data and simulations are necessary*** *to test the applicability and usefulness of these new methods for applied researchers* (Davidov et al., 2014, p. 68)

Background (II)

1. **Scales comparability in large scale cross-country surveys**
    1. Test procedures for detecting non-comparable items in large scale cross-national analyses.
    2. Determine the boundary conditions for making meaningful comparisons for different scales and methods.
    3. Test approximate measurement invariance approaches and compare them with partial equivalence approach.

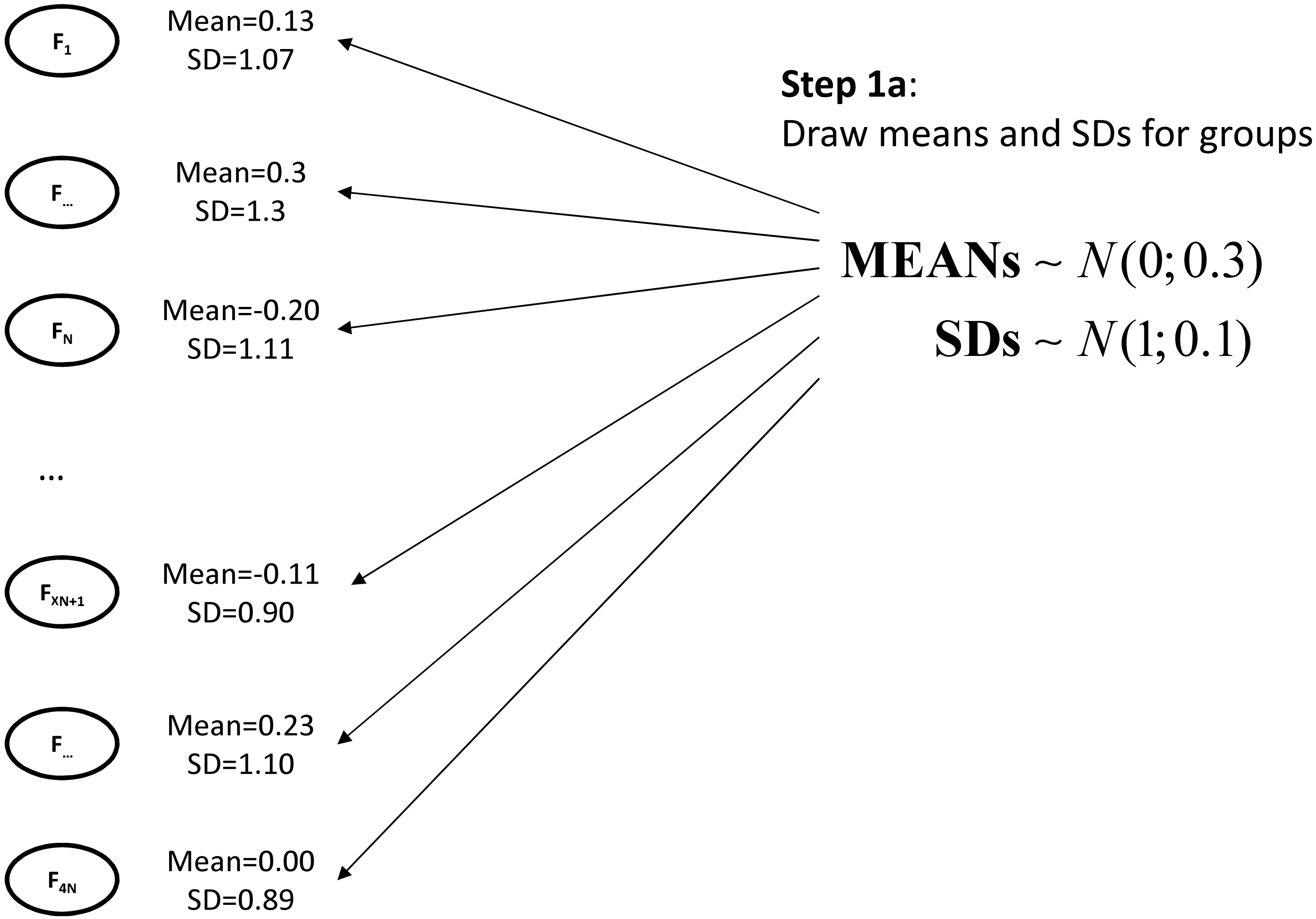
Research Grant
2016-2018

National Science Centre
Poland

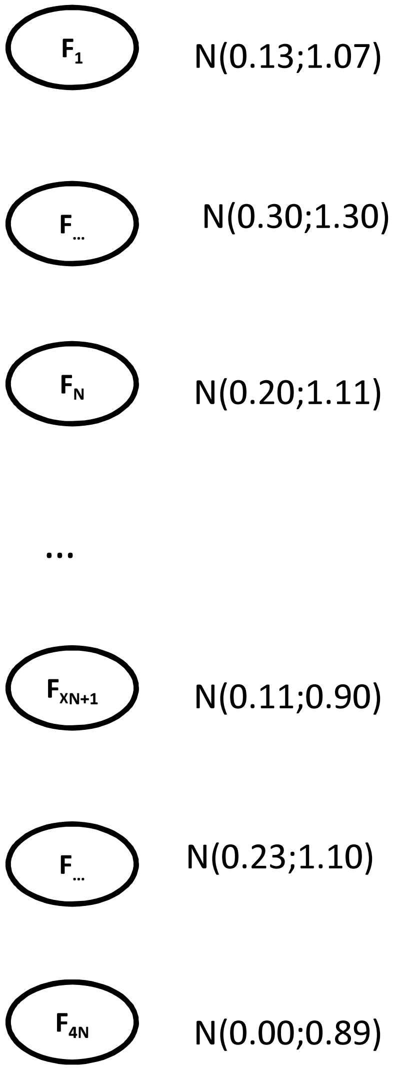
Background (III)

- Why detecting non-comparable items in large scale cross-national analyses?
  1. **Finding non-invariant items for instrument improvement**
  2. **Constructing partial measurement invariance models**

Monte Carlo simulations



Step 1. Generating continuous latent trait



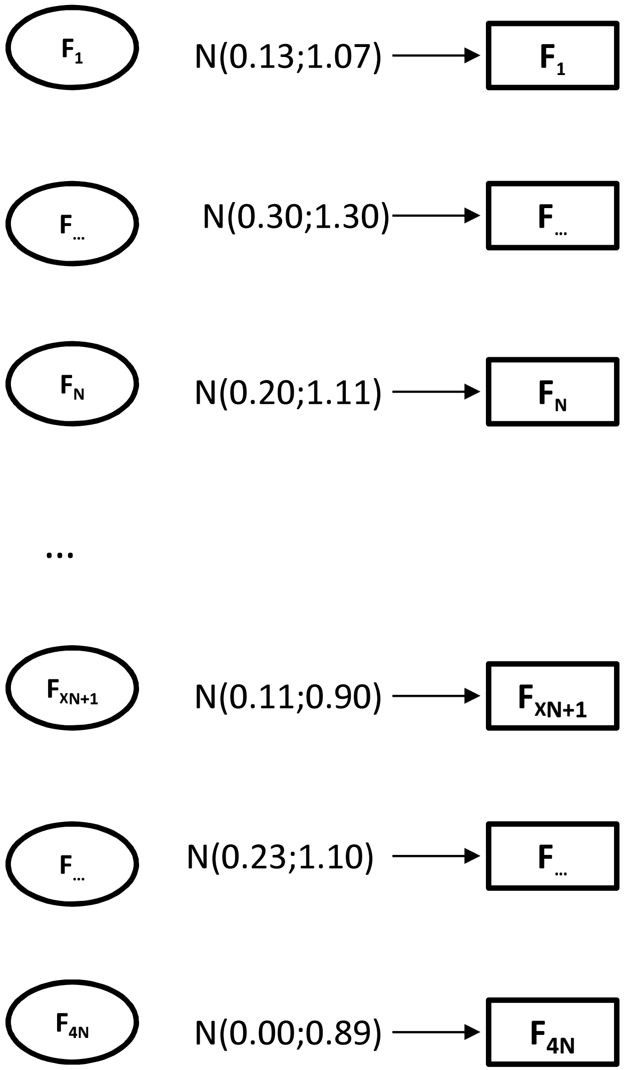
Step 1a:
Draw means and SDs for groups

$$\textbf{MEANs} \sim N(0; 0.3)$$

$$\textbf{SDs} \sim N(1; 0.1)$$

Step 1b:
Draw continuous variables reflecting true latent trait

Step 1. Generating continuous latent trait

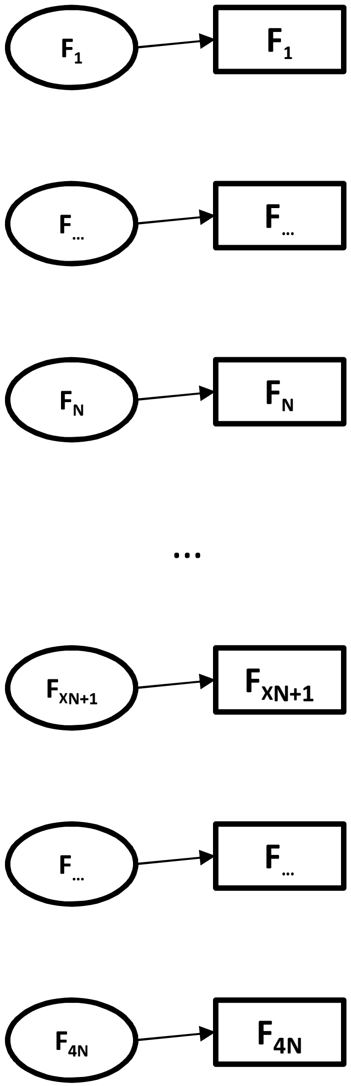


Step 1a:
Draw means and SDs for groups

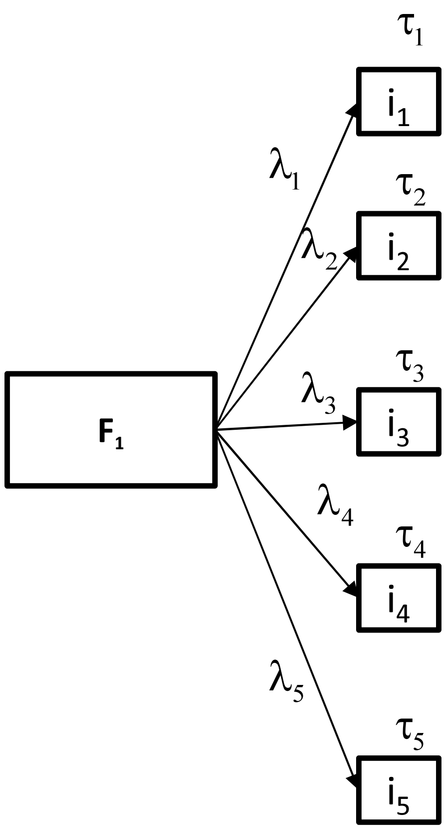
$$\mathbf{MEANs} \sim N(0; 0.3)$$
$$\mathbf{SDs} \sim N(1; 0.1)$$

Step 1b:
Draw continuous variables reflecting true latent trait

Step 1. Generating continuous latent trait



Step 1. Generating continuous latent trait



Step 2a:
Draw item parameters (the same for all groups)

$$\boldsymbol{\lambda} \sim U[0.4; 2.0]$$

$$\boldsymbol{\tau} \sim N(0; 1)$$

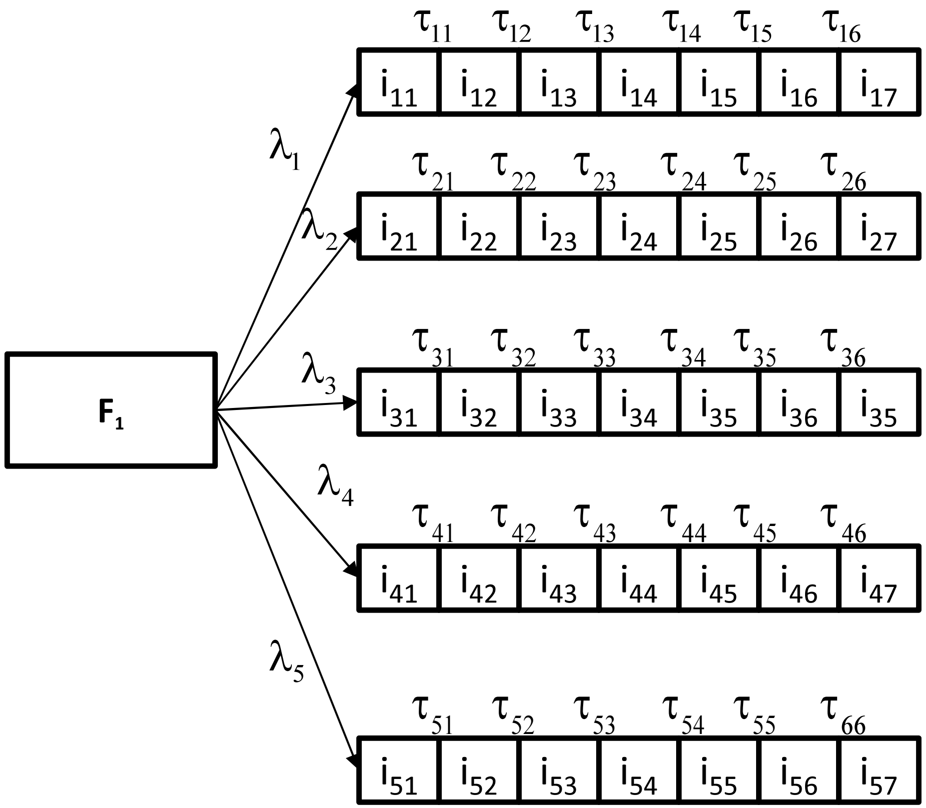
Optional Step:
Add small non-invariance (different random term for each item – approximate MI):

$$\boldsymbol{\lambda} + N(0; 0.05)$$

$$\boldsymbol{\tau} + N(0; 0.05)$$

Step 1. Generating continuous latent trait

Step 2. Generating item parameters



Step 2a:
Draw item parameters (the same for all groups)

$$\boldsymbol{\lambda} \sim U[0.4; 2.0]$$

$$\boldsymbol{\tau} \sim N(0; 1)$$

Step 2b:
Generate parameters for GRM

$$\tau_{i1} = \tau_i - 3.0$$

$$\tau_{i2} = \tau_i - 1.75$$

$$\tau_{i3} = \tau_i - 0.6$$

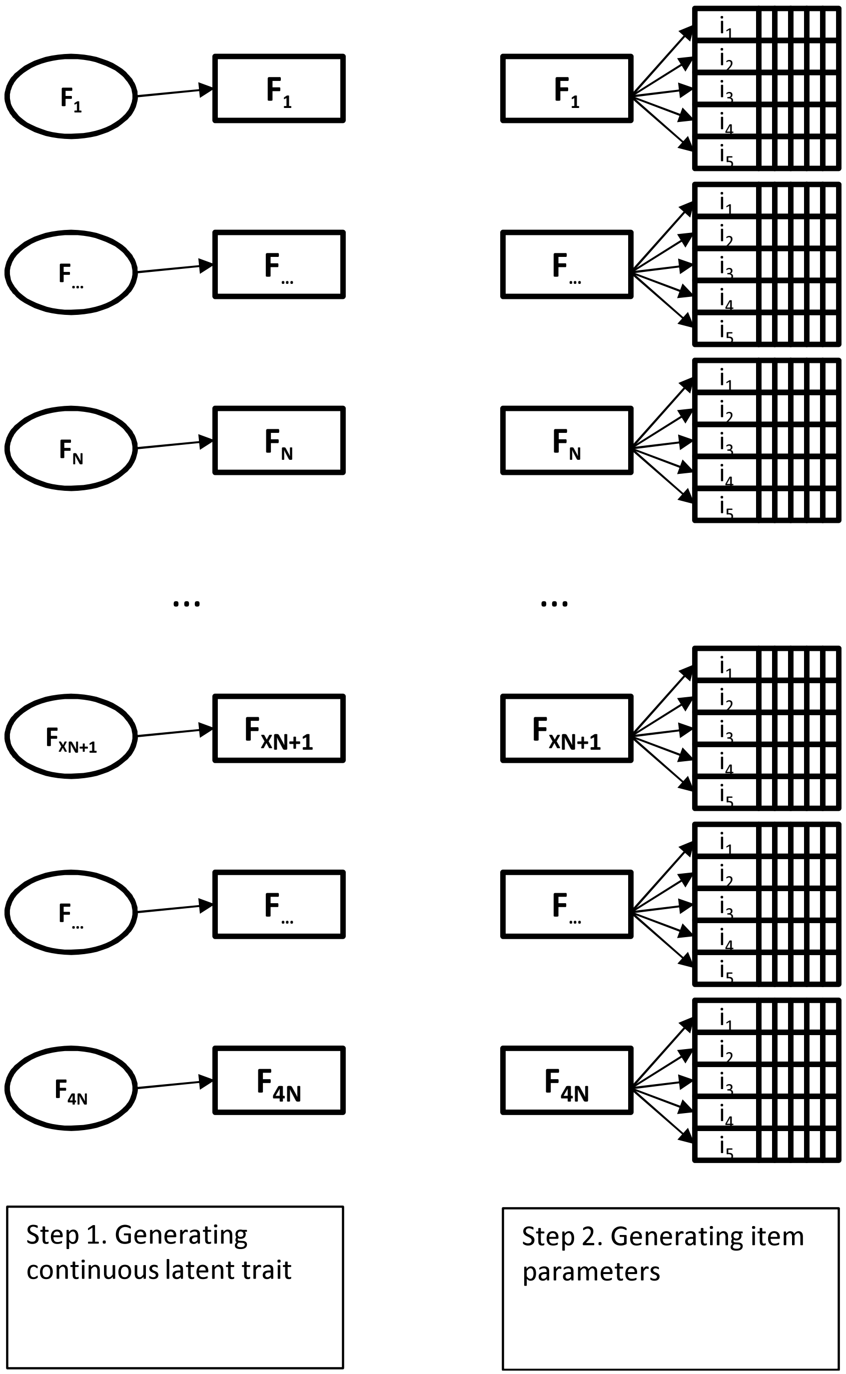
$$\tau_{i4} = \tau_i + 0.6$$

$$\tau_{i5} = \tau_i + 1.75$$

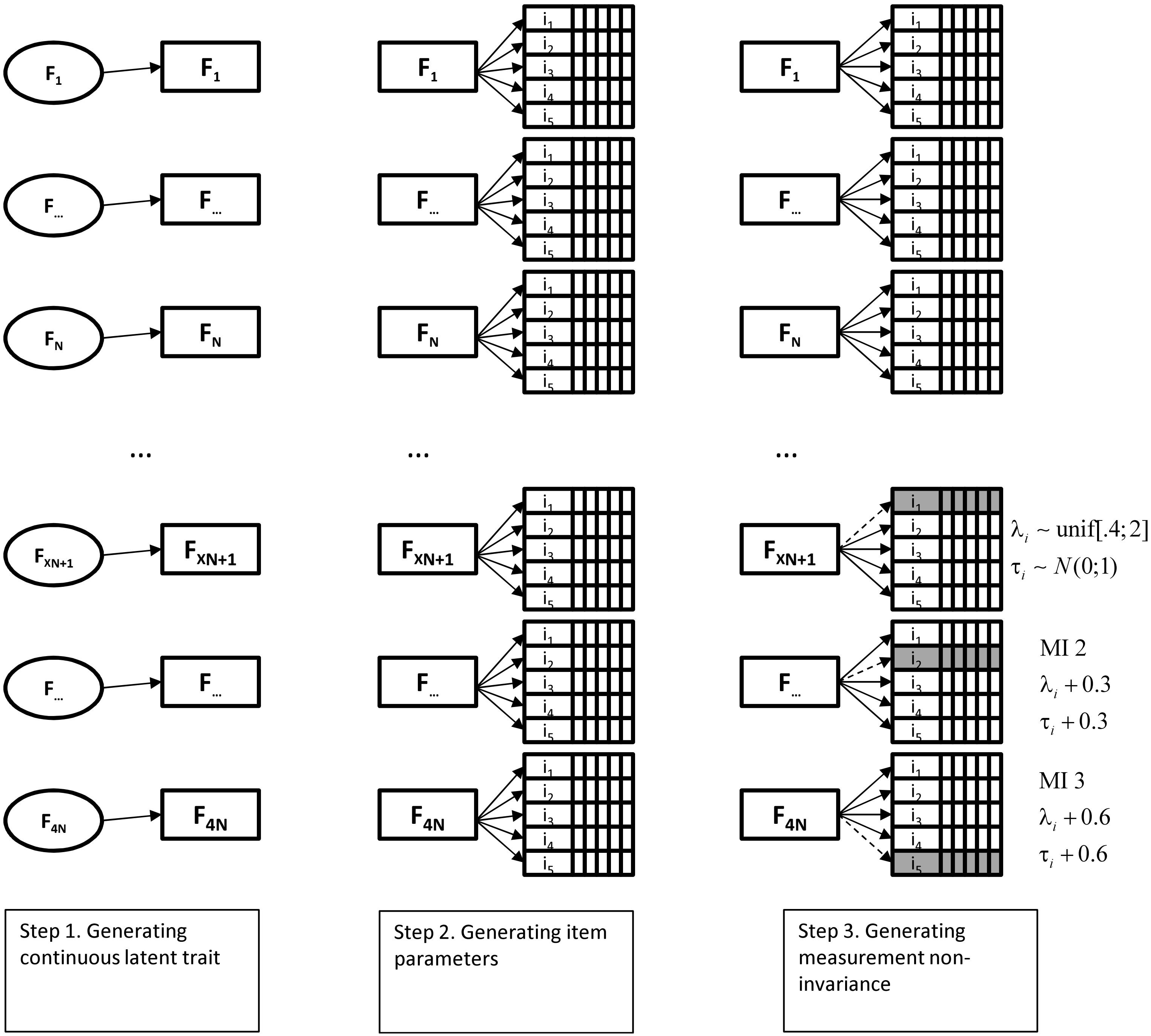
$$\tau_{i6} = \tau_i + 3.0$$

Step 1. Generating continuous latent trait

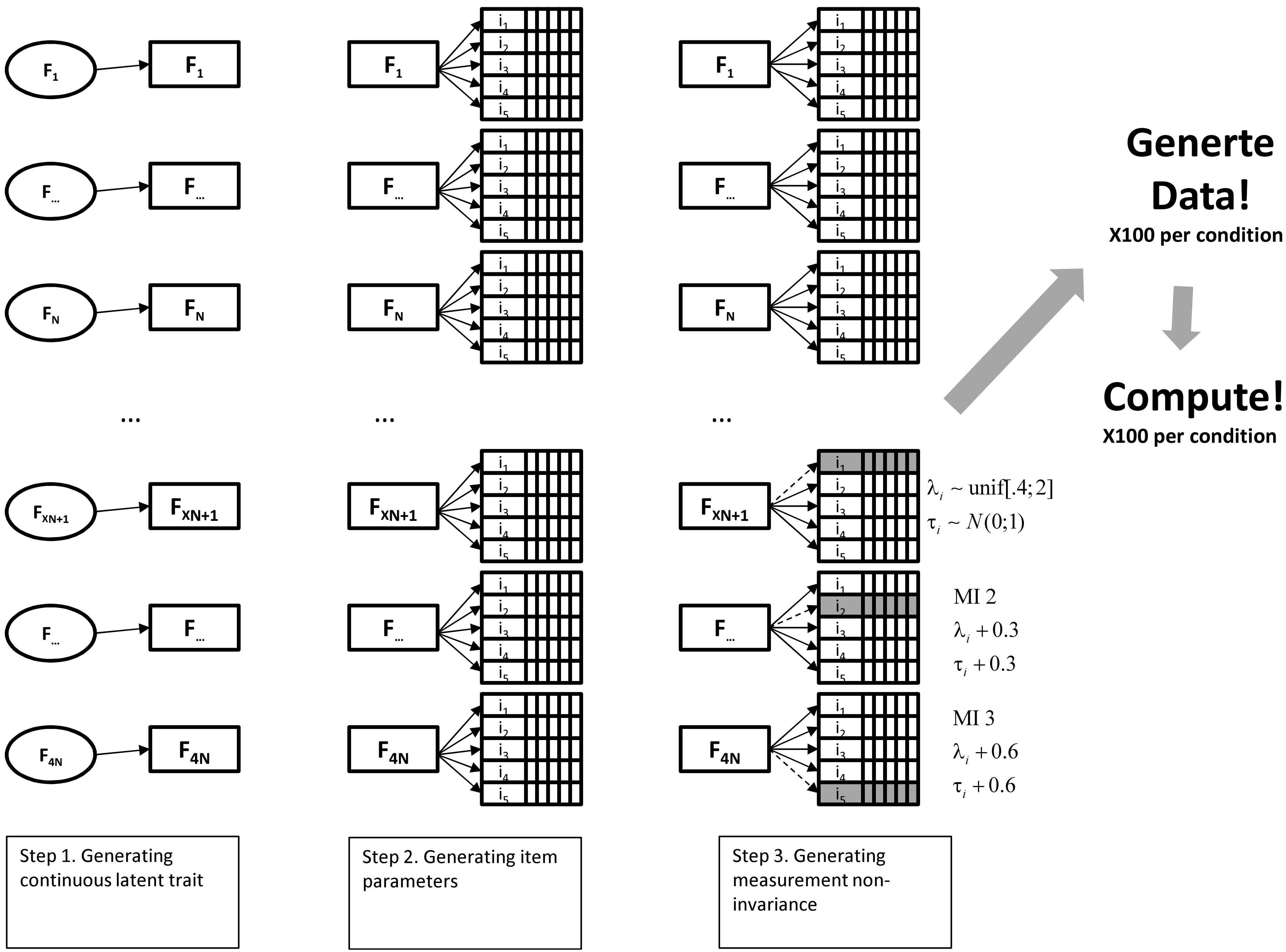
Step 2. Generating item parameters



F1
F…
FN
…
FXN+1
F…
F4N
i1
i2
i3
i4
i5
Step 1. Generating continuous latent trait
Step 2. Generating item parameters



F1
F…
FN
FXN+1
F4N
i1
i2
i3
i4
i5
λi ~ unif[.4; 2]
τi ~ N(0;1)
MI 2
λi + 0.3
τi + 0.3
MI 3
λi + 0.6
τi + 0.6
Step 1. Generating continuous latent trait
Step 2. Generating item parameters
Step 3. Generating measurement non-invariance



F1
F...
FN
...
FXN+1
F4N
i1
i2
i3
i4
i5
λi ~ unif[.4; 2]
τi ~ N(0;1)
MI 2
λi + 0.3
τi + 0.3
MI 3
λi + 0.6
τi + 0.6
Generte Data!
X100 per condition
Compute!
X100 per condition
Step 1. Generating continuous latent trait
Step 2. Generating item parameters
Step 3. Generating measurement non-invariance

Conditions for simulations

Number of groups	Sample size	Number of groups affected	MI type	Bias	Items affected per group
12	100	25%	Classic	random	1
24	300	50%	Aproximate	+0.3	…
48	1000	75%		+0.6	5

Statistical methods for detecting

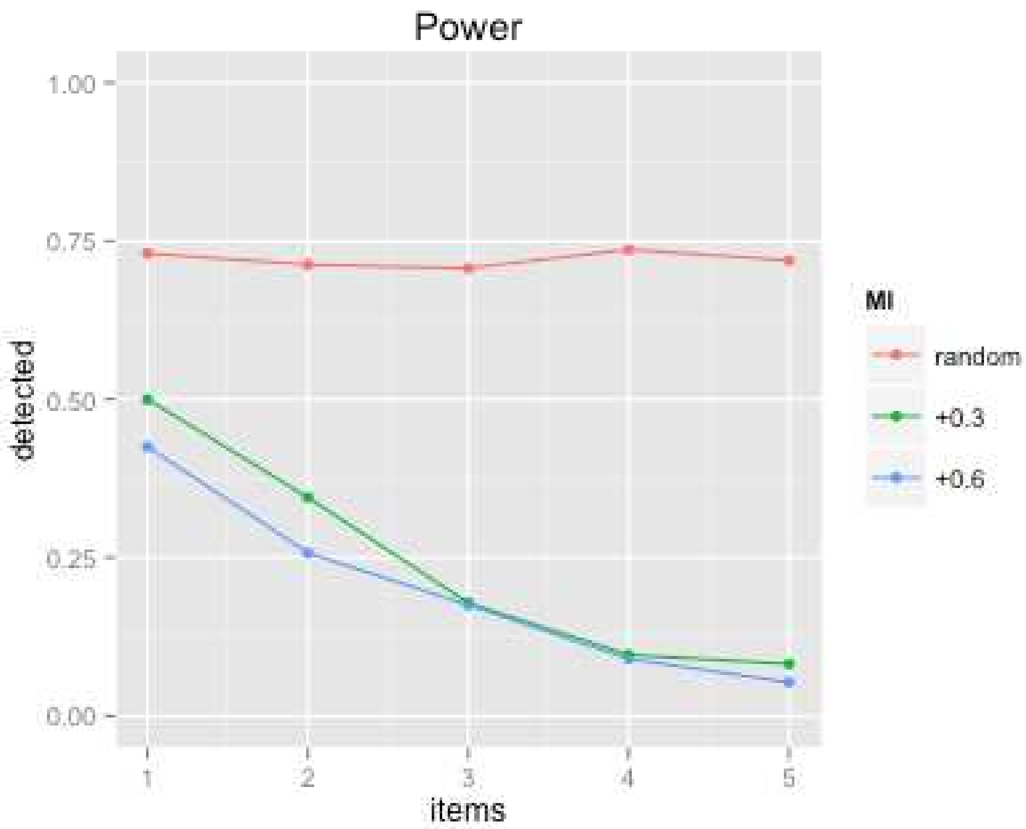
1. Modification Index (MI) and the Power of the Test
   – Saris, Satorra, van der Veld (2009); Cieciuch, J., Davidov, E., Oberski, D.L., & Algesheimer, R. (2015)
2. Bayesian Structural Equating Modeling measurement invariance analysis
   – De Jong et al. (2007); Muthén & Asparouhov (2013:9)
3. Alignment optimization invariance analysis
   – Asparouhov & Muthén (2014:7); Muthén & Asparouhov, (2014)
4. Multilevel Confirmatory Factor Analysis
   – Fox, (2010); Hox et al. (2010); Muthén & Asparouhov (2013)
5. DIF procedures (for instance Mantel, N., & Haenszel or Generalized Regression Approach)
   – Mantel & Haenszel (1959); Swaminathan & Rogers (1990)

Mplus

Some results!

MI and the Power of the Test

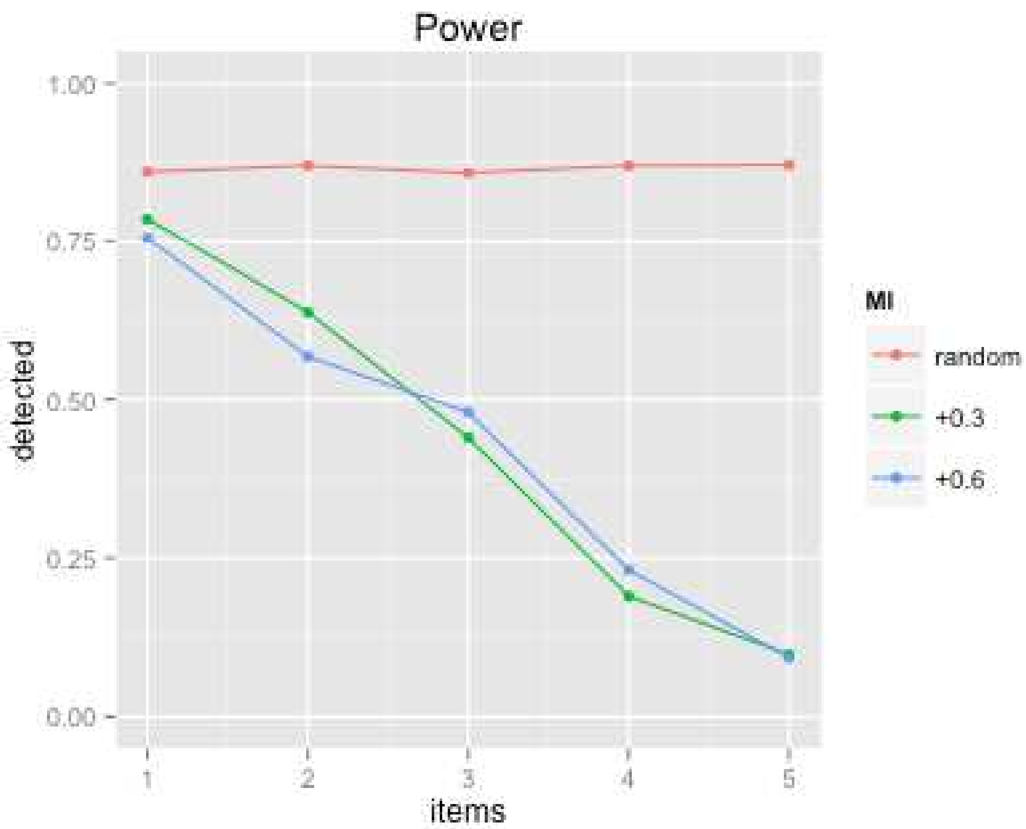
EPC=0.1, Alpha=0.05, Power=0.75   N=24, N=24x1000, N affected=25%



Average Type I Error = **10.32%**

MI and the Power of the Test

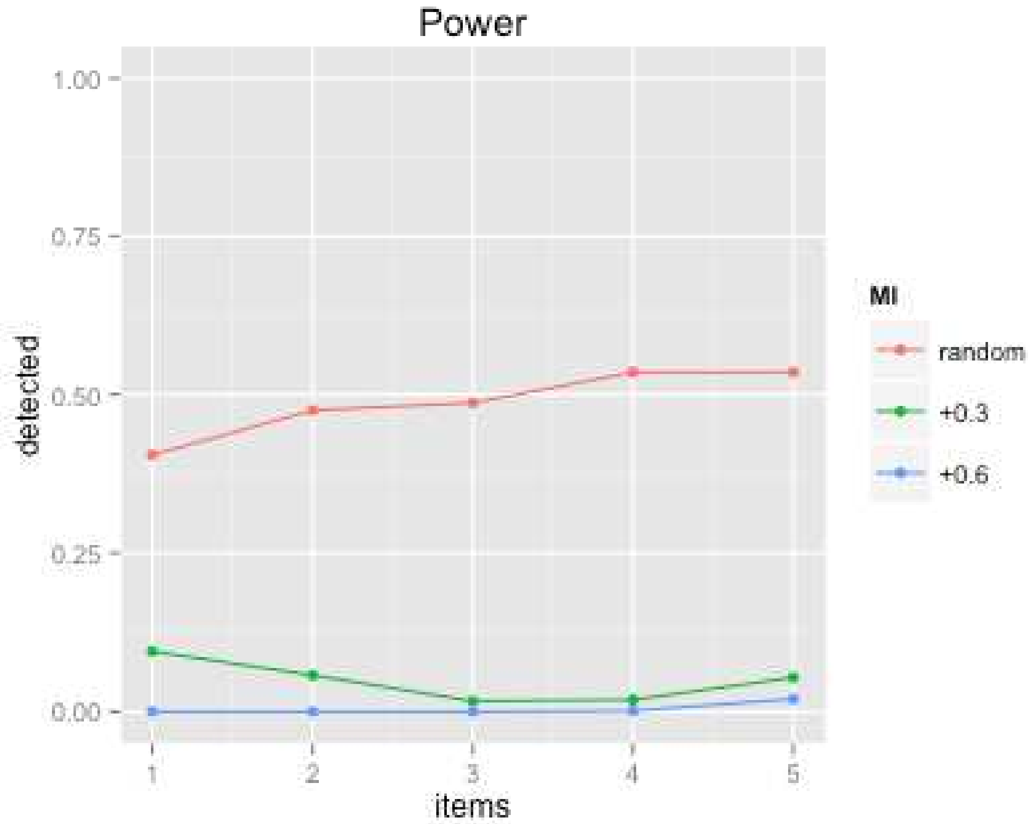
EPC=0.1, Alpha=0.05, Power=0.90   N=24, N=24x1000, N affected=25%



Average Type I Error = **13.65%**

MI and the Power of the Test

EPC=0.1, Alpha=0.05, Power=0.50 N=24, N=24x1000, N affected=25%

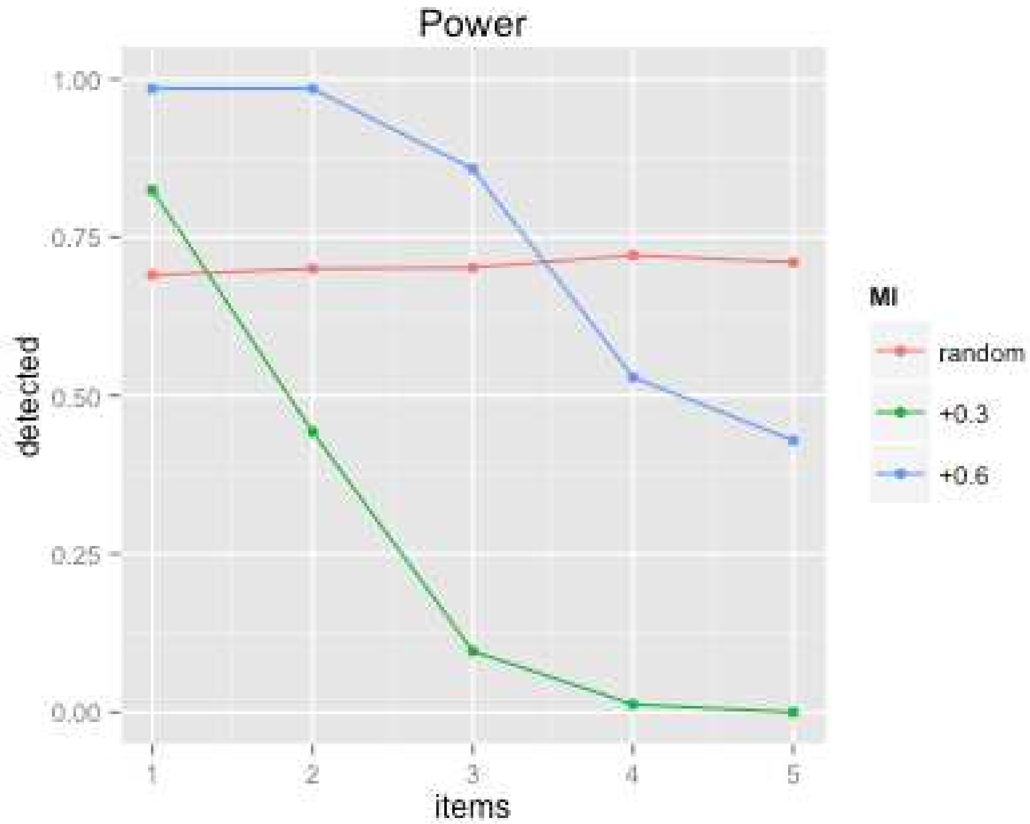


Average Type I Error = **7.92%**

BSEM

Priors ~ N(0,0.05)

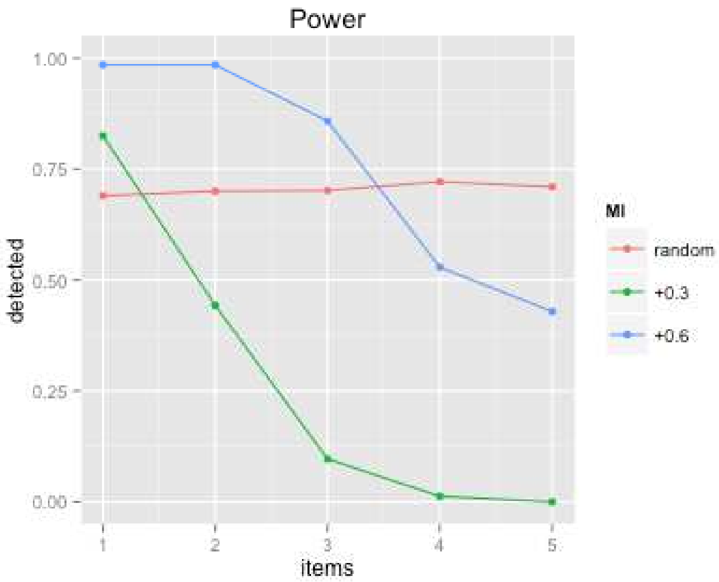
N=24, N=24x1000, N affected=25%



Average Type I Error = **3.32%**

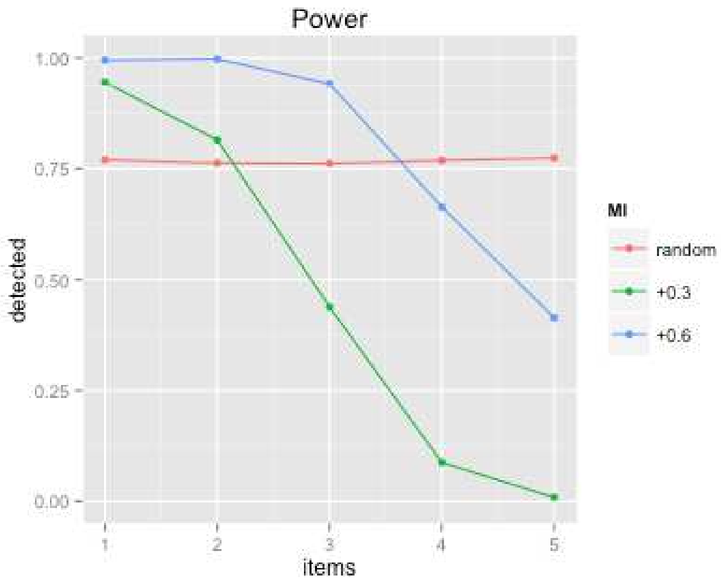
BSEM with different priors

DIFF(tau1_#-tau24_#)~N(0,0.05)



Average Type I Error = **3.32%**

DIFF(tau1_#-tau24_#)~N(0,0.01);

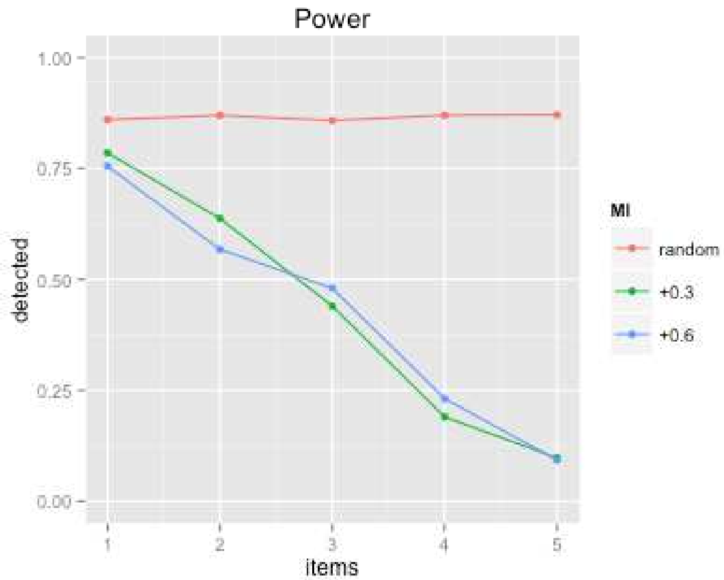


Average Type I Error = **4.26%**

MI vs BSEM

MI

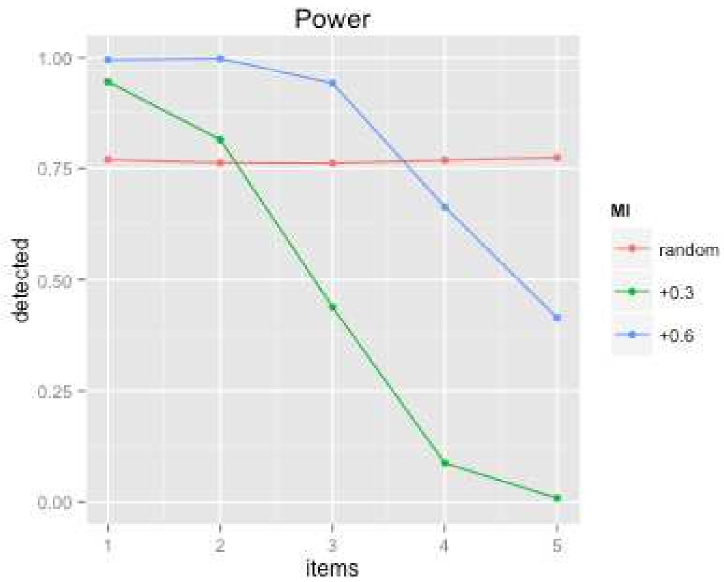
EPC=0.1, Alpha=0.05, Power=0.90



Average Type I Error = **13.32%**

BSEM

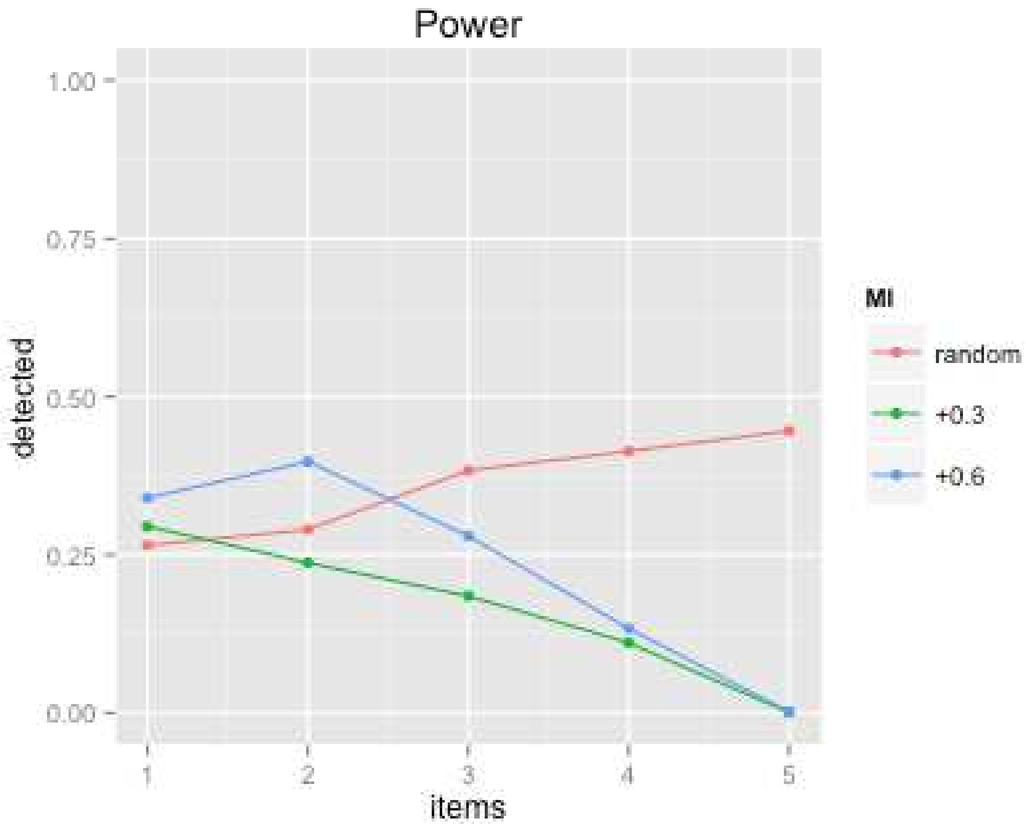
DIFF(tau1_#-tau24_#)~N(0,0.01);



Average Type I Error = **4.26%**

Multiple-Group FA Alignment (FREE)

N=24, N=24x1000, N affected=25%



Average Type I Error = **0.01%**

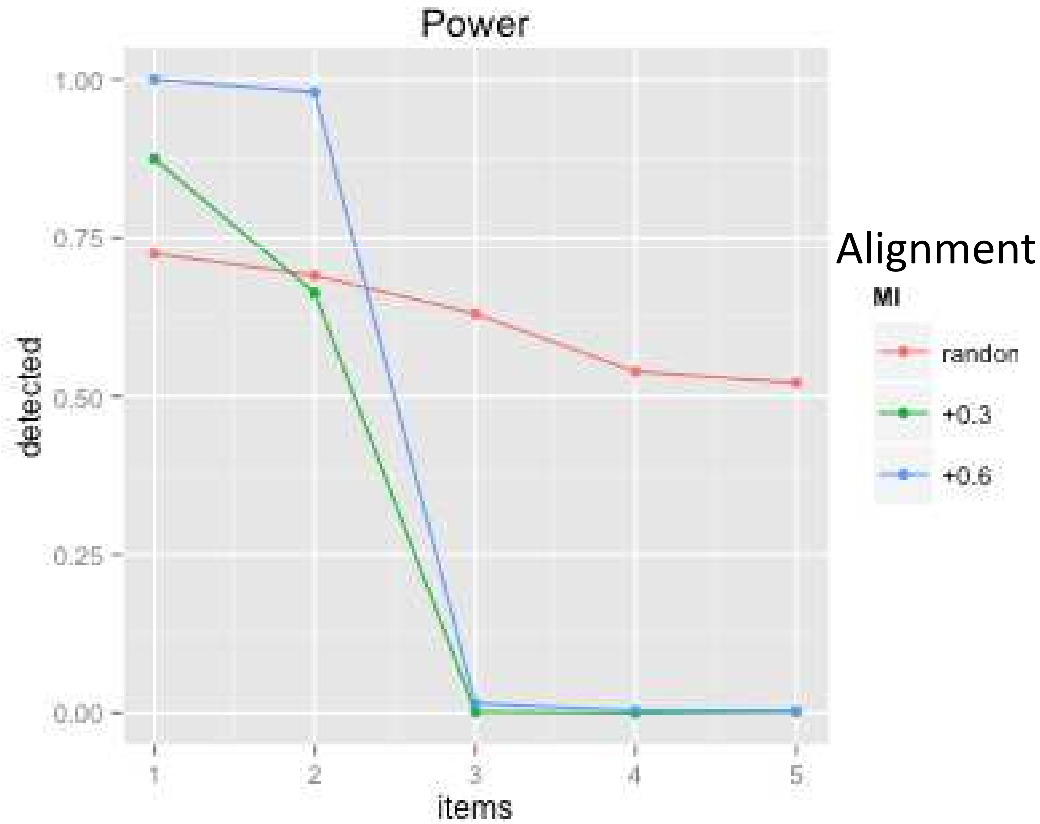
Multiple-Group FA Alignment (FREE)

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


THE MODEL ESTIMATION TERMINATED NORMALLY
```

Multiple-Group FA Alignment (FIXED)

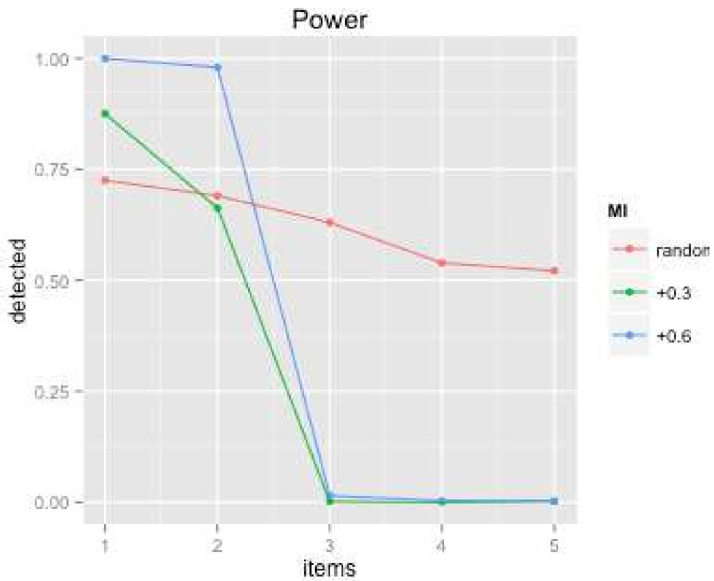
N=24, N=24x1000, N affected=25%



Average Type I Error = **0.01%**

Alignment vs BSEM

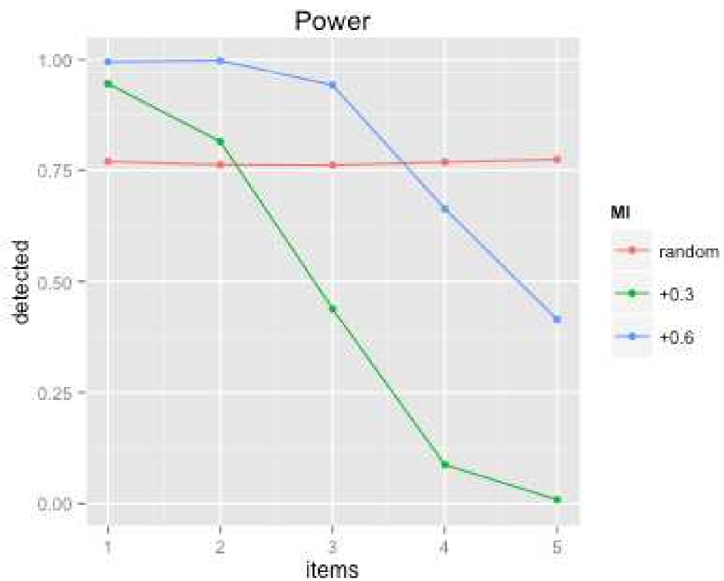
Alignment



Average Type I Error = **0.01%**

BSEM

DIFF(tau1_#-tau24_#)~N(0,0.01);

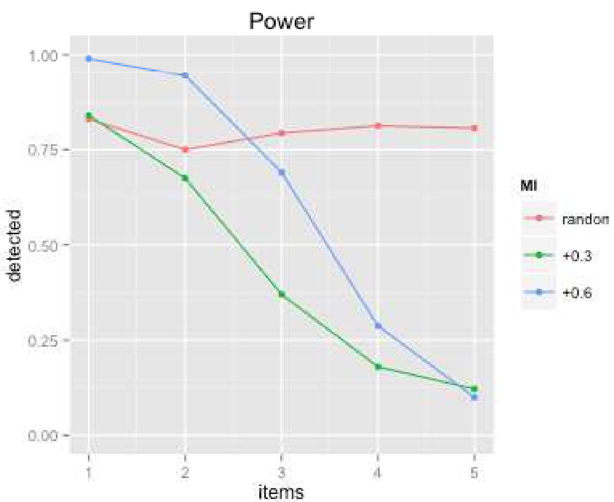


Average Type I Error = **4.26%**

Comparison of methods (power)

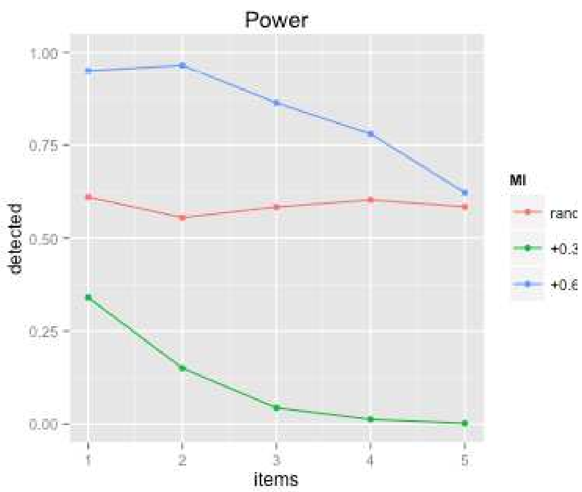
N=12, N=12x300, N affected=25%

MI



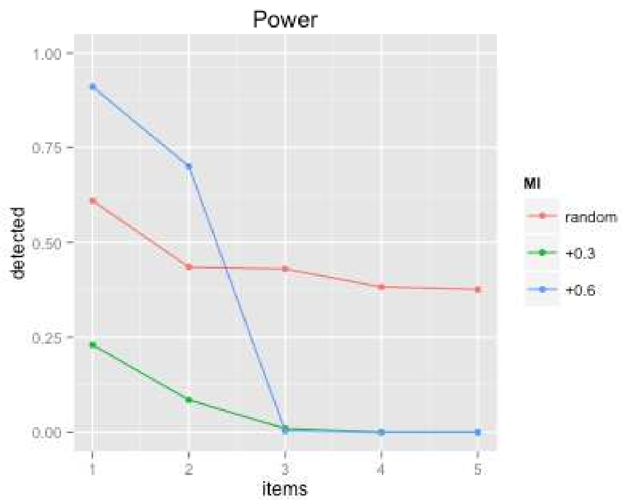
Average Type I Error = **9.79%**

BSEM



Average Type I Error = **4.12%**

Alignment



Average Type I Error = **0.01%**

Summary of simulations

- **MI and the Power of the Test**
  - Highest power for random dif
  - Highest type I error
- **BSEM**
  - High power for random dif (MI better for 12 groups)
  - High power for one direction dif (even when many items are affected)
  - Relatively robust to prior specification
- **Alignment**
  - Lowest type I error
  - Very effective when scale has only few large item biases
  - Lowest power
  - Mplus makes all decisions for you

Thank you!