Explaining cross-national inequivalence
A Bayesian Multilevel SEM approach

Bart Meuleman (University of Leuven) & Elmar Schlüter (Justus-Liebig-Universität Gießen)
Introduction

- Challenges in measurement equivalence testing
  - Large-scale surveys: increasing number of contexts
  - What to do if equivalence cannot be established?

- Outline:
  - MLSEM approach to measurement equivalence
    - The two-level CFA model
    - Measurement equivalence
    - Bayesian estimation
  - Illustration: citizenship conceptions in ISSP
    - Explaining random slope variation
    - Simulation study
Two-level CFA model

• A multilevel approach to CFA (Muthen 1994)
  – Starting point: population of individuals $i$ divided into $g$ groups
  – Decomposition of individual data into within group and between group components:
    \[ X_{ig} = X_W + X_B \]
    \[ X_{ig} = (X_{ig} - \overline{X}_g) + \overline{X}_g \]
  – Orthogonal decomposition of total covariance structure into within- and between-group covariance structures:
    \[ \Sigma_T = \Sigma_W + \Sigma_B \]
Two-level CFA model

- One model that simultaneously predicts within- and between-group components of the data:

\[ X_{ig} = \alpha_g + \Lambda_W \eta_W + \delta_{Wig} \]  
(1)

\[ \alpha_g = \nu + \Lambda_B \eta_B + \delta_{Bg} \]  
(2)

- Substitution of (2) into (1)

\[ X_{ig} = \nu + \Lambda_W \eta_W + \Lambda_B \eta_B + \delta_{Bg} + \delta_{Wig} \]

- This model implies:

\[ \Sigma_W = \Lambda_W \Phi_W \Lambda_W' + \Theta_W \]

\[ \Sigma_B = \Lambda_B \Phi_B \Lambda_B' + \Theta_B \]
Two-level CFA model

\[ \alpha_g = \nu + \Lambda_B \eta_B + \delta_{Bg} \]

\[ X_{ig} = \alpha_g + \Lambda_w \eta_w + \delta_{Wig} \]

\[ X_{ig} = \nu + \Lambda_w \eta_w + \Lambda_B \eta_B + \delta_{Bg} + \delta_{Wig} \]
Configural equivalence

Multigroup CFA vs. Multilevel CFA

- Equal factor structures across groups
- Equal structures of group-specific within matrices
- Analysis of pooled within matrix → implied by the model
Metric equivalence

Multigroup CFA vs. Multilevel CFA

- Equal factor loadings across groups

Group A

\[ \eta^A \]

Group B

\[ \eta^B \]

\[ \text{X1} \quad \text{X2} \quad \text{X3} \]

Multilevel CFA

- Single set of factor loadings for pooled within matrix → implied by the model
- But: can be overruled by including a random factor loading

\[ \eta^B \]

\[ \text{X1B} \quad \text{X2B} \quad \text{X3B} \]

\[ \text{X1W} \quad \text{X2W} \quad \text{X3W} \]

\[ \eta^W \]
Scalar equivalence

Multigroup CFA vs. Multilevel CFA

- Equal intercepts across groups

- Between-level residuals contain variations in item means not captured by the latent variable
- To test scalar equivalence, constrain between-level residuals to 0
Explaining inequivalence

\[ Z \rightarrow d4 \rightarrow X1B \]
\[ Z \rightarrow \eta B \rightarrow X2B \]
\[ Z \rightarrow X3B \]

\[ \text{BETWEEN} \]
\[ \text{WITHIN} \]

\[ X1W \rightarrow \eta W \rightarrow X2W \rightarrow X3W \]
Bayesian estimation

• Bayes’ theorem:

\[ P(\text{hypothesis} \mid \text{data}) \propto P(\text{data} \mid \text{hypothesis}) \times P(\text{hypothesis}) \]

Advantages:
1. More can be learned from parameters
2. **Better small-sample performance**
3. Computationally less demanding
4. New types of models can be analyzed
Application: citizenship conceptions

- Explaining slope differences in MLSEM:
  - International Social Survey Program
  - 2003 wave: focus on national identity
  - 32 societies
Application: citizenship conceptions

• Scale measuring **ethnic citizenship conceptions**:

  Some people say that the following things are important for being truly [NATIONALITY]. Others say they are not important. How important do you think each of the following is?

  • To have been born in the country (**BRN**)
  • To have [COUNTRY NATIONALITY] citizenship (**CTZ**)
  • To have lived in [COUNTRY] for most of one’s life (**LVD**)
  • To be a [RELIGION] (**RLG**)

• **Answer scale:** very important (1) – not important at all (4) (reversed)
Application: citizenship conceptions

- MGCFA for 32 groups:

<table>
<thead>
<tr>
<th></th>
<th>Chi2</th>
<th>Df</th>
<th>RMSEA</th>
<th>CFI</th>
<th>TLI</th>
<th>SRMR</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Configural equivalence</strong></td>
<td>359.658</td>
<td>64</td>
<td>0.063</td>
<td>0.992</td>
<td>0.976</td>
<td>0.018</td>
</tr>
<tr>
<td><strong>Metric equivalence</strong></td>
<td>2440.424</td>
<td>157</td>
<td>0.112</td>
<td>0.939</td>
<td>0.926</td>
<td>0.145</td>
</tr>
<tr>
<td><strong>Scalar equivalence</strong></td>
<td>10850.897</td>
<td>250</td>
<td>0.191</td>
<td>0.718</td>
<td>0.783</td>
<td>0.244</td>
</tr>
</tbody>
</table>
Application: citizenship conceptions

- Two-level CFA:
  - Bayesian estimation
    - Uninformative priors
    - 10,000 iterations

**WITHIN MODEL**

*Factor loadings*

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>BRN</td>
<td>0.787*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CTZ</td>
<td>0.693*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LVD</td>
<td>0.689*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RLG</td>
<td>0.426*</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Residual Variances*

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>BRN</td>
<td>0.381*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CTZ</td>
<td>0.52*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LVD</td>
<td>0.526*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RLG</td>
<td>0.819*</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**BETWEEN MODEL**

*Factor loadings*

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>BRN</td>
<td>0.973*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CTZ</td>
<td>0.694*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LVD</td>
<td>0.962*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RLG</td>
<td>0.747*</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Residual Variances*

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>BRN</td>
<td>0.053*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CTZ</td>
<td>0.518*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LVD</td>
<td>0.075*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RLG</td>
<td>0.442*</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Application: citizenship conceptions

• Set factor loading for BRN random:

\[ \text{Etn}_b \]

\[ \text{BRN} \quad \text{CTZ} \quad \text{LVD} \quad \text{RLG} \]

\[ s \quad \text{Etn}_b \quad \text{Etn}_w \]

\[ 0.019 \quad (0.006) \]
Application: citizenship conceptions

- Effect of % of immigrants (UNDP estimate) on slope variance:

\[
\begin{align*}
\text{BRN} & \quad \text{CTZ} \\
\text{LVD} & \quad \text{RLG}
\end{align*}
\]
Application: citizenship conceptions

Trace plot of: Parameter 19, %BETWEEN%: S1 ON MIGPER

Distribution of: Parameter 19, %BETWEEN%: S1 ON MIGPER

- Mean = 0.01010, Std Dev = 0.00338
- Median = 0.01015
- Mode = 0.01080
- 95% Lower CI = 0.00341
- 95% Upper CI = 0.01669
### Application: citizenship conceptions

- Monte Carlo simulation (500 replications)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>WITHIN</th>
<th>BETWEEN</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>bias</strong></td>
<td><strong>SE bias</strong></td>
<td>Coverage</td>
</tr>
<tr>
<td>Factor loadings</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item1</td>
<td>0.03%</td>
<td>1.89%</td>
</tr>
<tr>
<td>Item2</td>
<td>-0.03%</td>
<td>1.89%</td>
</tr>
<tr>
<td>Item3</td>
<td>-0.03%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Item 4</td>
<td>0.08%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Residual variances</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item1</td>
<td>0.04%</td>
<td>2.00%</td>
</tr>
<tr>
<td>Item2</td>
<td>-0.02%</td>
<td>-1.92%</td>
</tr>
<tr>
<td>Item3</td>
<td>-0.02%</td>
<td>-3.77%</td>
</tr>
<tr>
<td>Item4</td>
<td>0.04%</td>
<td>-5.66%</td>
</tr>
<tr>
<td>Random slope</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Effect of Z</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Conclusion

• In some cases, measurement inequivalence can be a source of information rather than just nuisance
• MLSEM can help to explain why slopes (and intercepts) vary across groups
• It might even be possible to explain inequivalence away
• Bayesian estimation produces sufficiently reliable inference, even with small sample sizes

• Unresolved issues:
  – Model fit indices in Bayesian MLSEM?
  – Set free multiple factor loadings (cfr. IRT approach)?
  – Small-variance priors on the factor loading variances?
Thank you for your attention!

Bart Meuleman
bart.meuleman@kuleuven.be
Twitter: @meuleb