A methodological study: Sensitivity analysis of the consequences of ignoring measurement inequivalence in latent variable modelling of cross-national survey data

Jouni Kuha, Irini Moustaki, Sally Stares
London School of Economics and Political Science

ESRC Research Methods Festival
Outline

- Sensitivity analysis
  - Latent Class Model
  - Latent Trait Model
- Methodology
- Results
- Tentative Conclusions
Notation

Observed data for one subject $l$:

- $p$ categorical response items $\mathbf{y}_l = (y_{l1}, \ldots, y_{lp})$ with $L_i$ levels for item $i = 1, \ldots, p$
- A response pattern is denoted by: $\mathbf{y}_l = (y_1 = a_1, \ldots, y_p = a_p)$ where $a_i$ indicate a response category.
- The group level variable is introduced with an explanatory variable $\mathbf{x}$ of $k$ levels.

Observed data for the full sample is a $(p + 1)$-way items-by-group contingency table with $L_1 \times \cdots \times L_p \times K$ cells and frequencies $(Y_k, n_k)$ for each $k = 1, \ldots, K$
Latent variables:

- $q$ latent variables $\eta = (\eta_1, \ldots, \eta_q)$
- $\eta$ can be either discrete with $J$ classes (latent class model) or continuous (latent trait model).
Multi-group Latent Class model

The probability of a response pattern conditional on the group variable \( x \)

\[
P(y \mid x = k) = \sum_{j=1}^{J} P(\eta = j \mid x = k)P(y \mid \eta = j, x = k),
\]

(1)

The probability of a positive response:

\[
\text{logit}P_k(y_i = 1 \mid \eta = j) = \alpha_{ijk}, \; i = 1, \ldots, p; \; j = 1, \ldots, J; \; k = 1, \ldots, K.
\]

(2)

The prior probability

\[
\text{logit}P_k(\eta = j) = \alpha_{0jk}, \; j = 1, \ldots, J; \; k = 1, \ldots, K.
\]
Multi-group Latent Trait model

The probability of a response pattern conditional on the group variable $x$

$$P(y | x) = P(y_1 = a_1, \ldots, y_p = a_p | x) = \int_{R(\eta_1)} \cdots \int_{R(\eta_q)} P(y | \eta, x) \phi(\eta | x) d\eta$$

(3)

where $\eta_k \sim N(\mu_k, \Phi_k)$.

The probability of a positive response:

$$\text{logit} P_k(y_i = 1 | \eta) = \alpha_{0ik} + \sum_{j=1}^{q} \alpha_{ijk} \eta_j, \quad i = 1, \ldots, p; \quad k = 1, \ldots, K.$$  (4)
The effect of non-equivalence using a simple example

Latent trait $\eta_k \sim N(\mu_k, \phi_k)$ in country $k$

Measurement model for binary item $y_i$ in country $k$:

$$\logit\left[ P(y_i = 1|\eta, x = k) \right] = \alpha_{0ik} + \alpha_{ik} \eta_k$$

Observed (unconditional) probabilities of item $y_i$ in country $k$:

$$\logit\left[ P(y_i = 1 | x = k) \right] \approx \frac{\alpha_{0ik} + \alpha_{ik} \mu_k}{\left[ 1 + 0.35(\alpha_{ik} \sqrt{\phi_k})^2 \right]^{1/2}}$$

- The unconditional probabilities depend on both the measurement parameters and the structural parameters.
- If the former are estimated incorrectly, then so are the latter.
What happens when measurement equivalence is ignored

- Instead of country-specific measurement parameters $\alpha_{0ik}$ and $\alpha_{ik}$, we use the same values $\alpha_{0i}$ and $\alpha_i$ for each country.
- There estimated values will be some kind of average of the country-specific values, which will be more or less wrong for any given country.
- For each item the assumption of equivalence may be correct or incorrect in different ways.
- The aggregate effect of assuming equivalence on estimates of the structural parameters $\mu_k$ and $\phi_k$ cannot be read from a simple formula.
What does incorrect model converge to?

The most common way of examining this: Simulation study

- Generate data sets from true model $p(Y|X = k)$
- Estimate analysis model $f(Y|X = k; \theta)$ for each of these, to get estimates $\hat{\theta}_i^*$
- Estimate $\theta^*$ as average of $\hat{\theta}_i^*$
- Kaplan and George (1995) and De Beuckelaer and Swinnen (2011) do this for questions on measurement equivalence like ours, for factor analysis models
An alternative approach

(Berk 1966; Huber 1967; Akaike 1973; White 1982):

- Parameter estimates of the analysis model converge to the value $\theta^*$ which maximizes the expectation $E_p \left[ \log f(Y|X = x; \theta) \right]$ under the true model.

Can be evaluated as follows:

- Calculate expected frequencies $E_k = n_k P(y_1, \ldots, y_p|X = k)$ under the true model for each $k = 1, \ldots, K$ and all values of $(y_1, \ldots, y_p)$, i.e. for the full items-group table

- Fit analysis model to $(E_k, n_k)$

- Estimate from this fit is $\theta^*$

- This way each true-analysis model combination requires just one model fit, not many as in a simulation

- Same idea used for sensitivity analyses in other contexts by Rotnitzky and Wypij (1994) and Heagerty and Kurland (2001)
Sensitivity Analysis

Experimental conditions:
- Length of scale.
- Number of groups (e.g. countries)
- Size of each group.
- Severity of non-equivalence (various rules)
- Values of thresholds and factor loadings.
Non-equivalence rules

Table: Rule: R1,.8,.1,.9, true value=0.8, three groups

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>...</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>G1</td>
<td>0.80</td>
<td>0.80</td>
<td>0.80</td>
<td>0.80</td>
<td>0.80</td>
<td></td>
<td>0.80</td>
</tr>
<tr>
<td>G2</td>
<td>0.45</td>
<td>0.47</td>
<td>0.49</td>
<td>0.51</td>
<td>0.53</td>
<td></td>
<td>0.85</td>
</tr>
<tr>
<td>G3</td>
<td>0.10</td>
<td>0.14</td>
<td>0.18</td>
<td>0.23</td>
<td>0.27</td>
<td></td>
<td>0.90</td>
</tr>
</tbody>
</table>

Table: Rule: R1,.2,.05,.3, true value=0.2, three groups

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>...</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>G1</td>
<td>0.20</td>
<td>0.20</td>
<td>0.20</td>
<td>0.20</td>
<td>0.20</td>
<td></td>
<td>0.20</td>
</tr>
<tr>
<td>G2</td>
<td>0.12</td>
<td>0.13</td>
<td>0.14</td>
<td>0.14</td>
<td>0.15</td>
<td></td>
<td>0.25</td>
</tr>
<tr>
<td>G3</td>
<td>0.05</td>
<td>0.06</td>
<td>0.08</td>
<td>0.09</td>
<td>0.10</td>
<td></td>
<td>0.30</td>
</tr>
</tbody>
</table>
Example: Latent class model with three classes.

### Measurement model

<table>
<thead>
<tr>
<th>Item (Y)</th>
<th>$P_k(y_i \mid \eta = 1)$</th>
<th>$P_k(y_i \mid \eta = 2)$</th>
<th>$P_k(y_i \mid \eta = 3)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(<em>0.8</em>)</td>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>2</td>
<td>0.8</td>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>3</td>
<td>0.8</td>
<td>0.2</td>
<td>(<em>0.2</em>)</td>
</tr>
<tr>
<td>4</td>
<td>0.8</td>
<td>0.8</td>
<td>0.2</td>
</tr>
<tr>
<td>5</td>
<td>0.8</td>
<td>0.8</td>
<td>0.2</td>
</tr>
<tr>
<td>6</td>
<td>0.8</td>
<td>0.4</td>
<td>0.2</td>
</tr>
</tbody>
</table>

### Structural model

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>$P(\eta = 1 \mid x = 1)$</th>
<th>$P(\eta = 2 \mid x = 2)$</th>
<th>$P(\eta = 3 \mid x = 3)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1000</td>
<td>0.3</td>
<td>0.5</td>
<td>0.2</td>
</tr>
<tr>
<td>2</td>
<td>1000</td>
<td>0.3</td>
<td>0.3</td>
<td>0.4</td>
</tr>
<tr>
<td>3</td>
<td>1000</td>
<td>0.3</td>
<td>0.1</td>
<td>0.6</td>
</tr>
</tbody>
</table>
Figure: Class probabilities conditional on the group, one item out of six in all classes non-invariant
Figure: Class probabilities conditional on the group, two items out of nine in all classes non-invariant
Figure: Class probabilities conditional on the group, three items out of ten in all classes non-invariant
Table: Latent class model with three classes, varying the group probabilities.

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>$P(\eta = 1 \mid x = 1)$</th>
<th>$P(\eta = 2 \mid x = 2)$</th>
<th>$P(\eta = 3 \mid x = 3)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1000</td>
<td>0.8</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>2</td>
<td>1000</td>
<td>0.7</td>
<td>0.1</td>
<td>0.2</td>
</tr>
<tr>
<td>3</td>
<td>1000</td>
<td>0.6</td>
<td>0.1</td>
<td>0.3</td>
</tr>
</tbody>
</table>
Figure: Class probabilities conditional on the group, 1 item out of ten in one class non-invariant
Example: Latent Trait model.

### Measurement model

<table>
<thead>
<tr>
<th>Item (Y)</th>
<th>$P(y_i = 1 \mid \eta = -2)$</th>
<th>$P(y_i = 1 \mid \eta = 2)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(<em>0.8</em>)</td>
<td>0.03</td>
</tr>
<tr>
<td>2</td>
<td>0.8</td>
<td>0.03</td>
</tr>
<tr>
<td>3</td>
<td>0.8</td>
<td>0.03</td>
</tr>
<tr>
<td>4</td>
<td>0.8</td>
<td>0.36</td>
</tr>
<tr>
<td>5</td>
<td>0.8</td>
<td>0.36</td>
</tr>
</tbody>
</table>

### Structural model

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>$\mu_k$</th>
<th>$\phi_k$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1000</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>1000</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>1000</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

*Rule: 0.8, 0.95, 0.03*
Figure: Item characteristic curves for 4 senaria
Figure: Estimated latent means and variances for the three groups, 1 item out of five in one class non-invariant
Conclusions

- In most studies we found that what matters is the proportion of non-equivalent items.
- When that proportion is kept small the effect on the structural parameters is negligible.
- The presence of a class with small probabilities across all groups even when one item is non-invariant the effect can be severe.
- The only way to minimize that effect is by increasing the length of the scale.
- The number of groups was not found to change the effect of a high proportion of non-equivalence.
- The size of each group was not found to change the effect of a high proportion of non-equivalence.
- The structural parameters of the latent trait model were found more sensitive to non-equivalence.