Handling item non-response in Structural Equation Modelling with ordinal variables

Mysrini Katsikatsou, Irini Moustaki
Department of Statistics, London School of Economics, UK

1. Framework

- Ordinal variables (items), cross-sectional data, N independent observations.
- Structural equation modelling (SEM)

Let yi be the vector of ordinal items of dimension p, η the vector of factors, and yi the underlying continuous variable of the ordinal variable yi, where yi = a ↔ τi-1 < yi < τi, i = 1,...,p. a is the a-th response category of variable yi, a = 1,...,c, τi is the a-th threshold, τi-1 = -∞, and τi = +∞.

2.2 Treatment of missing values under PL

- All types of missing pattern (monotone/non-monotone) is allowed.

3. Research Objective

- Study the performance of CP-PL and AC-PL under MAR via a simulation study.
- Performance criteria:
  - Standardised Bias of parameter estimates, \( \sqrt{\text{bias}^2 + \text{se}^2} \),
  - Root Mean Square Error (RMSE) of parameter estimates \( \sqrt{\sum (\hat{\theta}_j - \theta_j)^2} \),
  - Coverage rate of the 95% confidence interval (CI),
  - Bias of standard errors \( \sum (\hat{\text{se}}_j - \text{se}_j) \),
  - Type I error rate of pairwise likelihood ratio test (PLRT) for overall goodness-of-fit at 5% and 1% significance level.

4. Simulation study design

10 experimental conditions
Sample size 300 & 1000 in each of the 5 experimental conditions below.

<table>
<thead>
<tr>
<th>Experimental Conditions</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factors</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Items</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Item loadings</td>
<td>0.5 for all y's</td>
<td>0.8 for y1</td>
<td>0.6 for all y's</td>
<td>0.3, 0.5, 0.7, 0.8, 0.9</td>
<td>0.3, 0.6</td>
</tr>
<tr>
<td>Factor correlation</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
</tr>
</tbody>
</table>

5. Simulation results

Loadings and factor correlation
- All three methods, CP-PL, AC-PL, and MI-DWLS, show acceptable performance regarding all performance criteria in all conditions.
- However, CP-PL and AC-PL exhibit lower standardised bias and a better coverage rate than MI-DWLS.
- CP-PL and AC-PL exhibit standardised bias and coverage rate fairly close to those of PL with complete data, especially for sample size 1000.
- A sample size increase seems to be associated with better performance in all criteria.
- A smaller factor loading for y1 which determines the level of MAR, seems to be associated with slightly larger bias of both estimates and standard errors but this is the case in PL with complete data as well.
- Smaller loadings in all items seem to be associated with larger RMSE.
- A higher factor correlation seems to be associated with improved coverage rate.

Thresholds
- MI-DWLS clearly outperforms CP-PL and AC-PL in all criteria.
- No clear preference between AC-PL and CP-PL as:
  - AC-PL exhibits acceptable standardised bias (average per condition up to 11%), while standardised bias in CP-PL may exceed 40%.
  - But, AC-PL systematically under-estimates the standard errors leading to unacceptable coverage rate. CP-PL shows similar levels of standard errors bias as MI-DWLS and acceptable coverage in most occasions.
- A hybrid PL, which uses the AC-PL threshold estimates and the corresponding CP-PL standard errors, exhibits acceptable coverage rate in all conditions.

PLRT for overall goodness-of-fit
- Both CP-PL and AC-PL show rates of Type I error very close to the nominal levels 5% and 1% with two exceptions:
  - in Ex. Con. 6 with sample size 300, where the rates are smaller than the nominal levels, and
  - in Ex. Con. 3, where the rates are a bit larger than the nominal ones, but actually this occurs in PL with complete data as well.

6. Discussion

- The general result that CP-PL and AC-PL yield biased estimates do not seem to hold in SEM, especially for loadings and factor correlations.
- Potential advantages of CP-PL and AC-PL over MI-DWLS: a) MI-DWLS requires a model for imputing, b) in MI-DWLS, it is no clear how to use the fit indices to judge overall fit.
- Worthy to develop a doubly-robust PL (Molenberghs et al., 2011) and compared it to CP-PL and AC-PL.

References

Acknowledgments: Many thanks to Prof. Yves Rosseel who incorporated our R code into his R package lavaan. The research is supported by ES/L009838/1 ESRC grant.