General structural model – Part 1: comparing groups, missing values, partitioning of effects

Comparing groups

- All modeling so far considered a single group --- multiple-group analysis fits a model simultaneously to several groups, typically with some equality constraints across groups (how different from modeling with a covariate for grouping?)

- Testing for group invariance built on the additive nature of the chi-square statistic:

$$\chi_T^2 = \sum_{g=1}^{G} \chi_g^2, \qquad df_T = \sum_{g=1}^{G} df_g$$

- Levels of comparability
  - Model form --- comparison mostly qualitative or best fitting model derived per group
  - Hierarchy of invariance --- given the same model form, sort out parameters; primary interest vs. secondary, observed vs. latent

Testing hierarchy: path modeling part

- There is no gold standard for order of invariance testing
- Any ordering effect is not arbitrary in that the LR test is subject to the common part of the nested and the nesting model
- For <u>strict</u> invariance/equivalence, lower-level invariance necessary for higher-level invariance, given a chosen order of testing

$$H_{\text{form}} : \text{same form}, \quad g = 1, \ldots, G$$

$$H_{\mathrm{B}\Gamma} : \mathbf{B}_g = \mathbf{B}_T,\ \boldsymbol{\Gamma}_g = \boldsymbol{\Gamma}_T$$

$$H_{\mathrm{B}\Gamma\Psi} : \mathbf{B}_g = \mathbf{B}_T,\ \boldsymbol{\Gamma}_g = \boldsymbol{\Gamma}_T,\ \boldsymbol{\Psi}_g = \boldsymbol{\Psi}_T$$

$$H_{\mathrm{B}\Gamma\Psi\Phi} : \mathbf{B}_g = \mathbf{B}_T,\ \boldsymbol{\Gamma}_g = \boldsymbol{\Gamma}_T,\ \boldsymbol{\Psi}_g = \boldsymbol{\Psi}_T,\ \boldsymbol{\Phi}_g = \boldsymbol{\Phi}_T$$

Testing hierarchy: measurement part

- Same hierarchy, but regression weights are factor loadings instead of latent path coefficients

$$H_{\text{form}} : \text{same form}, \quad g = 1, \ldots, G$$

$$H_{\Lambda_x} : \mathbf{\Lambda}_x^{(g)} = \mathbf{\Lambda}_x^{(T)}$$

$$H_{\Lambda_x \Theta_\delta} : \mathbf{\Lambda}_x^{(g)} = \mathbf{\Lambda}_x^{(T)}, \ \mathbf{\Theta}_\delta^{(g)} = \mathbf{\Theta}_\delta^{(T)}$$

$$H_{\Lambda_x \Theta_\delta \Phi} : \mathbf{\Lambda}_x^{(g)} = \mathbf{\Lambda}_x^{(T)}, \ \mathbf{\Theta}_\delta^{(g)} = \mathbf{\Theta}_\delta^{(T)}, \ \mathbf{\Phi}^{(g)} = \mathbf{\Phi}^{(T)}$$

- For y-measurement models, invariance on $\mathbf{\Phi}$ irrelevant since covariances between $\boldsymbol{\eta}$'s are modeled as functions of parameters, not parameters themselves
- How should the LV scales be fixed for multiple groups?

Invariance in mean structure

- When structural equations include mean-structure, there is more flexibility in testing order of different kinds of parameters
- Yet, wide consensus on <u>minimum</u> requirement: invariance in model form and loading pattern needed for mean-structure invariance since $E(\mathbf{x}) = \boldsymbol{\nu}_x + \boldsymbol{\Lambda}_x\boldsymbol{\kappa},\ \boldsymbol{\Sigma}(\boldsymbol{\theta}) = \boldsymbol{\Lambda}_x\boldsymbol{\Phi}\boldsymbol{\Lambda}'_x,$ thus

$$H_{\text{form}} : \text{same form,} \quad g = 1, \ldots, G$$

$$H_{\Lambda_x} : \boldsymbol{\Lambda}_x^{(g)} = \boldsymbol{\Lambda}_x^{(T)}$$

$$H_{\Lambda_x \nu_x} : \boldsymbol{\Lambda}_x^{(g)} = \boldsymbol{\Lambda}_x^{(T)},\ \boldsymbol{\nu}_x^{(g)} = \boldsymbol{\nu}_x^{(T)}$$

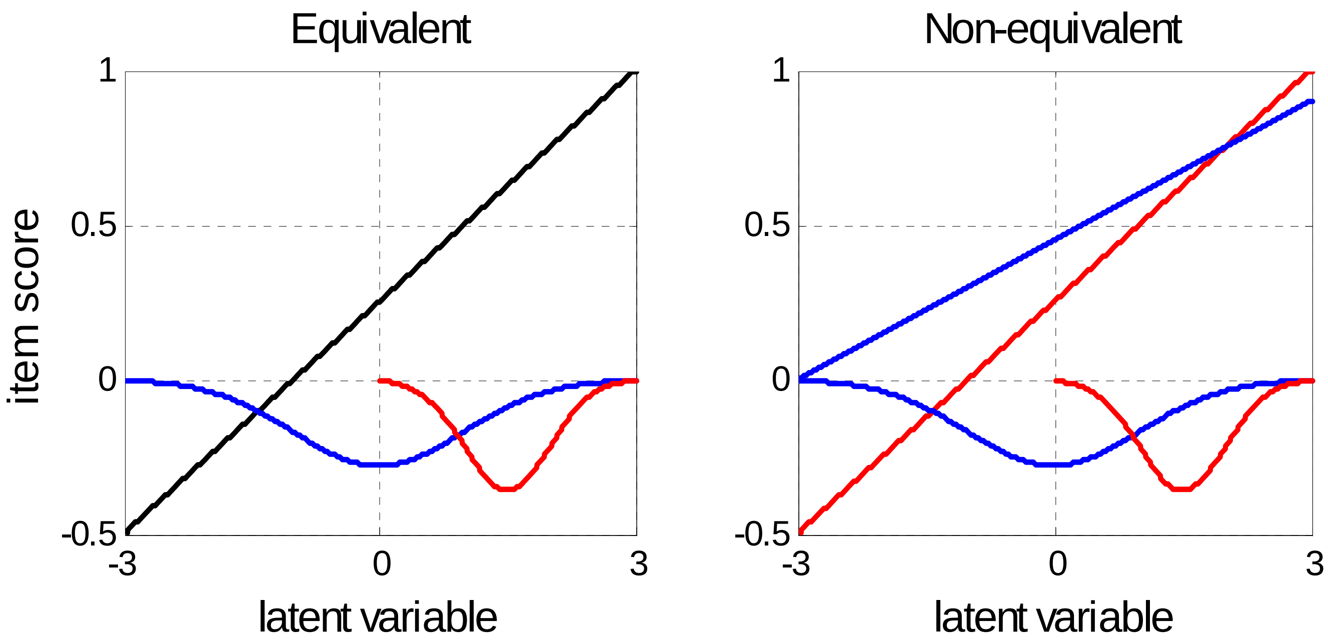
$$H_{\Lambda_x \nu_x \kappa} : \boldsymbol{\Lambda}_x^{(g)} = \boldsymbol{\Lambda}_x^{(T)},\ \boldsymbol{\nu}_x^{(g)} = \boldsymbol{\nu}_x^{(T)},\ \boldsymbol{\kappa}^{(g)} = \boldsymbol{\kappa}^{(T)}$$

- The Morale data (morale.xls; Table 8.7, p. 367) available for practice

Measurement equivalence

- Measurement of an LV is equivalent/invariant across distinctive groups if subjects (as random observational units) with identical LV scores are expected to have the same scores on indicators, regardless of group membership

- What's invariant: measurement parameters, not LV distribution



Why bother?

- It's crucial to establish measurement equivalence for any comparison involving LVs to be meaningful since the scale of LVs are arbitrarily chosen

  - Cases when measurement equivalence is needed --- distinctive groups (e.g., males vs. females, 3rd vs. 5th graders as different age cohorts), repeated (longitudinal) measures (e.g., children measured at 3rd and 5th grades), alternative forms or raters, etc.

- Testing measurement equivalence boils down to testing equality of the loadings and intercepts (i.e., measurement parameters under the linear model)

$$H_0 : \boldsymbol{\Lambda}_x^{(g)} = \boldsymbol{\Lambda}_x^{(T)}, \quad \boldsymbol{\nu}_x^{(g)} = \boldsymbol{\nu}_x^{(T)}, \quad g = 1, \ldots, G$$

Equating for distinctive groups

- Given estimated loadings and intercepts of a reference and a focal group (with conventional scaling within groups, e.g., 0 mean of LVs and same metric as one indicator for all compared groups), we need to find two constants for a linear transformation of the focal group's LV metric to the reference's

- Any equating procedure assumes a null hypothesis of equivalent measurement relationships up to a linear rescaling

- The two transformation constants are sought so that accordingly transformed parameter sets of the focal group maximally resemble the reference group's

Linear equating

- Suppose the same LV is arbitrarily scaled separately within groups to have a convenient scale

$$X_j^{(r)} = \nu_j^{(r)} + \lambda_j^{(r)} \xi^{(r)} + e_j^{(r)}, \quad X_j^{(f)} = \nu_j^{(f)} + \lambda_j^{(f)} \xi^{(f)} + e_j^{(f)}$$

- Since the group-specific metrics are arbitrarily chosen, we want to find a linear transformation of the focal group $\xi^{(f)} = c\xi^{(r)} + d$ such that the measurement equivalence holds maximally

  After substitution, we have the focal group's data represented in the reference group's metric $\tilde{\xi}^{(f)}$ and accordingly equated $\tilde{\nu}_j^{(f)}$ and $\tilde{\lambda}_j^{(f)}$

$$X_j^{(f)} = \left(\nu_j^{(f)} + \lambda_j^{(f)} d\right) + c\lambda_j^{(f)} \tilde{\xi}^{(f)} + e_j^{(f)}$$

$$\tilde{\nu}_j^{(f)} = \nu_j^{(f)} + d\lambda_j^{(f)}, \quad \tilde{\lambda}_j^{(f)} = c\lambda_j^{(f)}$$

- Under the linear measurement model, optimal transformation constants can be found by the least-squares estimator

$$\hat{c}_f = \boldsymbol{\lambda}^{(r)\prime}\boldsymbol{\lambda}^{(f)}\left(\boldsymbol{\lambda}^{(f)\prime}\boldsymbol{\lambda}^{(f)}\right)^{-1}$$

$$\hat{d}_f = \left(\mathbf{v}^{(r)} - \mathbf{v}^{(f)}\right)'\boldsymbol{\lambda}^{(f)}\left(\boldsymbol{\lambda}^{(f)\prime}\boldsymbol{\lambda}^{(f)}\right)^{-1}, \quad f = 1,\ldots,G-1$$

which minimizes

$$k_1^f = \sum_{j=1}^{q}\left(\tilde{\lambda}_j^{(f)} - \lambda_j^{(r)}\right)^2, \quad k_2^f = \sum_{j=1}^{q}\left(\tilde{\nu}_j^{(f)} - \nu_j^{(r)}\right)^2$$

(f), (r) --- focal and reference group’s parameters, respectively (arbitrarily scaled)

tilde --- transformed parameters to the reference group’s metric

Missing mechanisms

- MCAR (missing completely at random) --- probability of missing does not depend on value of X whatsoever

$$\Pr\left(\text{missing}\middle|X_{\text{com}}\right)=\Pr\left(\text{missing}\right),\qquad X_{\text{com}}=\left(X_{\text{obs}},X_{\text{mis}}\right)$$

- MAR (missing at random) --- probability of missing does not depend on the values of missing X, but may depend on values of other variables

$$\Pr\left(\text{missing}\middle|X_{\text{com}}\right)=\Pr\left(\text{missing}\middle|X_{\text{obs}}\right)$$

- Non-ignorable --- missing dependent on what's missing

$$\Pr\left(\text{missing}\middle|X_{\text{com}}\right)\neq\Pr\left(\text{missing}\middle|X_{\text{obs}}\right)$$

Non- or less-optimal missing-value treatment

- Listwise-deletion --- consistent and inefficient under MCAR; often severe data loss

- Pairwise-deletion --- less data loss, but overall N is unknown; also consistent and inefficient under MCAR

- Mean imputation --- poorest; equivalent to say “all I can predict is the univariate mean given incomplete information on covariation with other variables”, equivalent to imputing with expectation of implied covariances equal to 0, and thus leading to inconsistent results

Better treatments /w conditional imputation

- Multiple-group approach --- each missing pattern split into a separate group and treated as distinctive; an optimal way of missing treatment but practical only for a small number of missing patterns, impractical in most SEM cases; assumes MAR
- FIML (full information ML) --- likelihood function separately computed per subgroup with the same missing pattern, more feasible, assumes MAR (offered in AMOS)
- Multiple imputation: not only optimal imputation under MAR but also several imputations to provide more accurate standard errors

Effects partitioning

- Direct effect: influence of one variable on another, not mediated by anything else

- Indirect effects: causal influences mediated by at least one other variable

- Total effect = direct effect + all indirect effects

$$\mathbf{T}_{\eta\eta} = \underset{\substack{\uparrow \\ \mathbf{D}_{\eta\eta}}}{\mathbf{B}} + \underbrace{\mathbf{B}^2 + \mathbf{B}^3 + \cdots \mathbf{B}^{\infty}}_{\mathbf{H}_{\eta\eta}}$$

- If $\mathbf{B}^{\infty}$ converges to $\mathbf{0}$, indirect $(\mathbf{H})$ and total $(\mathbf{T})$ effects are defined; refer to Table 8.9 (p. 382) for various decomposed effects

Specific indirect effects

- Refers to indirect effects only through a particular mediator
  - Step 1 --- compute decomposed indirect effects with altered $\mathbf{B}$ (0 replacing column and row for the chosen mediator) and $\mathbf{\Gamma}$ (0 replacing the corresponding row) for the effects not through the chosen mediator
  - Step 2: subtract the resulting effects from the (total) indirect effects
- Similar change can be done for a particular pathway (with only the entries of $\mathbf{B}$ and $\mathbf{\Gamma}$ that are included in the pathway are replaced by 0) so as to obtain specific indirect effects only through that pathway