InterModel Vigorish for Model Comparison in Confirmatory Factor Analysis with Binary Outcomes

Lijin Zhang[1], Zhiyong Zhang[2], Klint Kanopka[1],

Esther Ulitzsch[3], Ben Domingue[1]

[1] Graduate School of Education, Stanford University

[2] Department of Psychology, University of Notre Dame

[3] Leibniz Institute for Science and Mathematics Education

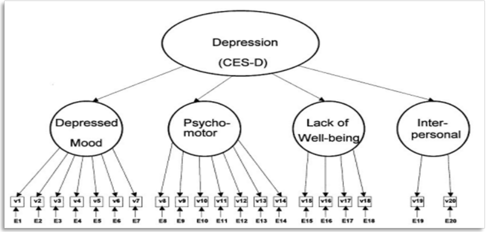
July 5, 2023

2023 ISDSA Meeting, Shanghai

Content

- Introduction
  - Confirmatory Factor Analysis with Binary Outcomes
  - Traditional Model Fitting Indices
- InterModel Vigorish in CFA
- Simulation Studies
  - Simulation 1: InterModel Vigorish in Model Selection
  - Simulation 2: IMV vs Traditional Fitting indices
- Empirical Study
- Discussion

Confirmatory Factor Analysis



A CFA model for J items $\mathbf{y}_i = (y_{i1}, \cdots, y_{iJ})^T$:

$$\mathbf{y}_i = \boldsymbol{\mu} + \boldsymbol{\Lambda}\boldsymbol{\omega}_i + \boldsymbol{\epsilon}_i, \quad i = 1, 2, \cdots, n, \tag{1}$$

- $\boldsymbol{\Lambda}$ factor loading
- $\boldsymbol{\omega}_i$ factor score
- $\boldsymbol{\mu}$ intercept
- $\boldsymbol{\epsilon}_i$ residual error

- Has been widely used to assess the fit of a theoretical measurement model to observed data.

Traditional Fitting Indices

Comparative Fit Index (CFI), Tucker-Lewis Index (TLI):

- > 0.95 indicates good model fitting

Root Mean Square Error of Approximation (RMSEA), Standardized Root Mean Squared Residual (SRMR):

- < 0.08 indicates good model fitting

Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC)

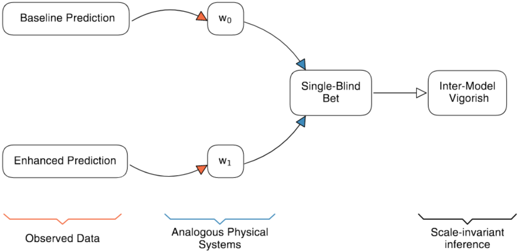
χ^2 Difference Test

Limitations

	CFI	TLI	RMSEA	SRMR	χ^2	AIC	BIC
Fixed cutoff lacks generalizability	✓	✓	✓	✓	-	✓	✓
Hard to interpret the size of ΔIndex	✓	✓	✓	✓	✓	✓	✓
No penalty for model complexity	-	-	-	✓	✓	-	-
Cannot provide item-level information	✓	✓	✓	✓	✓	✓	✓
Require nested model comparison	-	-	-	-	✓	-	-

Note. ✓: *the index has this limitation.*

InterModel Vigorish



- Build a fair bet with a weighted coin (w_0) based on the baseline prediction: \$1 (you) vs $\$\frac{1-w_0}{w_0}$ (your opponent).
- Updated the coin with the enhanced prediction (w_1).
- What is your expected win / lost?

$$\text{IMV} = \frac{1 - w_0}{w_0} * w_1 - 1 * (1 - w_1) = \frac{w_1 - w_0}{w_0} \tag{2}$$

Weight and Prediction

Connect weight with prediction: Suppose there is a binary variable $\mathbf{x}$ and the predicted probability from a model is p_i of $x_i = 1$. The mean log-likelihood would be:

$$A = \frac{1}{n} \sum_i (x_i \log p_i + (1 - x_i) \log(1 - p_i)) \tag{3}$$

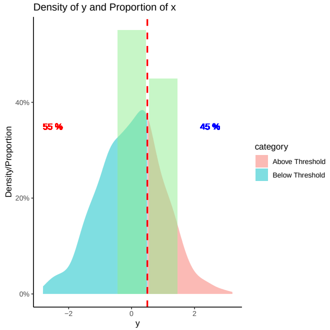
Create a coin with weight w that can produce the same log-likelihood:

$$\text{argmin}_w [|w \log(w) + (1 - w) \log(1 - w) - A|]. \tag{4}$$

Confirmatory Factor Analysis with Binary Outcomes

A CFA model for J dichotomous items $\mathbf{x}_i = (x_{i1}, \cdots, x_{iJ})^T$ can be conducted by introducing underlying continuous variables $\mathbf{y}_i$:

$$\begin{aligned} & \mathbf{y}_i = \boldsymbol{\mu} + \boldsymbol{\Lambda}\boldsymbol{\omega}_i + \boldsymbol{\epsilon}_i, \quad i = 1, 2, \cdots, n, \\ \text{where} \quad & x_{ij} = 1 \quad \text{if} \quad y_{ij} > \pi_j \\ & x_{ij} = 0 \quad \text{otherwise} \end{aligned} \tag{5}$$



Delta Parameterization

- μ_j is set at zero
- $Var(\mathbf{y}_i)$ is set at one

InterModel Vigorish in CFA

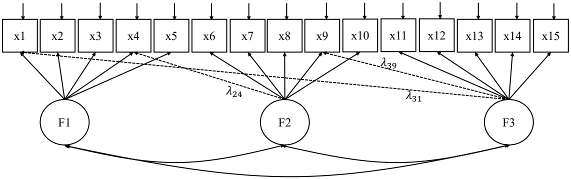
- Separate the datasets into the training set and the test set
- Identify parameter estimates with data in the training set
- Predict the responses in the test set (recall that the probability $Pr(x_{ij} = 1)$ is crucial in calculating the IMV values):

$$\begin{aligned} Pr(x_{ij} = 1 | \boldsymbol{\omega}_i, \boldsymbol{\Lambda}, \psi_{\epsilon j}, \pi_j) &= Pr(y_{ij} > \pi_j | \boldsymbol{\omega}_i, \boldsymbol{\Lambda}, \psi_{\epsilon j}, \pi_j) \\ &= \boldsymbol{\Phi}^* \left[\left(\boldsymbol{\Lambda}_j / \psi_{\epsilon j}^{1/2} \right) \boldsymbol{\omega}_i - \pi_j / \psi_{\epsilon j}^{1/2} \right] \end{aligned} \tag{6}$$

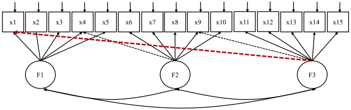
IMV versus Traditional Indicies

- Portability and Interpretability
  IMV offers an intuitive approach to assessing the benefits and costs of prediction accuracy on a standardized scale of $1.
- Item-level Information
  Could be useful in identifying specific model mis-specifications
- Avoid Model Overfitting
  shifts the researcher's focus from explanation to prediction
- Traditional Fitting Indicies in CFA with Binary Outcomes
  Robust χ^2 difference test: nested model comparison
  AIC BIC: not avaliable
  CFI TLI RMSEA SRMR: not designed for model comparison

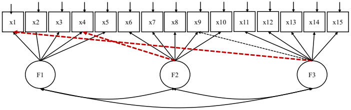
Simulation Design



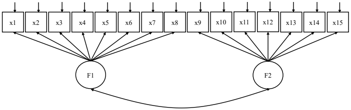
- Sample Size: 250, 500, 800, 1000, 200
- Effect Size of Cross-loadings: 0.1, 0.2, 0.3, 0.4, 0.5



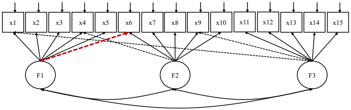
Condition 1



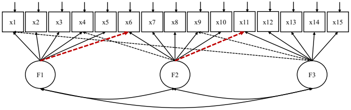
Condition 2



Condition 3



Condition 4



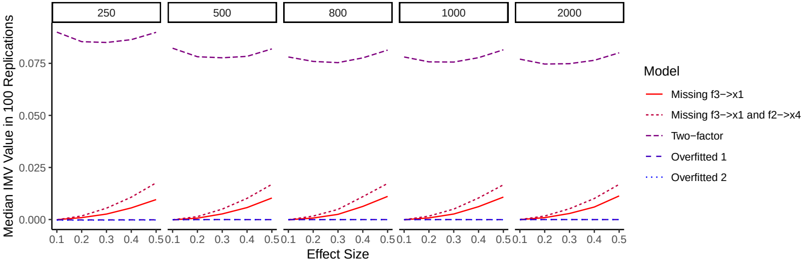
Condition 5

Model Comparison

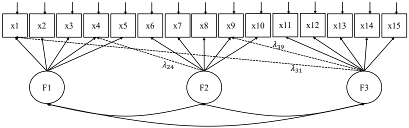
1. IMV(p_{M_1}, p_{M_0}): M_1 is the model missing one cross-loading (λ_{31});
2. IMV(p_{M_2}, p_{M_0}): M_2 is the model missing two cross-loading (λ_{31});
3. IMV(p_{M_3}, p_{M_0}): M_3 is a two-factor model;
4. IMV(p_{M_0}, p_{M_4}): M_4 is a over-fitted model which includes one more cross-loading (λ_{16});
5. IMV(p_{M_0}, p_{M_5}): M_5 is a over-fitted model with two more cross-loadings (λ_{16} and $\lambda_{2,11}$);

Scale-level IMV

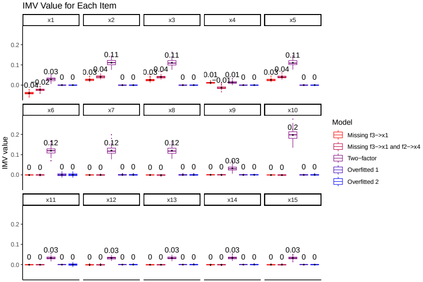
- IMV values effectively reflect the extent of model mis-specification.

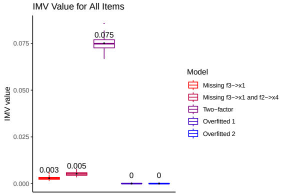


Item-level IMV ($N = 2000$, effect size $= 0.3$)



True Model





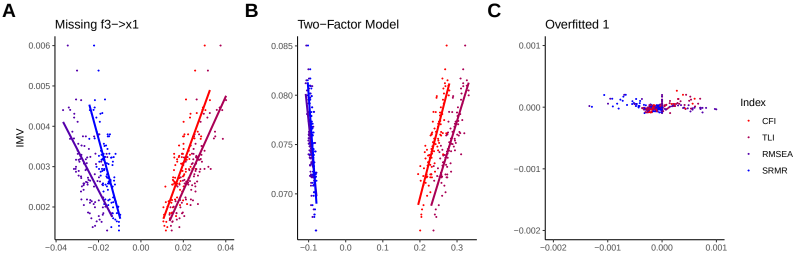
- IMV value for x_{10} (0.2) is much higher than $x_9, x_{11} - x_{15}$ (0.03).
- True model vs Model omits λ_{31}: Prediction accuracy for other items in F1 increased ($x_2 - x_4$, IMV $\approx$ .02).

Simulation Study 2

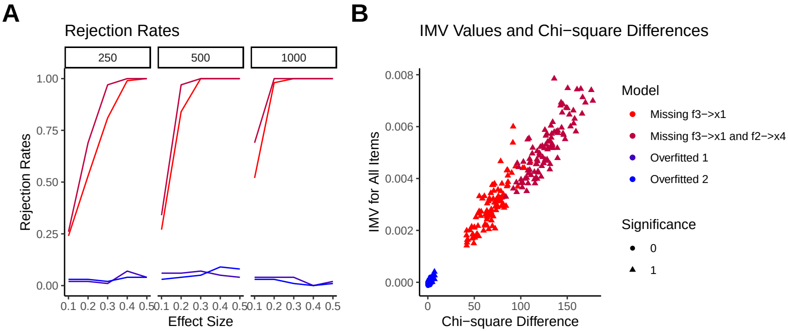
- Traditional Fitting Indicies: CFI, TLI, RMSEA, SRMR, and χ^2 (Robust versions were used)
- Fitness of individual models:
  - CFI TLI should be larger than 0.95, RMSEA, SRMR should be smaller than 0.08
  - In modeling conditions 1, 2, 4, and 5, almost all indicies show an excellent model fit.
  - These cutoffs lack generalizability.

Correlation: IMV and the Traditional Indicies

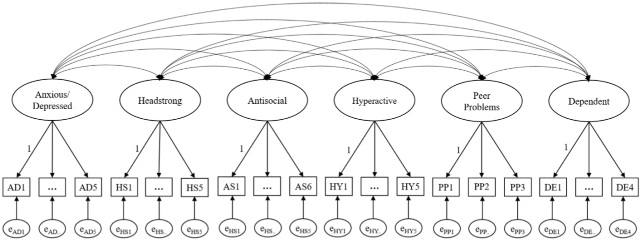
- IMV exhibits a strong correlation with changes in traditional indices, while being much more straightforward to interpret.
- SRMR supports the overfitted model over the true model.



Robust χ^2 Difference Test



Empirical Study



- Behavior Problem Index Scale
- Data (Kim et al., 2021): $N = 469$, 28 items, six factors

Model Comparison

1. Six-factor model vs Outcome prevalance (ignoring correlation between items): IMV = .179
2. Randomly combined two factors (peer problems and dependent):

Table: Comparison between the 5-factor and 6-factor Models.

Indicies	5-factor Model	6-factor Model
CFI	.958	.962
TLI	.954	.958
RMSEA	.039	.037
SRMR	.094	.089
IMV for all items	.005	
IMV for pp1	.020	
IMV for pp2	.011	
IMV for pp3	.009	
IMV for de1	.012	
IMV for de2	.019	
IMV for de3	.054	
IMV for de4	.021	

Discussion

- Portability across different contexts
- IMV does not rely on a single cutoff to determine if the model fit has significantly improved or not. Instead, it prioritizes the magnitude or "effect size" of the improvement.
- Traditional indices and IMV evaluate different facets of model fitting. the former focuses on how well the model fits the current data, while the latter emphasizes the model prediction.
- Item-level information facilitates targeted model modifications.

- Slides: `https://lijinzhang.com/share/230705_imv.pdf`

Thanks!