Title: The People Are Weird Fit Index

Author:

David Dueber

[david.dueber@uky.edu](mailto:david.dueber@uky.edu)

251C Dickey Hall, KY 40506

University of Kentucky

Department of Educational, School, and Counseling Psychology

Co-Author:

Michael D. Toland, University of Kentucky, toland.md@uky.edu.

Poster Session

Division – 5, Quantitative and Qualitative Methods

The People Are Weird Fit Index

“Nature hath framed strange fellows in her time.” – William Shakespeare

A debate rages within the heart of the structural equation modeling community regarding model fit, with some arguing that only the chi-square test of exact fit can provide a scientifically rigorous assessment of the model (Barrett, 2007; Hayduk, 2014a), whereas others argue that the chi-square test is too stringent, and models with only minor misspecification are worth retaining (Asparouhov & Muthén. 2018; Yuan et al., 2015). Within this great conversation, assessment of model fit is generally performed using chi-square tests, approximate fit indices based on chi-square (e.g., RMSEA, CFI, TLI), and indices based on residual covariances (e.g., SRMR). As such, misfit is necessarily posited to be caused by un-modeled dependencies amongst variables (Hayduk, 2014b) or unexplainable idiosyncracy in the data (Kaplan, 2008; Kline, 2015).

Yet, when stepping back from this battleground, the thought occurs that perhaps blame for misfit is being attributed to the wrong suspect; sometimes, misfit may simply be the result of the presence of weird people in the sample. Hypothetically, consider including a group of Pastafarians as participants in a survey measuring religiosity. Surely, their inclusion would induce misfit because Pastafarianism stands in direct conflict to traditional religions. With this in mind, we propose the People Are Weird Fit Index (PAWFI), which is the proportion of the sample which must be removed for the model to exhibit exact fit.

To compute PAWFI[[1]](#footnote-1), participants are sequentially removed from the sample until the model exhibits exact fit. Participants are selected for removal by constructing a jackknife distribution of the model chi-square; the person whose removal results in the largest decrease in model chi-square is selected for removal. This process is repeated until the model chi-square is non-significant. PAWFI is the ratio of the number of removed persons to the original sample size.

To illustrate the use of PAWFI, data was simulated in which two groups with different population models were combined and analyzed together using the population model structure of the larger group. Results (Table 1) show that the list of people removed by PAWFI are sufficient to suggest that the two groups do not share a common measurement model. Thus, the PAWFI removal list is used in an exploratory fashion to identify groups for which the model may not be working as desired, despite the group size not being large enough for measurement invariance testing. Additionally, PAWFI may identify members of a misfitting group that is not considered a relevant grouping a priori, so that information from the PAWFI removal list may be used to inform a theory of why some people are misfitting.

Survey data collected from four summer camps was analyzed using PAWFI. One site had over 40% (19/46) of its participants removed by PAWFI. This led to researchers questioning the survey administration protocol. A larger proportion of participants at the site were flagged by PAWFI than by the IRT Zh person fit index (Drasgow et al., 1985).

Simulations were performed to compare the performance of PAWFI and common approximate fit indices when some of the people in the sample are mis-fitting. The proportion of mis-fitting persons in the sample and the total sample size were manipulated; results for PAWFI, RMSEA, CFI, and SRMR can be found in Table 2. Somewhat surprisingly, PAWFI is much smaller than the proportion of mis-fitting persons; only approximately one tenth of mis-fitting persons were identified by PAWFI. The approximate fit indices tend to show better ability to detect model misfit. However, since one type and severity of person-misfit was considered in these simulations, no general conclusions can be reached about PAWFI’s ability to identify mis-fitting persons.

Simulations were performed to determine whether PAWFI can serve as an approximate fit index. For conditions in which the analysis model was identical to the population model, PAWFI is very near zero (Table 3). This is not surprising since PAWFI is only non-zero when the chi-square test of model fit is significant. For conditions in which the analysis model differed from the population model, PAWFI showed lower sensitivity to misfit, but similar sensitivity at larger sample size. Unfortunately, PAWFI seems sensitive to sample size (Table 4) and therefore would not make an effective approximate fit index.

Unfortunately, as the rather preliminary results presented herein demonstrate, PAWFI is not a terribly useful index. Research into a more robust person-based fit index is ongoing. The next iteration of PAWFI under consideration utilizes only a single round of jackknife chi-square values. The difficulty with this approach is in determining an appropriate cutoff for flagging persons as misfitting. Perhaps using a simulation based approach for creating a distribution of person removal chi-square change will yield an appropriate cutoff (e.g., 95th percentile of that distribution). Besides its practical use, the development of an effective person oriented approximate fit index would help broaden the conversation about model fit.

**References**

Asparouhov, T, & Muthén, B. (2018). *SRMR in Mplus*. Technical Report. May 2, 2018. http://www.statmodel.com/download/SRMR2.pdf

Barrett, P. (2007). Structural equation modelling: Adjudging model fit. *Personality and Individual Differences, 42*(5), 815-824.

Drasgow, F., Levine, M. V., & Williams, E. A. (1985) Appropriateness measurement with polychotomous item response models and standardized indices. *British Journal of Mathematical and Statistical Psychology*, 38, 67-86.

Hayduk, L. A. (2014a). Shame for disrespecting evidence: the personal consequences of insufficient respect for structural equation model testing. *BMC Medical Research Methodology, 14*(1), 124-133.

Hayduk, L. (2014b). Seeing perfectly fitting factor models that are causally misspecified: Understanding that close-fitting models can be worse. *Educational and Psychological Measurement, 74*(6), 905-926.

Kaplan, D. (2008). *Structural equation modeling: Foundations and extensions* (Vol. 10). Sage.

Kline, R. B. (2015). *Principles and practice of structural equation modeling*. Guilford.

Yuan, K. H., Chan, W., Marcoulides, G. A., & Bentler, P. M. (2016). Assessing structural equation models by equivalence testing with adjusted fit indexes. *Structural Equation Modeling: A Multidisciplinary Journal, 23*(3), 319-330.

Table 1

*Detection of Participants from Group with Mis-specified Model*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Type of misfit in group 2** | ***N*  (group 1)** | ***N*  (group 2)** | **Average PAWFI** | **Total % of removed persons from group 2** | **% of replications with at least 5 people removed from group 2** |
| Loading set to zero for four items | 375 | 125 | .016 | 78.0 | 67.0 |
| Two factor model instead of one | 320 | 80 | .029 | 92.5 | 93.0 |

*Note*. *N* = Sample size. Population model for Group 1 was an eight item unidimensional factor model with standardized loadings of 0.707. Population model for Group 2 was the same as Group 1 except for misfit described in table. Analysis model was the configural population model for Group 1. PAWFI = People Are Weird Fit Index.

Table 2

*Comparison of PAWFI and Common Approximate Fit Indices in the Presence of a Mis-fitting Subgroup*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Proportion Misfits** | **Sample Size** | **Mean PAWFI** | **Mean RMSEA** | **Mean CFI** | **Mean SRMR** |
| 10% | 250 | .01 | .05 | .97 | .03 |
| 10% | 500 | .01 | .05 | .98 | .03 |
| 20% | 250 | .02 | .09 | .94 | .05 |
| 20% | 500 | .03 | .09 | .94 | .05 |

Table 3

*Results for PAWFI When Applied to Correctly Specified Models*

|  |  |  |  |
| --- | --- | --- | --- |
| **Model** | ***N*** | **Average PAWFI** | **Proportion of replications with PAWFI = 0** |
| Unidimensional | 250 | < .001 | .92 |
| Unidimensional | 1000 | .001 | .88 |
| Two-Factor | 250 | < .001 | .93 |
| Two-Factor | 1000 | < .001 | .95 |

*Note*. *N* = Sample size. Unidimensional model was a 12-item unidimensional model with standardized loadings of 0.707. Two-factor model consisted of two eight-item factors with standardized loadings of 0.707 and an inter-factor correlation of 0.2. PAWFI = People Are Weird Fit Index.

Table 4

*Results for PAWFI When Applied to Mis-specified Models*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Type of misfit** | **Sample Size** | **Average PAWFI** | **Average RMSEA** | **Average CFI** | **Average SRMR** |
| Moderate error covariance | 250 | .058 | .068 | .958 | .036 |
| Moderate error covariance | 1000 | .110 | .069 | .959 | .028 |
| Moderate crossloading | 250 | .113 | .069 | .929 | .090 |
| Moderate crossloading | 1000 | .199 | .070 | .928 | .085 |

*Note*. Population model for condition with moderate error covariance was a twelve item unidimensional model with standardized loadings of 0.707 and an error correlation of 0.5 between two items; analysis model was a unidimensional model with no correlated errors. Population model for condition with moderate crossloading consisted of two eight-item factors with standardized loadings of 0.707 and an inter-factor correlation of 0.2 except for a single item which had a standardized loading of 0.577 on both factors; analysis model was a two-factor model with no crossloadings. PAWFI = People Are Weird Fit Index.

1. R code for PAWFI as well as code for the simulations conducted for this abstract are available on github: https://github.com/ddueber/PAWFI [↑](#footnote-ref-1)