Alavi M, Visentin D, Deepandra T, Hunt G, **Watson R**, Cleary M Referencing to be tidied up later

Chi-square for model fit in confirmatory factor analysis

1. What is model fit?

Confirmatory Factor Analysis (CFA) aims to confirm a theoretical model using empirical data and is an element of the broader multivariate technique Structural Equation Modelling (SEM) (Alavi et al., 2020). A fundamental characteristic of CFA is its hypothesis driven approach (Brown, 2015). The researcher first establishes a hypothesis regarding the model structure expressed as particular factor(s) underlying a set of items. Analysis is then performed to determine how much of the covariance between the items would be captured by the hypothesized factor structure. In addition to assessing the covariance captured by the model evaluating the goodness of fit of the proposed model, which reflects how well the model fits the observed data, is a critical step in CFA (Hooper, Coughlan, & Mullen, 2008).

Model fit is evaluated using a range of model fit indices which assess the relationship between the observed data and the data which would be expected from the model. Model fit indices can be used with either thresholds or hypothesis testing to reject or retain the proposed model (Sarmento & Costa, 2019).

2. Categories of model fit indices

There are two types of fit indices available for CFA; global and local fit indices (Brown, 2015; Kline, 2005). Global model fit indices attempt to quantify the overall recovery of the observed data without considering specific components of the mean and covariance structure. In addition to global fit indices, local fit indices are evaluated to examine model components including factor correlations, inter-item residual covariance, and suggested model re-specification statistics.

Global model fit indices fall into three categories: absolute, incremental (also known as comparative or relative), and parsimony fit indices (Hooper et al., 2008; Kline, 2005). Absolute fit indices assess the overall theoretical model against the observed data. They are often a function of the test statistic, which quantifies global fit to the population covariance structure. Absolute fit indices can also be a function of the model residuals. They assess how well the model fits the data compared to no model. In addition to the chi-square (χ^2) statistic other examples of absolute fit indices are , Goodness-of-Fit Index (GFI), Adjusted Goodness-of-Fit Index (AGFI), Root Mean Square Error of Approximation (RMSEA) , Root Mean Square Residual (RMR) and Standardized Root Mean Square Residual (SRMR) (K G Jöreskog & Sörbom, 1989; Steiger, 2007).

Incremental fit indices compare a candidate model against a baseline model. The baseline model is a minimal model that specifies no relationships between the variables and contains only variances for observed variables. Hence, the baseline model represents the hypothesis that there are no meaningful relationships between variables. Incremental fit indices represent the improved fit for the model compared to the assumption of independence of variables. Examples are Comparative Fit Index (CFI), Normed-Fit Index (NFI) and Non-Normed Fit Index (NNFI) (Bentler & Bonett, 1980)(Bentler, 1990).

As parameters are added to a model, the model fit will improve. Parsimonious fit indices aim to address this issue for more complex models by adding a penalty for the inclusion of additional parameters. This introduces a trade-off of between model fit and degrees of freedom. Parsimony Goodness-of-Fit Index (PGFI) (Mulaik et al., 1989) and the Parsimony Normed Fit Index (PNFI) (James, Mulaik, & Brett, 1982) are examples. It is not expected that all indices would be used or reported in fitting any hypothesized model. The application of each fit index depends on the purpose of analysis and characteristics of the fit indices.

3. The chi-square fit index

The chi-square fit index assesses the fit between the hypothesised model and data from a set of measurement items (the observed variables). The model chi-square is the chi-square statistic obtained using maximum likelihood. When a model is estimated using maximum likelihood, the respective test statistic (i.e. likelihood ratio) is commonly used to assess the overall goodness of fit (K. G. Jöreskog, 1969; Maydeu-Olivares, Fairchild, & Hall, 2017). Assuming the hypothesised model is correctly specified, the likelihood ratio test statistic would approach a central chi-square distribution. The model chi-square test is the most commonly used global fit index in CFA and is also a component of several other fit indices. It tests the null hypothesis that the model-implied covariance matrix represents the population covariance. Generally, chi-square is used as an absolute fit index, with a low chi-square value relative to the degrees of freedom (and higher p-value) indicating better model fit. Since the test is used to reject a null hypothesis representing perfect fit, chi-square is often referred to as a 'badness of fit' or 'lack of fit index' (Kline, 2005).

Chi-square statistic using likelihood ratio tests can also be used to assess nested models, where the free parameters of one model are a subset of an alternative model. In this situation, the difference in model fit is expressed as the difference in model chi-square values, which is also distributed according to a chi-square distribution. The degrees of freedom for nested likelihood ratio tests are the added parameters for the less parsimonious model (Kline, 2005; Tomarken & Waller, 2003).

As with all tests the assumptions of the chi-square model fit index must be met. These assumptions include multivariate normality of data, no systematic missing data, adequate sample size and correct model specification.

4. Limitations for the chi-square model-fit-index

There are limitations with using the chi-square statistic as a model fit index. Firstly, it is sensitive to sample size with larger sample sizes decreasing the p-value. The chi-square test can be ‘significant’ in cases where there is only a trivial misfit, especially in large samples (Babyak & Green, 2010). Overemphasis on model chi-square may lead to a preference for smaller samples in which the null hypothesis (of good fit) is not rejected. This is more likely to accept poor models and may yield inaccurate or imprecise parameter estimations. Parameter estimates should be given consideration rather than merely model fit indices, as they often hold substantive clinical interest.

The model statistic does not always follow a chi-square distribution particularly in cases where data are not multivariate normal or when the sample size is small. As with any statistical test it is often interpreted as a binary result, in this case a fit or no-fit decision resulting in the model being rejected or retained. Assessment of the test statistic itself which indicates the degree to which a model is discrepant should be preferred.

Chi-squared model fit is a non-parsimonious approach and hence the model fit improves as the model size increases (Schermelleh-Engel, Moosbrugger, & Müller, 2003). Increasing the number of parameters may provide unnecessarily complex models which are likely to be accepted than parsimonious ones. The complexity of the model needs to be considered when assessing model fit using chi-square.

5. Summary and recommendations

Given that the chi-squared fit statistic is affected by large samples, the ratio of the chi-square statistic to the respective degrees of freedom is preferred (Wheaton, Muthen, Alwin, & Summers, 1977). A ratio of ≤ 2 indicates a superior fit between the hypothesized model and the sample data (Cole, 1987). Nevertheless, the chi-square statistic can be useful when a CFA model fails to fit. It is common then to enter an exploratory phase which involves inspecting the modification indices of all the pairs of error terms and correlating those pairs with the largest indices until the model fits (Watson, Chen, Ip, Smith, Wong, & Deary, 2013). As better fitting models are achieved the fit indices and the RMSEA improve by increasing and decreasing, respectively. The chi-square statistic should decrease but in with large sample sizes it will most probably remain statistically significant. Considering the sensitivity of the chi-square statistic to sample size, a wide variety of other indices have been suggested to assess model adequacy. In practice, the chi-square test is "not always the final word in assessing fit" (West, Taylor, & Wu, 2012, p. 211). Kline (2005) suggests that at a minimum the following indices should be reported and assessed in combination: chi-square; Root Mean Square Error of Approximation; Comparative Fit Index; and Standardized Root Mean Square Residual. The use of multiple fit indices provides a more holistic view of goodness-of-fit, accounting for sample size, model complexity and other considerations relevant to the particular study.

References

Commented [MC1]: This will be tidies up next draft

Alavi, M., Visentin, D. C., Thapa, D. K., hunt, G. E., Watson, R., & Cleary, M. (2020). Exploratory factor analysis and principal component analysis in clinical studies: Which one should you use? *Journal of Advanced Nursing, Article in press*

Babyak, M. A., & Green, S. B. (2010). Confirmatory factor analysis: an introduction for psychosomatic medicine researchers. *Psychosom Med, 72*(6), 587-597. doi:10.1097/PSY.0b013e3181de3f8a

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*(2), 238-246. doi:10.1037/0033-2909.107.2.238

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88*(3), 588-606. doi:10.1037/0033-2909.88.3.588

Brown, T. A. (2015). *Confirmatory Factor Analysis for Applied Research, Second Edition*. New York: Guilford Publications.

Cole, D. A. (1987). Utility of confirmatory factor analysis in test validation research. *Journal of consulting and clinical psychology, 55*(4), 584.

Hooper, D., Coughlan, J., & Mullen, M. R. (2008). Structural equation modelling: Guidelines for determining model fit. *Electronic journal of business research methods, 6*(1), 53-60.

James, L., Mulaik, S., & Brett, J. M. (1982). *Causal analysis: Assumptions, models, and data*. Beverly Hills, CA: Sage.

Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika, 34*(2), 183-202. doi:10.1007/BF02289343

Jöreskog, K. G., & Sörbom, D. (1989). *LISREL 7: A Guide to the Program and Applications*

Chicago, IL: SPSS.

Kline, R. B. (2005). *Principles and practice of structural equation modeling, 2nd ed*. New York, NY, US: Guilford Press.

Maydeu-Olivares, A., Fairchild, A. J., & Hall, A. G. (2017). Goodness of Fit in Item Factor Analysis: Effect of the Number of Response Alternatives. *Structural equation modeling: a multidisciplinary journal, 24*(4), 495-505. doi:10.1080/10705511.2017.1289816

Mulaik, S. A., James, L. R., Van Alstine, J., Bennett, N., Lind, S., & Stilwell, C. D. (1989). Evaluation of goodness-of-fit indices for structural equation models. *Psychological Bulletin, 105*(3), 430. doi:10.1037/0033-2909.105.3.430

Sarmento, R. P., & Costa, V. (2019). Confirmatory Factor Analysis--A Case study. *arXiv preprint arXiv:1905.05598*.

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the Fit of Structural Equation Models: Tests of Significance and Descriptive Goodness-of-Fit Measures. *Methods of Psychological Research, 8*(2), 23-74.

Steiger, J. H. (2007). Understanding the limitations of global fit assessment in structural equation modeling. *Personality and Individual Differences, 42*(5), 893-898. doi:10.1016/j.paid.2006.09.017

Tomarken, A. J., & Waller, N. G. (2003). Potential problems with" well fitting" models. *Journal of abnormal psychology, 112*(4), 578.

Watson, R., Chen, Y., Ip, M.Y.K., Smith, G.D., Wong, T.S.K., & Deary, I.J. (2013). The structure of stress: confirmatory factor analysis of a Chinese version of the stressors in Nursing Students Scale (SINS). *Nurse Education Today*, 33(2) 160-165.

West, S. G., Taylor, A. B., & Wu, W. (2012). Model fit and model selection in structural equation modeling. *Handbook of structural equation modeling, 1*, 209-231.

Wheaton, B., Muthen, B., Alwin, D. F., & Summers, G. F. (1977). Assessing reliability and stability in panel models. *Sociological methodology, 8*, 84-136. doi:doi.org/10.2307/270754