Deciding Between Competing Models: Chi-Square Difference Tests

1 Research Situation

Using structural equation modeling to investigate a research question, the simplest strategy would involve constructing just a single model corresponding to the hypotheses, test it against empirical data, and use a model fit test and other fit criteria to judge the underlying hypotheses. However, frequently, it is not particularly satisfactory to analyze a single model, but more appropriate to analyze several *competing models* and *compare the results*, e. g. in the following situations:

- The favorite model fits the data well, but there is a competing model based on different hypotheses which may explain the observed relationships as well. If possible, this competing model should be rejected (it should fit the data worse compared to the favorite model).
- There are several competing models which are all theoretically plausible. Differences in model fit would be the only criteria to decide which model to prefer.
- The original, presumed model did not fit the data well. It has been modified and it is to be shown that the modifications actually result in better model fit.

In particular, the following model comparisons may be made:

- A model with an additional path compared to an otherwise identical model without this path: Is there an effect between two latent variables or not? Is there a direct effect of a ξ-variable on an η-variable, or an indirect effect only?
- A model assuming a relationship between two latent variables compared to a model where these latent variables are presumed to be unrelated: Are the factors ξ_1 and ξ_2 independent of each other or not?
- A model with an additional loading of a manifest variable on a latent variable, compared to a model without such an additional loading: Is the manifest variable x_1 an exclusive indicator of construct ξ_1, or does it also measure aspects of a different latent variable ξ_2 at the same time?

A decision between competing models may be clear-cut if there are completely obvious differences in model fit criteria, or if a parameter in question turns out to be both insignificant ($|t| < 1.96$) and of marginal size (close to zero in the completely standardized solution).

However, frequently differences in model fit are more subtle, and an objective criterion for a decision between competing models may be desired. For this purpose, different models can be compared with regard to their model fit by computing a χ^2 **difference test**. This test allows to decide whether a given model fits significantly better or worse than a competing model.

2 Assumption

A χ^2 difference test is meaningful only if the models in question are **nested models**, i. e. one of the models could be obtained simply by *fixing/eliminating parameters* in the other model. When comparing models, this is frequently the case, e. g. in the situations described above where one model just contains

- an *additional* path in the structural model
- an *additional* loading in a measurement model
- an *additional* correlation/covariance between latent variables

which the other model does not contain (where the parameter in question is fixed to zero).

This test is not directly applicable to non-nested models containing *structurally different* parameters, e. g. a model in which x_1 serves as an indicator of latent variable ξ_1, compared to an alternative model where x_1 is an indicator of ξ_2 *instead.* In this case, there are the following alternatives:

- Testing each of the non-nested models against a common parent model in which all models in question are nested. Example: To compare a model assuming that x_1 is an indicator of the latent variable ξ_1 against a model that assumes x_1 to be an indicator of ξ_2 *instead*, a common parent model would contain *both* loadings (x_1 on ξ_1 as well as x_1 on ξ_2). Each of the models of interest could then be compared to the common parent model using χ^2 difference tests. Ideally, only one of the models would significantly differ from the parent model in terms of model fit, allowing to indirectly compare the non-nested models and decide between them.
- Descriptive model comparison using criteria suitable for non-nested models, for example AIC (Akaike Information Criterion). This does not allow significance testing, though.

3 Procedure

To compute a χ^2 difference test, the **difference of the χ^2 values** of the two models in question is taken as well as the **difference of the degrees of freedom**. Frequently, models under investigation differ from each other by just one more free parameter or one more fixed parameter, respectively, so in these cases the difference of the degrees of freedom is 1.

$$\chi^2_{\text{diff}} = \chi^2_{\text{s}} - \chi^2_{\text{l}} \quad \text{and} \quad df_{\text{diff}} = df_{\text{s}} - df_{\text{l}}$$

Here, s denotes the "smaller" model with fewer parameters and therefore more degrees of freedom, whereas l denotes the "larger" model with more parameters and therefore fewer degrees of freedom.

This χ^2_{diff}-value is distributed with df_{diff} degrees of freedom and can be checked manually for significance using a χ^2 table.

If the χ^2_{diff}-value is significant, the "larger" model with more freely estimated parameters fits the data better than the "smaller" model in which the parameters in question are fixed. So it "pays off" to estimate the additional parameters and to prefer the "larger" model. In case the χ^2_{diff}-value is insignificant, both models fit equally well statistically, so the parameters in question can be eliminated from the model (fixed to zero) and the "smaller" model can be accepted just as well.

Chi-square difference tests applied to nested models have essentially the same strengths and weaknesses as χ^2-tests applied to any single model: They are directly affected by sample size, and for large samples even trivial differences may become significant.

References

Bollen, K. A. (1989). *Structural equations with latent variables.* New York: John Wiley & Sons.

Kline, R. B. (1998). *Principles and practice of structural equation modeling.* New York: Guilford.

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 Online*, *8*, 23–74. Available from `http://www.dgps.de/fachgruppen/methoden/mpr-online/`