

A Terse Explanation to the Slides

Carel F.W. Peeters

8 februari 2016

The slides contained in **CFWP_APAC.2015.pdf** were used to shortly present my dissertation research. This dissertation (Peeters, 2012) received the 2015 Anne Anastasi Distinguished Dissertation Award. Below you will find a terse explanation to each slide, giving an idea of the storyline as used in the presentation.

1. Title page
2. The shortest possible summary of the dissertation: It's a layer-cake (also see Slide 8).
3. The dissertation considers the factor model, essentially a multivariate regression model with unobserved (latent) predictors. Factor analysis (FA) assumes that a vector of observed items consists of correlated variables that can be grouped into a lower number of latent factors. FA comes in two general flavors: Exploratory and Confirmatory. In exploratory factor analysis (EFA) both the dimensionality of the latent vector and the meaning of latent factors are unknown (left-hand figure). In the exploratory sense, FA is a theory-generating technique used for the identification of meaningful latent factors. Confirmatory FA (CFA) is a theory-testing technique. An a priori factor structure is assumed, with a given number of latent variables, with a pre-specified loadings matrix in which exclusion constraints indicate which variables are indicators of which latent factor(s), and with possibly correlated factors and error variances (right-hand figure). The λ parameter indicates a factor loading which is essentially a regression parameter.
4. The dissertation first seeks to contribute to EFA-efforts by proposing a Bayesian model-selection approach to the selection of the optimal dimension of the latent vector. Essentially, this means one is constructing a stopping rule: Assess all possible dimensions and choose the one that is optimal. In order to do this successfully, the approach must be able to cope with distortions due to overfactoring (i.e., fitting the FA model under more latent factors than the data-generating mechanism). The figure shows, represented by the bivariate densities on two factor loadings, what happens when one fits a 5-factor model to data generated under a 3-factor model: The (posterior) density falls apart in multiple separated regions of non-negligible posterior probability for which transition probabilities are low (or nil). This problem of rank deficiency accounts for the oft-observed fact that the likelihood-ratio test and information criteria have a tendency to overestimate model size in EFA, i.e., they tend to retain too many factors. The dissertation advocates an approach that copes with this problem. In passing, a truly Bayesian EFA is proposed.
5. The dissertation also seeks to contribute to CFA-efforts. The motivation can be found in the traditional specification of the loadings matrix in CFA-models. One can collect all λ parameters in the loadings matrix $\mathbf{\Lambda}$. The slide gives an example of a loadings matrix in a model with 6 items and 2 latent factors. In traditional CFA fixed-value equality restrictions are usually used to incorporate theory into the factor model at the level of the factor loadings. These restrictions mostly take the form of exclusion restrictions (i.e., setting factor loadings to 0) to specify a factor-pattern that emulates sparse structure and that indicates which item loads on which latent factor. Such restrictions imply a loss of information and may, e.g., induce bias in the estimates of the free parameters. Importantly, researchers often have informed ideas regarding direction and magnitude of parameter effects that cannot be expressed using exclusion restrictions. Usage of inequalities directly allows one to express the direction and (relative) strength of factor loadings and as such would bring more specificity to competing models of factor structure. The desire is then to construct an approach to Bayesian model selection that allows one to express competing factor analytic structures using inequality constraints rather than exclusion restrictions.
6. This slide contain some geometric intuition on how Bayesian inequality-constrained-model selection would work for the factor analytic model. The schematic considers a unit circle, representing the

two-dimensional prior space on 2 factor loadings bound by the communality of the standardized FA solution. The concentric circles represent the (location of the) posterior probability mass. Two competing inequality-constrained models are then considered. Model 1 represents the feasible space (in grey) defined by $\lambda_{jk} > |\lambda_{jk'}|$. Model 2 represents the feasible space (in blue) defined by $\lambda_{jk'} > |\lambda_{jk}|$. Both models are of the same complexity as for both models the feasible space comprises 1/4 of the total parameter space. The posterior mass, however, is mostly located in the feasible region defined by Model 1. Assume for simplicity that 7/8 of the posterior mass is located in the feasible region of Model 1 and that 1/8 of the posterior mass is located in the feasible region of Model 2. The Bayes factor (the main Bayesian model selection criterion) for an inequality-constrained model to the unconstrained model boils down to the ratio of the posterior probability mass over the prior probability mass satisfying the constraints. Hence, the Bayes factor of Model 1 to the unconstrained model, $B_{10} = (7/8) \div (1/4) = 7/2$. The Bayes factor of Model 2 to the unconstrained model, $B_{20} = (1/8) \div (1/4) = 1/2$. From these Bayes factors we may find the Bayes factor of model 1 to Model 2 as: $B_{12} = (7/2) \div (1/2) = 7$. This conveys that there is positive evidence of Model 1 against Model 2, or, loosely speaking, Model 1 is seven times as likely for the data at hand than Model 2.

7. The provision from the previous slide can be used to select, from a batch of competing inequality-constrained factor structures, the one structure most supported by the data in terms of a balancing of model fit and model complexity. The strategy is to formulate a base-model that imposes only minimal exclusion restrictions in order to achieve (rotational) identification of the factor model. These restrictions are chosen such that they convey preconceived theoretical meaning. Substantive factor analytic theory is then not represented by imposing additional exclusions restrictions, but by imposing (competing) systems of inequality constraints on and between the free parameters in the loadings matrix. This slide then gives a simple example of the possibilities. Again a situation is considered with 6 items and 2 latent factors. Two competing inequality-constrained structures are given. It should be clear that the ability to directly compare these models against each other allows for the testing of factor structure with higher specificity.
8. Back to the cake. The base layer is formed by the development of Bayes factors and accompanying computation strategies for constrained-model selection. These strategies are subsequently utilized in the second layer: The reformulation of EFA and CFA in the Bayesian framework from a constrained-model perspective. The top layer then uses these developments to propose an alternative factor analytic strategy that aims to merge EFA and CFA. This strategy is used in the analysis of real data from various fields of interdisciplinary (psychological) inquiry.
9. The strategy is, e.g., used in the reanalysis of data regarding anthropometric measurements on overweight and obese children. These measurements consider phenotypic expressions of the metabolic syndrome (MBS). MBS refers to a clustering of risk factors of metabolic origin and is thought to be a precursor for the development of diabetes and cardiovascular disease. The pathophysiologic constellation of the MBS has become an active area of research. The reanalysis using the alternative strategy suggests two pathophysiologic constellations indicating abnormalities in the glucose and lipid metabolism, respectively. These findings may give inroads for behavioral and educational approaches towards MBS prevention.
10. Another application considers handedness in captive Chimpanzees. The data considered various tasks that were designed to elicit dominant hand preference. It was found that the clustering of tasks could be explained by factors representing the complexity of motor demands on hand musculature.
11. Acknowledgements: Supervisors.
12. Acknowledgements: People who provided data for the applications in the dissertation.
13. Acknowledgements: Wife and collaborator (not necessarily in that order).