UNIVERSITY OF TWENTE.

SEM Meeting,
Amsterdam

Confirmatory Composite Analysis

Florian Schuberth[1] Jörg Henseler[1]
Theo K. Dijkstra[2]

[1]University of Twente

[2]Unversity of Groningen

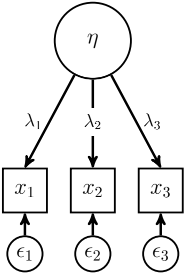
March 16, 2018

Overview

Latent variables

Type of theoretical construct

Criterion:	Latent variable
Dominant statistical model:	Common factor model



Fundamental scientific question:	Does the latent variable exist?
Scientific paradigm:	Positivism
Examples:	Abilities, attitudes, traits

Artifacts

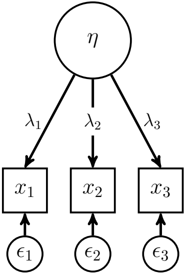
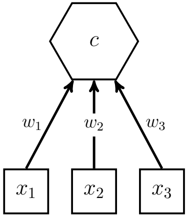
Many disciplines deal with an interplay of behavioral (latent variable) and design constructs (artifacts) such as

Discipline	Latent variable	Artifact
Marketing:	Consumer brand attitude	Advertising mix
Criminology:	Intention to commit a crime	Prevention strategy
Education:	Pupil's knowledge base	Teaching program
Psychotherapy:	Mental illness	Psychiatric treatment

$\rightarrow$ How to model these artifacts?

Two kinds of constructs

Type of theoretical construct

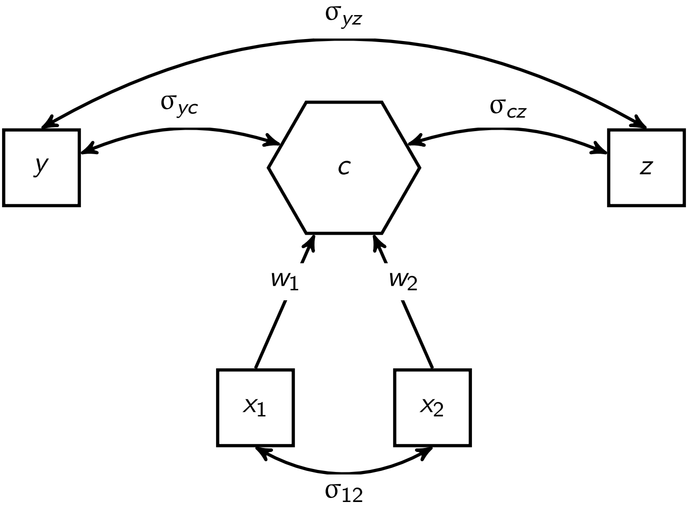
Criterion:	Latent variable	Artifact
Dominant statistical model:	Common factor model	Composite model
		
Fundamental scientific question:	Does the latent variable exist?	Is the artifact useful?
Scientific paradigm:	Positivism	Pragmatism
Examples:	Abilities, attitudes, traits	Indices, therapies, intervention programs

Confirmatory Composite Analysis

The confirmatory composite analysis (CCA) consists of 4 steps:

1. Specification of the composite model
2. Identification of the composite model
3. Estimation of the composite model
4. Assessment of the composite model

Specification of the composite model



Minimal composite model

Is this a statistical model?

Consider the model-implied indicator population covariance matrix:

$$\mathbf{\Sigma} = \begin{pmatrix} \underline{y} & \underline{x_1} & \underline{x_2} & \underline{z} \\ \sigma_{yy} & & & \\ \lambda_1 \sigma_{yc} & \sigma_{11} & & \\ \lambda_2 \sigma_{yc} & \sigma_{12} & \sigma_{22} & \\ \sigma_{yz} & \lambda_1 \sigma_{cz} & \lambda_2 \sigma_{cz} & \sigma_{zz} \end{pmatrix},$$

where $\lambda_1 = \mathrm{cov}(x_1, c)$ and $\lambda_2 = \mathrm{cov}(x_2, c)$.
This matrix has rank-one constraints, which can be exploited in statistical testing.
$\rightarrow$ Indeed, it is a statistical model

Identification of the composite model

Identification of composite models is straightforward:[1]

- Normalization of the weights, e.g., $\boldsymbol{w}_j'\boldsymbol{\Sigma}_{jj}\boldsymbol{w}_j = 1$
- Each composite must be connected to at least one composite or variable not forming the composite

$\rightarrow$ All model parameters can be uniquely retrieved from the population indicator covariance matrix

[1] We ignore trivial regularity assumptions such as weight vectors consisting of zeros only; and similarly, we ignore cases where intra-block covariance matrices are singular.

Estimation of the composite model

For determining the weights, several methods have been proposed:

- Sum scores
- Expert weighting
- Approaches to generalized canonical correlation analysis (GCCA) such as MAXVAR [Kettenring, 1971]
- Regularized general canonical correlation analysis (RGCCA) [Tenenhaus & Tenenhaus, 2011]
- Partial least squares path modeling (PLS-PM) [Wold, 1975]
- Generalized structured component analysis (GSCA) [Hwang & Takane, 2004]

Assessment of the composite model

The overall model fit can be assessed in two non-exclusive ways:

- Measures of fit (heuristic rules)
- Test for overall model fit

Assessment of the composite model

To test the overall model fit, a bootstrap-based test can be used ($H_0 : \boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\theta)$) [Beran & Srivastava, 1985, Bollen & Stine, 1992] in combination with various discrepancy measures such as

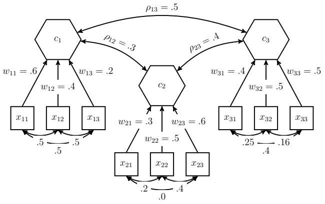
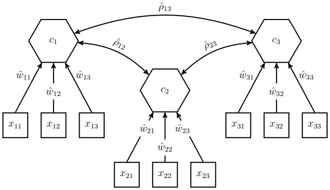
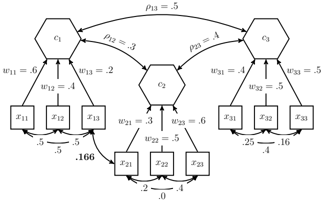
- Standardized root mean squared residual (SRMR)
- Geodesic distance (d_G)
- Euclidean distance (d_L)

Is the test for overall model fit capable to detect misspecifications in the composite model such as

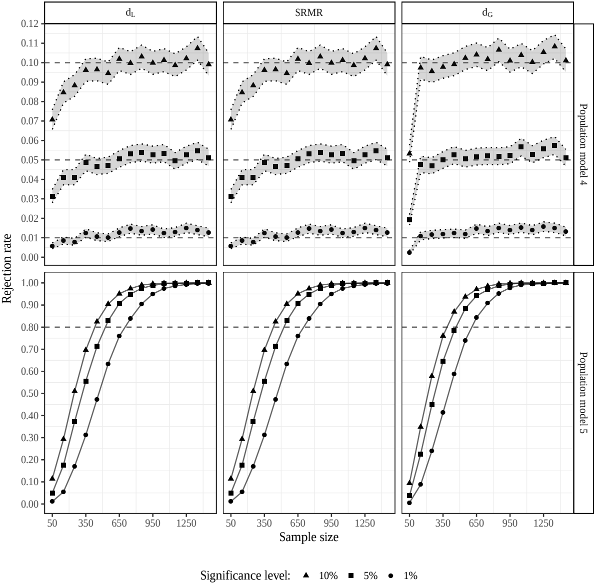
- Wrongly assigned indicators
- Correlations between indicators of different blocks that cannot be fully explained by the composites

$\rightarrow$ Monte Carlo simulation, where we use MAXVAR to determine the weights

Monte Carlo simulation

Experimental condition	Population model	Specified model
4) No misspecification		
5) Unexplained correlation		

Rejection rates



Thank you!

Florian Schuberth
email: f.schuberth@utwente.nl
UNIVERSITY OF TWENTE.

References

Beran, R. & Srivastava, M.S. (1985)
Bootstrap tests and confidence regions for functions of a covariance matrix
The Annals of Statistics 13(1) 95 – 115.

Bollen, K. A. & Stine, R. A. (1992)
Bootstrapping goodness-of-fit measures in structural equation models
Sociological Methods & Research 21(2) 205 – 229.

Hwang, H. & Takane, Y. (2004)
Generalized structured component analysis
Psychometrika 69(1) 81 – 99.

References

Kettenring, J.R. (1971)
Canonical analysis of several sets of variables
Biometrika, 58(3), 433 – 451.

Pearson, K. (1901)
On lines and planes of closest fit to systems of points in space
Philosophical Magazine Series 6 2(11) 559 – 572.

Tenenhaus, A. & Tenenhaus, M. (2011)
Regularized generalized canonical correlation analysis
Psychometrika 76(2) 257 – 284.

References

Wold, A.O.H. (1975)
Path models with latent variables: The NIPALS approach. In H. Blalock, A. Aganbegian, F. Borodkin, R. Boudon, & V. Capecchi (Eds.)
Quantitative Sociology, 307 - 357, New York Academic Press.