UNIVERSITY OF TWENTE.

Confirmatory Composite Analysis

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

[1]University of Twente

[2]Unversity of Groningen

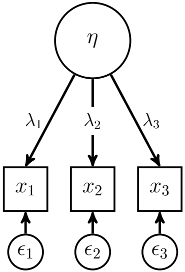
November 1, 2018

Overview

1. Motivation
2. Confirmatory composite analysis
   - Model Specification
   - Model Identification
   - Model Estimation
   - Model Assessment
3. Monte Carlo simulation
4. Extensions

Latent Variable

Type of construct:	Latent variable
Dominant statistical model:	Common factor model



Fundamental scientific question:	Does the latent variable exist?
Scientific paradigm:	Positivism
Statistical approach:	Confirmatory factor analysis
Examples:	Abilities, attitudes, traits

Artifacts

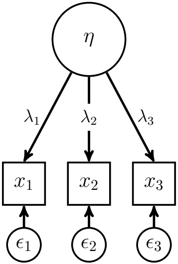
Many disciplines deal with design constructs (artifacts) and their interplay with behavioral constructs (latent variables)

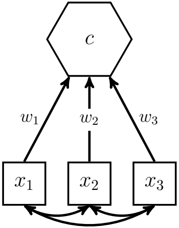
Discipline	Latent variable	Artifact
Marketing:	Consumer brand attitude	Advertising mix
Criminology:	Intention to commit a crime	Prevention strategy
Education:	Pupil's knowledge base	Teaching program
Psychotherapy:	Mental illness	Psychiatric treatment

$\rightarrow$ How to model and assess these artifacts?

Latent Variables & Artifacts

Type of construct:	Latent variable	Artifact
Dominant statistical model:	Common factor model	Composite model





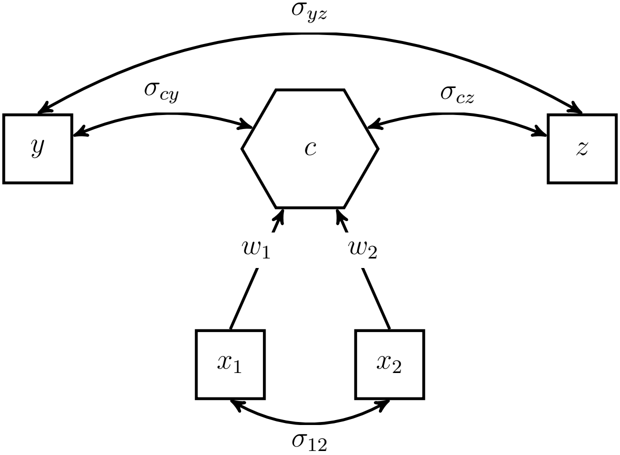
Fundamental scientific question:	Does the latent variable exist?	Is the artifact useful?
Scientific paradigm:	Positivism	Pragmatism
Statistical approach:	Confirmatory factor analysis	**Confirmatory composite analysis**
Examples:	Abilities, attitudes, traits	Indices, therapies, intervention programs

Confirmatory Composite Analysis

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?

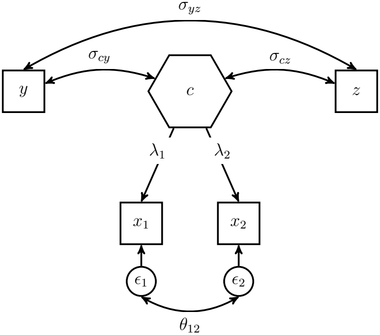
Consider the model-implied indicators' population covariance matrix:

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

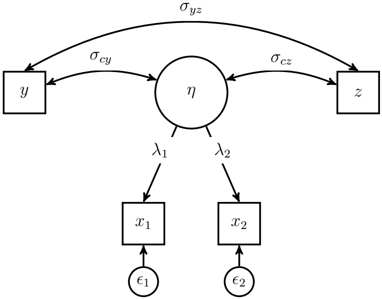
where $\lambda_1 = \text{cov}(x_1, c)$ and $\lambda_2 = \text{cov}(x_2, c)$.

This matrix has rank-one constraints, which can be exploited in statistical testing.
$\rightarrow$ Indeed, it is a statistical model

Composite Model vs. Common Factor Model



(a) Composite factor model



(b) Common factor model

Composite Model vs. Common Factor Model

Model-implied indicators' covariance matrix of the...

...composite factor model:

$$\begin{array}{cc} & \begin{array}{cccc} \underline{x_1} & \underline{x_2} & \underline{y} & \underline{z} \end{array} \\ \boldsymbol{\Sigma} = & \begin{pmatrix} \lambda_1^2+\theta_1 & & & \\ \lambda_1\lambda_2+\theta_{12} & \lambda_2^2+\theta_2 & & \\ \lambda_1\sigma_{cy} & \lambda_2\sigma_{cy} & \sigma_{yy} & \\ \lambda_1\sigma_{cz} & \lambda_2\sigma_{cz} & \sigma_{yz} & \sigma_{zz} \end{pmatrix} \end{array}$$

...common factor model:

$$\begin{array}{cc} & \begin{array}{cccc} \underline{x_1} & \underline{x_2} & \underline{y} & \underline{z} \end{array} \\ \boldsymbol{\Sigma} = & \begin{pmatrix} \lambda_1^2+\theta_1 & & & \\ \lambda_1\lambda_2 & \lambda_2^2+\theta_2 & & \\ \lambda_1\sigma_{cy} & \lambda_2\sigma_{cy} & \sigma_{yy} & \\ \lambda_1\sigma_{cz} & \lambda_2\sigma_{cz} & \sigma_{yz} & \sigma_{zz} \end{pmatrix} \end{array}$$

$\Rightarrow$ The common factor model is nested in the composite model [Henseler et al. 2014]

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

In case of composites embedded in a structural model, also the structural model needs to be identified [Dijkstra, 2017]

[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.

Model Identification: Degrees of Freedom

For the composite model the degrees of freedom are calculated as follows:

$$
\begin{aligned}
\text{df} = \ & \# \text{ non-redundant off-diagonal elements of the indicator covariance matrix} \\
- \ & \# \text{ free correlations among the composites} \\
- \ & \# \text{ free covariances between the composites and indicators not forming a composite} \\
- \ & \# \text{ covariances among the indicators not forming a composite} \\
- \ & \# \text{ free non-redundant off-diagonal elements of each intra-block covariance matrix} \\
- \ & \# \text{ weights} \\
+ \ & \# \text{ blocks}
\end{aligned}
$$

For our minimal composite example:

$$\text{df} = 6 - 0 - 2 - 1 - 1 - 2 + 1 = 1$$

Estimation of the Composite Model

To determine the weights, several methods have been proposed:

- Predetermined weights such as unit weights or weights obtained by experts
- 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]

GCCA: MAXVAR

MAXVAR maximizes the largest eigenvalue of the composite correlation matrix to obtain the weights
$\Rightarrow$ The total variation of the composites is explained as well as possible by one underlying "principal component"

Advantage over other approaches to GCCA that it has a closed form expression

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

Fit Measures

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

- Standardized root mean squared residual (SRMR)
- Root mean squared residual covariance matrix (RMS_{Θ})
- Normed fit index (NFI)
- ...

More research is required to assess their performance in case of composite models

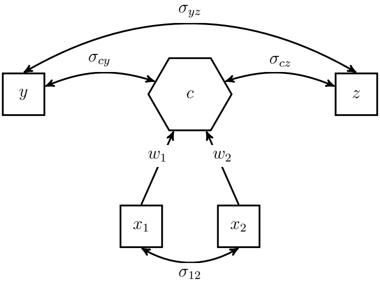
Test for Overall Model Fit

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

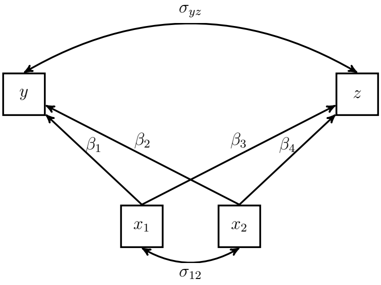
- SRMR
- Geodesic distance
- Squared Euclidean distance

Test for Overall Model Fit

It compares the model-implied indicators' covariance matrix of the composite and a saturated model:



(c) Composite model



(d) Saturated model

If the test is not rejected empirical evidence for the usefulness of the artifact is obtained

Monte Carlo Simulation

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 to assess the performance

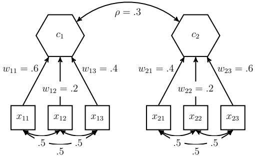
Simulation setup:

- 5 population models
- weights are calculated by MAXVAR
- 10,000 runs
- 200 bootstrap runs
- normally distributed datasets
- various sample sizes from 50 to 1,450

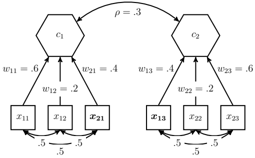
Monte Carlo Simulation: Population Models

Experimental condition	Population model	Estimated model

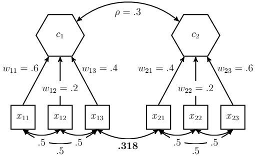
1) No misspecification



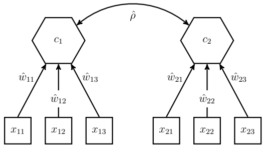
2) Confounded indicators



3) Unexplained correlation



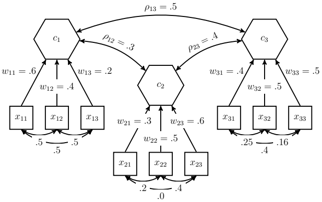
Estimated model:



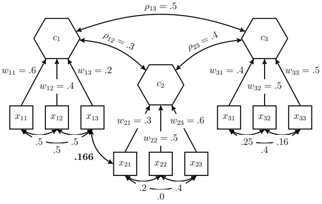
Monte Carlo Simulation: Population Models

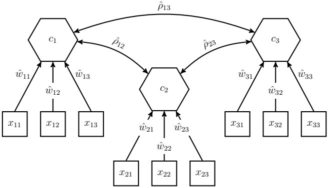
Experimental condition	Population model	Specified model

4) No misspecification

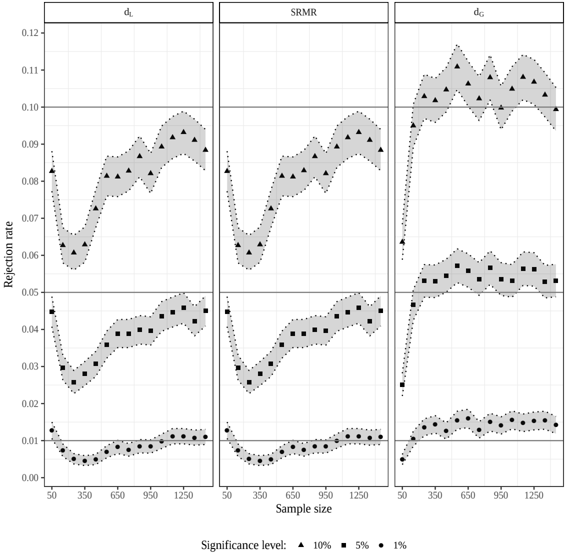


5) Unexplained correlation

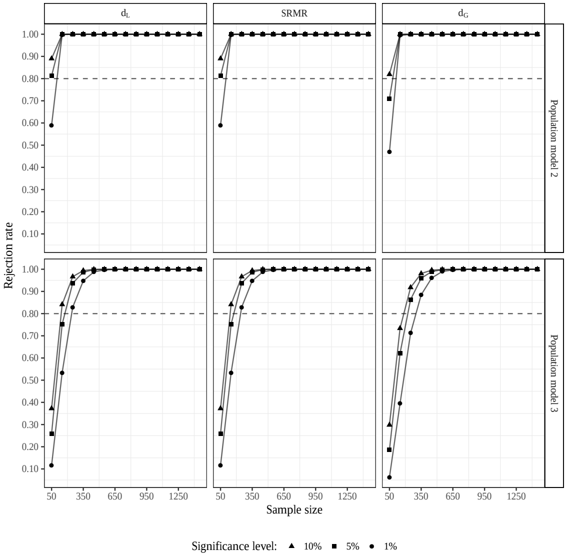




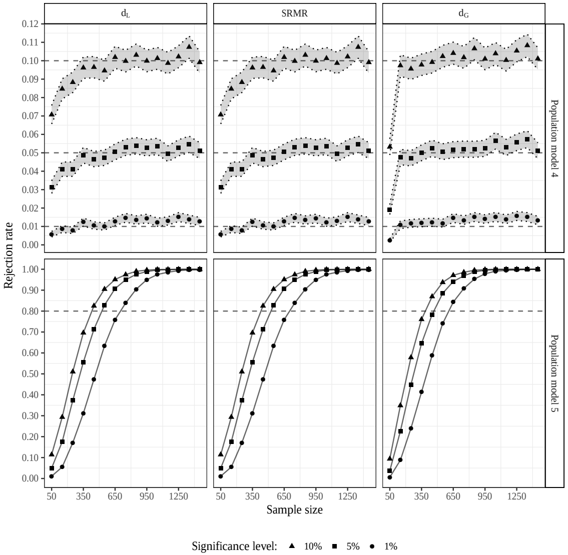
Monte Carlo Simulation: Rejection Rates



Monte Carlo Simulation: Rejection Rates

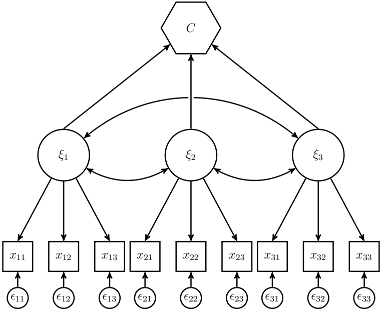


Monte Carlo Simulation: Rejection Rates

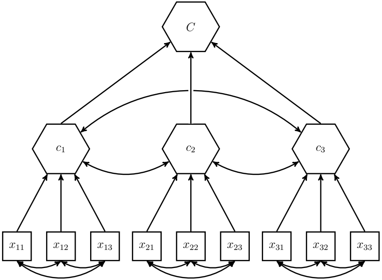


Extension: Second-order Composites

Artifacts that are built of other constructs can be modeled and tested such as an artifact built of latent variables [Van Riel et al., 2017] or artifacts [Schuberth & Henseler, 2018]



(e) Composite of common factors



(f) Composite of composites

Extension: Multigroup Comparison

It can be assessed whether artifacts are built the same across groups (MICOM) [Henseler et al. 2016].

It can be assessed whether the built artifact's behavior is the same across groups, i.e., comparing the model-implied indicator variance-covariance matrix across groups using a permutation test [Klesel et al., in press].

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.

Dijkstra, T. K. (2017)
The perfect match between a model and a mode. In H. Latan & R. Noonan (Eds.)
Partial Least Squares Path Modeling: Basic Concepts, Methodological Issues and Applications, 55 – 80, Springer.

References

Henseler, J., Dijkstra, T. K., Sarstedt, M., Ringle, C. M., Diamantopoulos, A., Straub, D. W., Ketchen, D. J., Hair, J. F., Hult, G. T. M. & Calantone, R. J. (2014)
Common beliefs and reality about PLS: Comments on Rönkkö and Evermann (2013)
Organizational Research Methods 17 182 – 209.

Henseler, J., Ringle, C., & Sarstedt, M. (2016)
Testing measurement invariance of composites using partial least squares
International Marketing Review 33(3) 405 – 431.

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

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

References

Klesel, M., Schuberth, F., Henseler, J., & Niehaves, J. (in press)
A test for multigroup comparison in partial least squares path modeling
Internet Research, 1 – 17.

Schuberth, F. & Henseler, J. (2018)
Dealing with hierarchical models containing composites of composites using PLS path modeling
Working Paper.

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

Van Riel, A. C. R., Henseler, J., Kemény, I., & Sasovova, Z (2017)
Estimating hierarchical constructs using Partial Least Squares: the case of second order composites of factors
Industrial Management & Data Systems, 117(3) 459 – 477.

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.