Confirmatory Composite Analysis

Florian Schuberth\textsuperscript{1} \quad Jörg Henseler\textsuperscript{1}
Theo K. Dijkstra\textsuperscript{2}
\textsuperscript{1}University of Twente
\textsuperscript{2}University of Groningen
June 18, 2019
Overview

1 Motivation

2 Confirmatory composite analysis
   • Model Specification
   • Model Identification
   • Model Estimation
   • Model Assessment

3 Monte Carlo simulation

4 Extensions
# Behavioral concepts

<table>
<thead>
<tr>
<th>Type of concept:</th>
<th>Behavioral concept</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type of construct:</td>
<td>Latent variable</td>
</tr>
<tr>
<td>Dominant statistical model:</td>
<td>Common factor model</td>
</tr>
</tbody>
</table>

![Diagram of common factor model]

**Scientific paradigm:** Scientific Realism  
**Statistical approach:** Confirmatory factor analysis  
**Examples:** Abilities, attitudes, traits
Artifacts

Many disciplines deal with design concepts, so-called artifacts and their interplay with behavioral concepts:

<table>
<thead>
<tr>
<th>Discipline</th>
<th>Behavioral concept</th>
<th>Artifact</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marketing:</td>
<td>Consumer brand attitude</td>
<td>Advertising mix</td>
</tr>
<tr>
<td>Criminology:</td>
<td>Intention to commit a crime</td>
<td>Prevention strategy</td>
</tr>
<tr>
<td>Education:</td>
<td>Pupil's knowledge base</td>
<td>Teaching program</td>
</tr>
<tr>
<td>Psychotherapy:</td>
<td>Mental illness</td>
<td>Psychiatric treatment</td>
</tr>
</tbody>
</table>

→ How to model and assess these artifacts?
## Behavioral concepts & Artifacts

<table>
<thead>
<tr>
<th>Type of concept:</th>
<th>Behavioral concept</th>
<th>Artifact</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type of construct:</td>
<td>Latent variable</td>
<td>Emergent variable</td>
</tr>
<tr>
<td>Dominant statistical model:</td>
<td>Common factor model</td>
<td>Composite model</td>
</tr>
</tbody>
</table>

- Scientific paradigm: Scientific Realism
- Statistical approach: Confirmatory factor analysis
- Confirmatory composite analysis
- Examples: Abilities, attitudes, traits
- Indices, therapies, intervention programs

**Diagram 1:**
- $\eta$
- $\lambda_1, \lambda_2, \lambda_3$
- $x_1, x_2, x_3$
- $\epsilon_1, \epsilon_2, \epsilon_3$

**Diagram 2:**
- $c$
- $w_1, w_2, w_3$
- $x_1, x_2, x_3$
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:

\[
\Sigma = \begin{pmatrix}
  \sigma_{11} & \sigma_{12} & \lambda_1 \sigma_{cy} & \lambda_1 \sigma_{cz} \\
  \sigma_{12} & \sigma_{22} & \lambda_2 \sigma_{cy} & \lambda_2 \sigma_{cz} \\
  \lambda_1 \sigma_{cy} & \lambda_2 \sigma_{cy} & \sigma_{yy} & \sigma_{yz} \\
  \lambda_1 \sigma_{cz} & \lambda_2 \sigma_{cz} & \sigma_{yz} & \sigma_{zz}
\end{pmatrix},
\]

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.

→ 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:  \[ \Sigma = \begin{pmatrix}
  x_1 & x_2 & y & z \\
  \lambda_1^2 + \theta_1 & \lambda_1 \lambda_2 + \theta_{12} & \lambda_2^2 + \theta_2 & \\
  \lambda_1 \sigma_{cy} & \lambda_1 \sigma_{cz} & \lambda_2 \sigma_{cy} + \sigma_{yy} & \lambda_2 \sigma_{cz} + \sigma_{yz} & \sigma_{zz}
\end{pmatrix} \]

...common factor model:

\[ \Sigma = \begin{pmatrix}
  x_1 & x_2 & y & z \\
  \lambda_1^2 + \theta_1 & \lambda_1 \lambda_2 & \lambda_2^2 + \theta_2 & \\
  \lambda_1 \sigma_{cy} & \lambda_1 \sigma_{cz} & \lambda_2 \sigma_{cy} & \lambda_2 \sigma_{cz} & \sigma_{yy}
\end{pmatrix} \]

\[ \Rightarrow \] 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., \(w_j^\prime \Sigma_{jj} w_j = 1\)
- 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:

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

For our minimal composite example:

\[
df = 6 - 0 - 2 - 1 - 1 - 2 + 1 = 1
\]
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]
- 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

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 ($\text{RMS}_\Theta$)
▶ 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 ($H_0 : \Sigma = \Sigma(\theta)$) [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:

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

⇒ 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

<table>
<thead>
<tr>
<th>Experimental condition</th>
<th>Population model</th>
<th>Estimated model</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) No misspecification</td>
<td>$\rho = .3$</td>
<td>$\hat{\rho}$</td>
</tr>
<tr>
<td></td>
<td>$w_{11} = .6$</td>
<td>$\hat{w}_{11}$</td>
</tr>
<tr>
<td></td>
<td>$w_{12} = .2$</td>
<td>$\hat{w}_{12}$</td>
</tr>
<tr>
<td></td>
<td>$w_{13} = .4$</td>
<td>$\hat{w}_{13}$</td>
</tr>
<tr>
<td></td>
<td>$w_{21} = .4$</td>
<td>$\hat{w}_{21}$</td>
</tr>
<tr>
<td></td>
<td>$w_{22} = .2$</td>
<td>$\hat{w}_{22}$</td>
</tr>
<tr>
<td></td>
<td>$w_{23} = .6$</td>
<td>$\hat{w}_{23}$</td>
</tr>
<tr>
<td></td>
<td>$\rho = .3$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$w_{11} = .6$</td>
<td>$\hat{w}_{11}$</td>
</tr>
<tr>
<td></td>
<td>$w_{12} = .2$</td>
<td>$\hat{w}_{12}$</td>
</tr>
<tr>
<td></td>
<td>$w_{13} = .4$</td>
<td>$\hat{w}_{13}$</td>
</tr>
<tr>
<td></td>
<td>$w_{12} = .2$</td>
<td>$\hat{w}_{12}$</td>
</tr>
<tr>
<td></td>
<td>$w_{13} = .4$</td>
<td>$\hat{w}_{13}$</td>
</tr>
<tr>
<td></td>
<td>$w_{21} = .4$</td>
<td>$\hat{w}_{21}$</td>
</tr>
<tr>
<td></td>
<td>$w_{22} = .2$</td>
<td>$\hat{w}_{22}$</td>
</tr>
<tr>
<td></td>
<td>$w_{23} = .6$</td>
<td>$\hat{w}_{23}$</td>
</tr>
</tbody>
</table>

2) Confounded indicators

3) Unexplained correlation
Monte Carlo Simulation: Population Models

<table>
<thead>
<tr>
<th>Experimental condition</th>
<th>Population model</th>
<th>Specified model</th>
</tr>
</thead>
<tbody>
<tr>
<td>4) No misspecification</td>
<td>![Diagram]</td>
<td>![Diagram]</td>
</tr>
<tr>
<td>5) Unexplained correlation</td>
<td>![Diagram]</td>
<td>![Diagram]</td>
</tr>
</tbody>
</table>
Monte Carlo Simulation: Rejection Rates

![Graph showing rejection rates for different significance levels and sample sizes. The graphs compare values for $d_L$, SRMR, and $d_G$, with rejection rates plotted against sample size. The significance levels are indicated by markers: ▲ for 10%, ▼ for 5%, and ● for 1%.]
Monte Carlo Simulation: Rejection Rates

![Graph showing rejection rates for different sample sizes and significance levels for population models 2 and 3.](image-url)
Monte Carlo Simulation: Rejection Rates

Graphs showing the rejection rates for different sample sizes and significance levels. The graphs compare population models 4 and 5. The rejection rates are depicted for different statistical measures such as $d_L$, SRMR, and $d_0$. The significance levels are marked as 10%, 5%, and 1%.
Emergent variables that are built of other constructs can be modeled and tested such as an emergent variable built of latent variables [Van Riel et al., 2017] or artifacts [Schuberth et al., in progress].
Extension: Multigroup Comparison

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

It can be assessed whether the built emergent variable’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

Bootstrapping goodness-of-fit measures in structural equation models
*Sociological Methods & Research* 21(2) 205 – 229.

The perfect match between a model and a mode. In H. Latan & R. Noonan (Eds.)
References


References

A test for multigroup comparison in partial least squares path modeling  
*Internet Research*, 1 – 17.

Schuberth, F., Rademaker, M., & Henseler, J. (in progress)  
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.

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.)