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
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

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

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)

<table>
<thead>
<tr>
<th>Discipline</th>
<th>Latent variable</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?
# Latent Variables & Artifacts

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

**Diagram:**

- **Latent Variable Diagram:**
  - Latent variable $\eta$
  - Factor loadings $\lambda_1$, $\lambda_2$, $\lambda_3$
  - Observed variables $x_1$, $x_2$, $x_3$
  - Error terms $\epsilon_1$, $\epsilon_2$, $\epsilon_3$

- **Artifact Diagram:**
  - Artifact $c$
  - Indicator variables $x_1$, $x_2$, $x_3$
  - Error terms $w_1$, $w_2$, $w_3$

**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:

\[
\Sigma = \begin{pmatrix}
  \sigma_{11} & \sigma_{12} & \sigma_{1y} & \sigma_{1z} \\
  \sigma_{12} & \sigma_{22} & \lambda_1 \sigma_{cy} & \lambda_2 \sigma_{cy} \\
  \sigma_{1y} & \lambda_1 \sigma_{cy} & \sigma_{yy} & \sigma_{yz} \\
  \sigma_{1z} & \lambda_2 \sigma_{cy} & \sigma_{yz} & \sigma_{zz}
\end{pmatrix},
\]

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:  

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

...common factor model:  

\[
\Sigma = \begin{pmatrix}
\lambda_1^2 + \theta_1 & \lambda_1 \lambda_2 & \lambda_2^2 + \theta_2 \\
\lambda_1 \sigma_{cy} & \lambda_1 \sigma_{cz} & \lambda_2 \sigma_{cy} & \lambda_2 \sigma_{cz} & \sigma_{yy} & \sigma_{yz} & \sigma_{zz}
\end{pmatrix}
\]

⇒ 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:\(^1\)

1. Normalization of the weights, e.g., \( w_j' \Sigma_{jj} w_j = 1 \)
2. Each composite must be connected to at least one composite or variable not forming the composite

→ 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:

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

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
⇒ 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 ($\text{RMS}_\Theta$)
- 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><img src="image1" alt="Diagram" /></td>
<td><img src="image2" alt="Diagram" /></td>
</tr>
<tr>
<td>2) Confounded indicators</td>
<td><img src="image3" alt="Diagram" /></td>
<td><img src="image4" alt="Diagram" /></td>
</tr>
<tr>
<td>3) Unexplained correlation</td>
<td><img src="image5" alt="Diagram" /></td>
<td><img src="image6" alt="Diagram" /></td>
</tr>
</tbody>
</table>
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><img src="image1" alt="Diagram 1" /></td>
<td><img src="image2" alt="Diagram 2" /></td>
</tr>
<tr>
<td>5) Unexplained correlation</td>
<td><img src="image3" alt="Diagram 3" /></td>
<td><img src="image4" alt="Diagram 4" /></td>
</tr>
</tbody>
</table>
Monte Carlo Simulation: Rejection Rates

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

Sample size vs. Rejection rate for different population models and significance levels.

Population model 2:
- d1
- SRMR
- d2

Population model 3:
- d1
- SRMR
- d2

Significance levels: ▲ 10%, ■ 5%, ● 1%
Monte Carlo Simulation: Rejection Rates

Population model 4
Population model 5

Sample size
Rejection rate
Significance level: ▲ 10% ■ 5% ● 1%

Significance level: ▲ 10% ■ 5% ● 1%
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].
Confirmatory Composite Analysis

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

Common beliefs and reality about PLS: Comments on Rönkkö and Evermann (2013)
*Organizational Research Methods* 17 182 – 209.

Testing measurement invariance of composites using partial least squares

Generalized structured component analysis
*Psychometrika* 69(1) 81 – 99.

Kettenring, J.R. (1971)
Canonical analysis of several sets of variables
A test for multigroup comparison in partial least squares path modeling  
*Internet Research*, 1 – 17.

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.)