



universität
wien

Diplomarbeit

Titel der Diplomarbeit

Simulationsstudie zur konfirmatorischen Faktorenanalyse: Evaluation der χ^2 -Teststatistik und anderer Fit-Indizes bei fehlspezifizierter Faktorenanzahl

Verfasser

Michael Themessl-Huber

Angestrebter akademischer Grad

Magister der Naturwissenschaften (Mag. rer. nat.)

Wien, 2014

Studienkennzahl: 298

Studienrichtung: Diplomstudium Psychologie

Betreuer: Univ.-Prof. i.R. Mag. Dr. Klaus D. Kubinger

Danksagung

Ich möchte an dieser Stelle all jenen danken, die durch ihre Unterstützung zum Gelingen dieser Diplomarbeit beigetragen haben.

Besonders bedanke ich mich bei Univ.-Prof. Mag. Dr. Klaus Kubinger für die Betreuung der Diplomarbeit, der Bereitstellung dieses interessanten Themas sowie die unkomplizierte und schnelle Hilfe bei Unklarheiten.

Mein Dank gilt auch Mag. Jan Steinfeld, der mir bei technischen Fragen und Problemen immer sofort weitergeholfen hat.

Ebenso möchte ich meiner Familie danken, insbesondere meinen Eltern, die mir dieses Studium ermöglicht haben und immer hinter mir gestanden sind.

Großer Dank gilt Katharina Schweigkofler für das gewissenhafte Korrekturlesen dieser Arbeit.

Schließlich möchte ich mich auch bei meinen Studienkollegen und Freunden bedanken, die mich während meiner Studienzeit begleitet haben.

Abstract – deutsche Version

Modellevaluation ist einer der wichtigsten Bestandteile der konfirmatorischen Faktorenanalyse. Dazu gibt es verschiedene Kriterien. Das klassische Kriterium ist die χ^2 -Statistik. Daneben haben sich verschiedene Fit-Indizes etabliert.

Da anhand der Fit-Indizes keine Signifikanztestung möglich ist, benötigt man Cut-Off-Werte, um eine Entscheidung bezüglich Modellannahme zu treffen. In einer umfangreichen Simulationsstudie ermittelten Hu und Bentler (1999) Cut-Offs für die gängigsten Fit-Indizes. Diese werden in der Literatur häufig empfohlen und kamen auch in vorliegender Simulationsstudie zum Einsatz.

Es wurde das tatsächlich eingegangene Risiko 1. Art und das Risiko 2. Art der χ^2 -Statistik bei fehlspezifizierter Faktorenstruktur untersucht. Dies geschah, indem ein richtig und zwei falsch spezifizierte Modelle getestet wurden. Ebenso wurde das Verhalten des *standardized root-mean-square residual (SRMR)*, des *root-mean-square error of approximation (RMSEA)* und des *comparative fit index (CFI)* untersucht. Es gab verschiedene Simulationsszenarien.

Um konkurrierende Modelle miteinander zu vergleichen und das besser passende auszuwählen, kann man das Akaike Informationskriterium (AIC), den χ^2 -Differenztest oder die F-Statistik heranziehen. Das Verhalten dieser Methoden wurde in der Simulation ebenfalls untersucht.

Es stellte sich heraus, dass der χ^2 -Test das nominelle Risiko 1. Art ($\alpha = 0,05$) vor allem bei der zugrunde gelegenen Stichprobe ($n = 150$) nicht einhält. Es liegt hier in allen Bedingungen bei über 20 %. Ist die Stichprobe größer ($n = 600$), nähert sich das Risiko 1. Art dem nominellen Risiko von 5%. Für den SRMR und RMSEA scheinen die Cut-Off-Werte bei vorliegender Fehlspezifikation nicht geeignet. Falsch spezifizierte Modelle werden häufig beibehalten. Für den CFI scheinen die Cut-Off-Werte passend. Er erzielte in nahezu allen Bedingungen bessere Ergebnisse als die anderen beiden Indizes.

Beim Modellvergleich gibt es keinen Unterschied zwischen dem χ^2 -Differenztest und dem F-Test. Der F-Test besitzt den Vorteil, dass er auch eingesetzt werden kann, wenn die Anzahl der Freiheitsgrade in den zu vergleichenden Modellen identisch ist.

Schlüsselwörter: Konfirmatorische Faktorenanalyse, Fit-Indizes, Fehlspezifizierung, Modellpassung, Faktorenstruktur, Modelltest, Strukturgleichungsmodell

Abstract – English version

Model evaluation is one of the most important parts in confirmatory factor analysis. There exist different criteria to evaluate a model. Beside the classical χ^2 -statistic, many fit indices have been developed.

Fit indices only allow a descriptive evaluation of the model fit - therefore cut-off values are needed, to decide, whether a model fits the data or not. In a simulation study, Hu and Bentler (1999) provided cut-off values for different fit indices. These cut-offs are often recommended by literature and were also used in this study.

Using data simulation the type I risk and the type II risk of the χ^2 -statistic were investigated under misspecified factor structure. Therefore, a correct specified model and two false specified models were tested under different simulation conditions. The behavior of the *standardized root-mean-square residual (SRMR)*, the *root-mean-square error of approximation (RMSEA)* and the *comparative fit index (CFI)* were also investigated.

To compare different models, the akaike information criterion (AIK), the χ^2 -difference-test or the F-statistic can be used. The behavior of these methods was investigated too.

It was shown, that the χ^2 -test did not hold the type I risk of 5 % in the small sample size condition ($n = 150$). It was over 20 %. In the larger sample size condition ($n = 600$), the type I risk is lower, but still tops 5 %. For the SRMR and the RMSEA different cut-off values should be used under present misspecification. Many false specified models were not rejected. In nearly all conditions the CFI showed better properties than the other indices - the cut-off value seems adequate.

There is no difference between the χ^2 -difference-test and the F-test. The F-test has the advantage, that it can be also used, if the number of degrees of freedom is the same in the compared models.

Keywords: confirmatory factor analysis, fit indices, misspecification, model fit, factor structure, model test, structural equation modeling

Inhaltsverzeichnis

I.	Einleitung	1
II.	Theoretischer Teil	3
1.	Faktorenanalyse	3
2.	Explorative Faktorenanalyse (EFA).....	4
3.	Konfirmatorische Faktorenanalyse (KFA)	6
3.1.	Modellidentifikation.....	8
3.2.	Parameterschätzung	10
4.	Modellbeurteilung	12
4.1.	Globaler Modell Fit.....	12
4.2.	Globale Fit Indizes	13
4.3.	Cut-Off-Werte.....	15
4.4.	Modellvergleich	16
III.	Simulationsstudie	20
5.	Ziel der Simulationsstudie	20
6.	Methoden	20
6.1.	Populationsmodelle.....	20
6.2.	Simulationsbedingungen.....	23
6.3.	Aufgestellte Modelle.....	23
6.4.	Auswertungsmethoden	25
7.	Ergebnisse	26
7.1.	Risiko 1. Art: Das 3-Faktoren-Modell	27
7.2.	Risiko 2. Art: Das 2-Faktoren-Modell	29
7.3.	Risiko 2. Art: Das 4-Faktoren-Modell	31
7.4.	Modellvergleich	33
8.	Diskussion.....	35
9.	Zusammenfassung.....	37
IV.	Verzeichnisse	39

10.	Literaturverzeichnis	39
11.	Tabellenverzeichnis.....	42
12.	Abbildungsverzeichnis.....	42
13.	R-Code	43
13.1.	Verwendetes Paket	43
13.2.	Populationsmodelle	43
13.3.	Aufgestellte Modelle	45
13.4.	Funktionen	47
13.5.	Funktionsaufrufe	51
14.	Lebenslauf	60

I. Einleitung

Die Faktorenanalyse ist ein häufig eingesetztes Instrument in den Sozialwissenschaften, vor allem in der Psychologie. Es erlaubt, von direkt beobachtbaren Eigenschaften auf latente Konstrukte (Faktoren) zu schließen. So ist beispielsweise das Konstrukt Intelligenz nicht direkt beobachtbar, kann aber durch geeignete Indikatoren operationalisiert werden. Indikatoren für Intelligenz wären sprachliche Fähigkeiten, mathematische Fähigkeiten und künstlerische Begabung. Diese Indikatoren könnten anhand der Schulnoten in Deutsch, Mathematik und Bildnerisches Gestalten gemessen werden. Die Faktorenanalyse wird häufig bei Fragebogen- und Testkonstruktionen eingesetzt.

Ziel der Faktorenanalyse ist es, die Anzahl und Beschaffenheit von Faktoren zu bestimmen, die für die gemeinsame Variation von gemessenen Variablen verantwortlich sind. Ein Faktor ist also eine unbeobachtete Variable, die mehr als eine manifeste Variable beeinflusst.

Während sich die explorative Faktorenanalyse (EFA) mit dem Auffinden von verdeckten Strukturen und einer Dimensionsreduktion befasst, ist es Aufgabe der konfirmatorischen Faktorenanalyse (KFA), die gefundenen Strukturen zu überprüfen.

Um zu untersuchen, ob die Faktorenstruktur die Daten ausreichend gut beschreibt, gibt es verschiedene Kriterien. Häufig werden die χ^2 -Teststatistik und verschiedene Fit-Indizes als Entscheidungsgrundlage für die Modellannahme oder –verwerfung herangezogen. Diese Kriterien reagieren jedoch auf diverse Modellfehlspezifikationen unterschiedlich sensibel.

In vergangenen Simulationsstudien wurde unter anderem das Verhalten der χ^2 -Statistik und anderer Fit-Indizes bezüglich falscher Faktorladungen, fehlspezifizierter korrelierter Fehlervarianzen und falscher Faktorenkorrelationen unter verschiedenen Simulationsbedingungen untersucht (Beauducel & Wittmann, 2005; Heene, Hilbert, Draxler, Ziegler & Bühner, 2011; Heene, Hilbert, Freudenthaler & Bühner, 2012; Hu & Bentler, 1999).

In dieser Arbeit soll das Verhalten der χ^2 -Statistik und der am häufigsten verwendeten Fit-Indizes bei fehlspezifizierter Faktorenanzahl untersucht werden. Dabei werden Modelle getestet, die entweder zu klein (ein Faktor zu wenig) oder zu groß (ein Faktor zu viel) sind. So kann das Risiko 2. Art eingeschätzt werden, indem der Anteil angenommener Modelle betrachtet wird. Ebenso wird das Risiko 1. Art untersucht, indem richtig spezifizierte Modelle, das heißt richtige Anzahl an Faktoren, getestet

werden. Außerdem werden die falsch spezifizierten Modelle mit dem richtig spezifizierten Modell verglichen.

Im theoretischen Teil wird die Faktorenanalyse beschrieben und Unterschiede zwischen explorativer und konfirmatorischer Faktorenanalyse aufgezeigt. Ebenso wird die konfirmatorische Faktorenanalyse von Strukturgleichungsmodellen abgegrenzt. Es wird beschrieben, wie Parameter im Modell geschätzt werden und welche Kriterien zur Modellbeurteilung bzw. zum Modellvergleich es gibt.

Der praktische Teil der Arbeit bezieht sich auf die durchgeführte Simulationsstudie. Dabei wird auf die verschiedenen Simulationsbedingungen genau eingegangen und Ergebnisse werden näher erläutert und diskutiert. Im Anhang befindet sich der Programmcode für die Simulation.

II. Theoretischer Teil

In diesem Teil der Arbeit werden wichtige Grundlagen erläutert, die für das Verständnis der durchgeführten Simulationsstudie wichtig sind. Zunächst werden die Unterschiede zwischen explorativer und konfirmatorischer Faktorenanalyse aufgezeigt. Im Anschluss wird die Parameterschätzung erklärt sowie jene Voraussetzungen, die dafür erfüllt sein müssen. Es werden verschiedene Kriterien zur Modellbeurteilung und zum Modellvergleich vorgestellt, die in der Simulationsstudie zum Einsatz kamen.

1. Faktorenanalyse

Die Faktorenanalyse wurde 1904 vom Psychologen Charles Spearman entwickelt. Er verwendete das Verfahren zur Auswertung von Intelligenztests (Spearman, 1904, 1927). Die Faktorenanalyse ist ein Verfahren der multivariaten Statistik. Ziel ist es, mit Hilfe von Indikatoren auf latente Konstrukte (Faktoren) zu schließen. Es ist also ein dimensionsreduzierendes Verfahren, da die Anzahl der Faktoren geringer ist als die Anzahl manifester Variablen.

Eine Grundannahme der Faktorenanalyse besagt, dass jede beobachtete Variable als lineare Funktion mindestens eines gemeinsamen Faktors und eines Fehlers ausgedrückt werden kann:

$$y_j = \lambda_{j1}\eta_1 + \lambda_{j2}\eta_2 + \dots + \lambda_{jm}\eta_m + \varepsilon_j$$

In dieser Formel steht y_j für den j -ten von p Indikatoren, λ_{jm} repräsentiert die Faktorladung der j -ten Beobachtung auf den m -ten Faktor η , mit ε_j ist die Varianz gemeint, die sich nur auf den j -ten Indikator bezieht und unabhängig von allen Faktoren η und allen anderen Fehlervarianzen ε ist.

Die Varianz eines jeden Indikators wird also in eine gemeinsame Varianz und in eine unique Varianz zerlegt. Die gemeinsame Varianz ist die Varianz, die mit anderen Indikatoren geteilt wird und durch den gemeinsamen Faktor erklärt wird. Die unique Varianz ergibt sich aus der für den Indikator spezifischen Varianz und einer Messfehlervarianz. Man spricht vom „common factor model“ (Thurstone, 1947).

Generell wird zwischen explorativer (EFA) und konfirmatorischer (KFA) Faktorenanalyse unterschieden. Auf diese beiden Methoden wird nun näher eingegangen.

2. Explorative Faktorenanalyse (EFA)

Die EFA ist ein hypothesengenerierendes Verfahren. Es brauchen keine a-priori Annahmen über die Anzahl der Faktoren oder das Faktorladungsmuster gemacht werden. Dies ist einer der Hauptunterschiede zwischen EFA und KFA. Das Ziel der EFA ist es, die geringstmögliche Anzahl an Faktoren zu finden, die für die Zusammenhänge der manifesten Variablen verantwortlich sind.

Man benötigt also eine Methode, um die Faktoren zu extrahieren. Im Anschluss werden dann diejenigen ausgewählt, die am stärksten zur Erklärung der Zusammenhänge zwischen den manifesten Variablen beitragen. Wichtig ist auch, dass die Faktoren inhaltlich sinnvoll interpretiert werden können. Die Datenbasis ist eine Korrelationsmatrix der beobachteten Variablen.

Es gibt verschiedene Methoden zur Faktorenextraktion. Bei zumindest intervallskalierten Daten werden am häufigsten die Hauptachsen- oder die Maximum-Likelihood-Methode (ML-Methode) eingesetzt. Letztere wird aufgrund einiger wünschenswerter Eigenschaften ebenso oft bei der KFA eingesetzt und wird später (Kapitel Parameterschätzung) näher erläutert. Der Vorteil der Hauptachsenmethode gegenüber der ML-Methode ist, dass die Indikatoren nicht multivariat normalverteilt sein müssen, sie ist also frei von Verteilungsannahmen (Brown, 2006).

Nach der Faktorenextraktion gilt es, die passende Anzahl an Faktoren zu finden. Dies stellt eine wichtige Entscheidung dar. Sowohl die Auswahl zu vieler Faktoren, als auch die Auswahl zu weniger Faktoren, können die Gültigkeit des Modells gefährden, wenn auch die Auswahl zu vieler Faktoren bei der EFA weniger folgenschwer sein kann (Fabrigar, Wegener, MacCallum & Strahan, 1999). Es gibt mehrere Möglichkeiten, die passende Anzahl an Faktoren zu finden, diese führen aber teilweise zu unterschiedlichen Ergebnissen. Zwei sehr populäre Kriterien basieren auf dem Konzept der Eigenwerte und Eigenvektoren. Vereinfacht kann man sagen, dass Eigenwerte die Varianz in den Indikatoren repräsentieren, die durch die Faktoren erklärt wird (Brown, 2006).

Die Kaiser-Gutmann Regel besagt, dass so viele Faktoren ausgewählt werden sollen, wie es Eigenwerte größer als 1 in der Indikatoren-Korrelationsmatrix gibt. Ist nämlich ein Eigenwert kleiner 1, bedeutet das, dass dieser zusätzliche Faktor weniger Varianz erklären würde als ein einzelner Indikator. Aufgrund ihrer Einfachheit ist diese Regel häufig in Verwendung, hat aber auch ihre Schwächen. Sind beispielsweise zwei Eigenwerte mit 1,01 und 0,99 gegeben, stuft das Kaiser-Kriterium den ersten Faktor als wichtig ein, während der andere nicht mehr aufgenommen wird (Brown, 2006).

Eine andere Methode zur Bestimmung der optimalen Faktorenanzahl ist der Scree-Test. Dies stellt ein grafisches Verfahren dar und wurde von Cattell entwickelt (1966). Auf der x-Achse wird dabei die Anzahl der Faktoren aufgetragen und auf der y-Achse die Eigenwerte. Der abfallende Graph wird dann nach einer Knickstelle untersucht (siehe Abbildung 1). Genau an dieser Stelle befindet sich die passende Anzahl an Faktoren. Die Faktoren nach der Knickstelle sind unbedeutend.

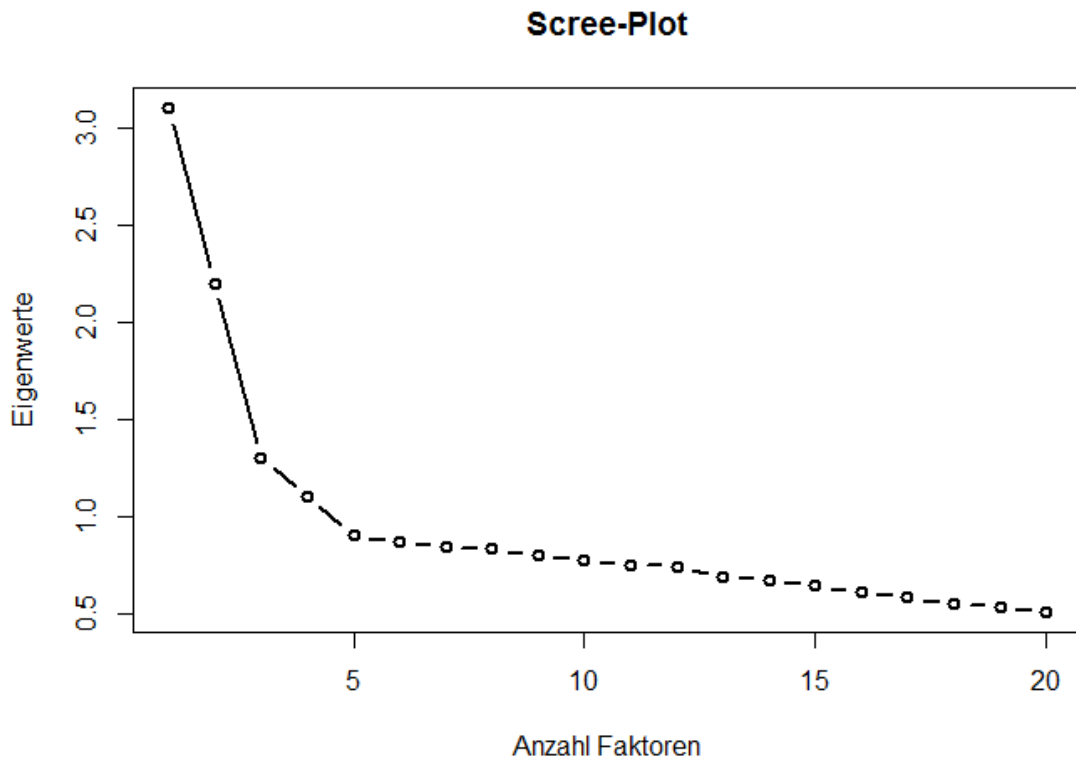


Abbildung 1: Beispiel eines Scree-Plots

Ein Nachteil des Scree-Tests ist seine geringe Objektivität. Besteht keine eindeutige Knickstelle, gibt es Interpretationsspielraum. In obiger Grafik könnte man von vier, aber auch von nur zwei Faktoren ausgehen, da sowohl an der Stelle des fünften Faktors als auch an der Stelle des dritten Faktors ein Knick zu erkennen ist.

Hat man nun die Anzahl der Faktoren bestimmt, geht es darum, sie inhaltlich sinnvoll zu interpretieren. Wurden zwei oder mehr Faktoren bestimmt, besteht die Möglichkeit, die Faktoren zu rotieren. Dies kann die inhaltliche Interpretation der latenten Konstrukte erleichtern.

Thurstone prägte den Begriff „simple structure“ (1947). Damit ist gemeint, dass auf jeden Faktor mehrere Variablen laden, aber jede Variable soll nur eine hohe Ladung auf einem einzigen Faktor haben und niedrige Ladungen auf allen anderen. Man spricht in diesem Zusammenhang auch von Primär- bzw. Sekundär- oder Nebenladungen. Rotation ermöglicht es, genau solche Ergebnisse zu erhalten, ohne

dass die Passung des Modells verändert wird. Es wird tatsächlich nur die Interpretation erleichtert, der Anteil an erklärter Varianz bleibt gleich.

Man unterscheidet orthogonale und schiefwinklige Rotation. Bei der orthogonalen Rotation wird der Winkel zwischen den Achsen des Koordinatensystems nicht verändert. Das heißt, die x- und y-Achse stehen weiterhin normal aufeinander. Das impliziert, dass die Faktoren nicht miteinander korrelieren. Bei der schiefwinkligen Rotation stehen die Achsen des Koordinatensystems nicht mehr orthogonal aufeinander. Somit kann der Winkel zwischen den Achsen Werte größer oder kleiner 90° annehmen. Die Höhe der Korrelation ergibt sich aus dem Cosinus des Winkels zwischen den beiden Achsen, z.B. $\cos(90)=0$. Die obige Beschreibung bezieht sich auf zwei Faktoren, das Prinzip ist aber auf mehrere Faktoren verallgemeinerbar.

Ob man eine orthogonale oder schiefwinklige Rotation verwendet, hängt von der Problemstellung ab. Manchmal ist davon auszugehen, dass Faktoren untereinander korreliert sind, wie bei einem Persönlichkeitsfragebogen der mehrere verwandte latente Dimensionen erfasst. In den meisten Fällen liefert eine schiefwinklige Rotation realistischere Abbildungen der Realität. Sind die Faktoren tatsächlich unkorreliert, so liefern schiefwinklige und orthogonale Rotation ähnliche Ergebnisse (Brown, 2006).

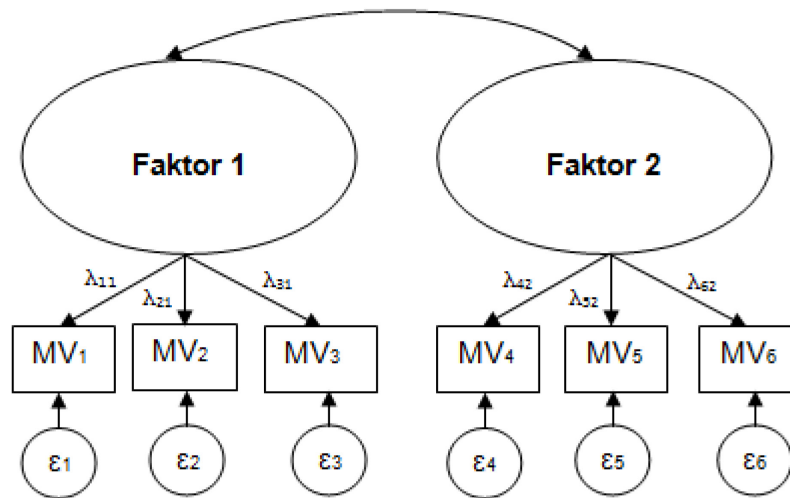
Zu den Methoden der Faktorenrotation zählen unter anderem die varimax Rotation (orthogonal) und die promax Rotation (schiefwinklig).

3. Konfirmatorische Faktorenanalyse (KFA)

Die KFA ist ein hypothesenprüfendes Verfahren. Die Datenbasis für die Analyse besteht aus einer Varianz-Kovarianz-Matrix (VC-Matrix) der Beobachtungen. Es überprüft ein Messmodell hinsichtlich der Faktorenanzahl, der Zuordnung der manifesten Variablen zu den Faktoren, der Höhe der Faktorladungen und der Beziehungen zwischen den Faktoren untereinander. Meist wird das Modell aus empirischem Vorwissen aufgestellt, indem beispielsweise bereits die Anzahl der Faktoren und das Faktorladungsmuster mittels EFA bestimmt wurden.

Ein großer Unterschied zwischen EFA und KFA besteht in der Handhabung von Nebenladungen. Wie bereits erwähnt, laden bei der EFA alle Indikatoren auf allen Faktoren. Mit Hilfe von Rotationsverfahren wird versucht, die Nebenladungen zu minimieren und die Primärladungen zu maximieren. Bei der KFA werden Sekundärladungen meist auf 0 gesetzt (siehe Abbildung 2).

Modell A: Konfirmatorische Faktorenanalyse



Modell B: Explorative Faktorenanalyse

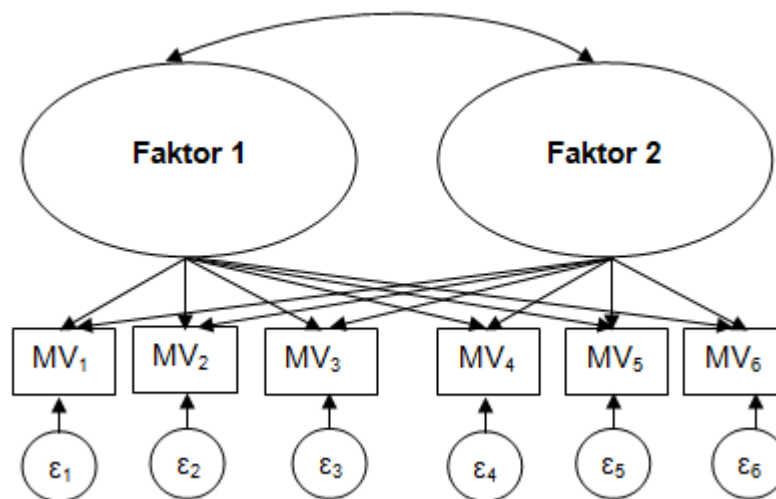


Abbildung 2: Pfaddiagramme der konfirmatorischen und explorativen Faktorenanalyse

In Abbildung 2 steht MV_i für die i -te manifeste Variable, ε_i für die i -te Fehlervarianz und λ_{ij} für die i -te Ladung auf dem j -ten Faktor. Die Faktorladungen sind in der Abbildung nur im Modell A angegeben. Der ungerichtete Pfad zwischen den beiden Faktoren deutet darauf hin, dass diese miteinander korreliert sind. Ebenso lässt sich erkennen, dass Modell A sparsamer ist als Modell B, da keine Nebenladungen geschätzt werden müssen. Im Modell A müssen nur sechs Hauptladungen, im Modell B zusätzlich sechs Nebenladungen geschätzt werden. In beiden Modellen kommen noch Schätzungen von sechs Fehlervarianzen und einer Faktorkovarianz dazu.

Zwei Modellgleichungen in Modell A würden beispielsweise wie folgt lauten:

$$MV_1 = \lambda_{11} \text{Faktor}_1 + \varepsilon_1$$

$$MV_6 = \lambda_{62} \text{Faktor}_2 + \varepsilon_6$$

Allgemeine Modellgleichung für Modell A:

$$MV_i = \lambda_{ij} \text{Faktor}_j + \varepsilon_i$$

Im Gegensatz zur KFA gibt es bei Strukturgleichungsmodellen auch gerichtete Pfade zwischen latenten Konstrukten. Die KFA spezifiziert lediglich das Messmodell, also die Beziehungen zwischen Indikatoren und Faktoren, während im Strukturmodell auch Beziehungen zwischen den latenten Dimensionen beschrieben werden (siehe Abbildung 3). In Diagrammen werden solche Beziehungen zumeist mit einem gerichteten Pfeil dargestellt. Ist der Pfeil ungerichtet, so ist eine Korrelation gemeint und man befindet sich im Messmodell.

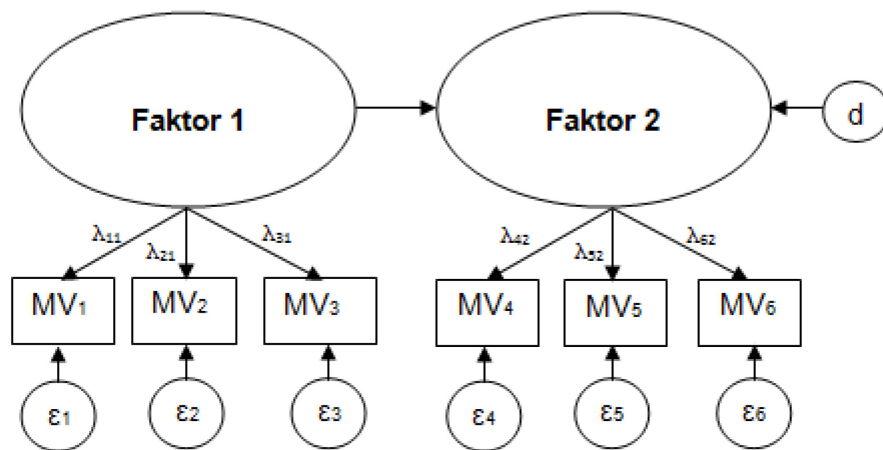


Abbildung 3: Strukturgleichungsmodell

Die KFA ist so gesehen ein Spezialfall eines Strukturgleichungsmodells. Aufgabe der KFA ist es, ein Messmodell mit entsprechenden Faktorladungen und Faktorkorrelationen zu finden, das die beobachteten Beziehungen zwischen den Indikatoren möglichst gut beschreibt (Brown, 2006).

Um Parameter im Modell schätzen zu können, muss ein Modell identifiziert sein. Deshalb befassen wir uns noch mit Modellidentifikation bevor wir zur Parameterschätzung übergehen.

3.1. Modellidentifikation

Um Modellidentifikation zu erreichen, muss den latenten Variablen eine Metrik zugewiesen werden. Zusätzlich müssen die Informationseinheiten der Varianz-

Kovarianz-Matrix (VC-Matrix), sprich die Anzahl an Varianzen und Kovarianzen, größer sein als die Anzahl der geschätzten Parameter im Modell.

Da latente Variablen unbeobachtet sind, haben sie keine definierte Metrik. Um eine Metrik für Faktoren zu definieren, gibt es zwei Möglichkeiten. Man kann eine Ladung pro Faktor auf den Wert 1 setzen. Somit wird ein gewisser Anteil der Varianz des Indikators auf den Faktor übertragen. Man nennt den entsprechenden Indikator dann Referenzindikator. Zumeist wird dabei der erste Indikator je Faktor ausgewählt. Die zweite Möglichkeit besteht darin, die Varianz des Faktors auf einen bestimmten Wert zu setzen, zumeist auf 1. Somit erhält man als Ergebnis eine standardisierte Lösung, ähnlich wie bei der EFA (Brown, 2006).

Übersteigt jedoch die Anzahl der zu schätzenden Parameter die Informationseinheiten der Ausgangsdaten, gilt ein Modell als nicht identifiziert. Ein unteridentifiziertes Modell kann nicht geschätzt werden, da es unendlich viele gleichwertige Lösungen gibt. Zur Veranschaulichung eine einfache Gleichung:

$$x + y = 5$$

Es gibt für diese Gleichung theoretisch unendlich viele Möglichkeiten für x und y , um eine Lösung zu erhalten, z.B. $x = 5$ und $y = 0$ oder $x = 4$ und $y = 1$.

Gerade identifizierte Modelle können zwar gelöst werden, aber es gibt nur genau eine Lösung, die die Varianz-Kovarianz-Matrix (VC-Matrix) der Beobachtungen perfekt reproduziert. Somit kann das Modell nicht getestet werden. Man spricht in diesem Fall von einem saturierten Modell. Zur Veranschaulichung folgendes Gleichungssystem:

$$x + y = 3$$

$$2x + y = 5$$

Die Anzahl der Unbekannten entspricht exakt der Anzahl an Bekannten, nämlich zwei Gleichungen. Die eindeutige Lösung lautet $x = 2$ und $y = 1$.

Ein Modell ist überidentifiziert, wenn die Anzahl der Bekannten die Anzahl der Unbekannten übersteigt. Bezogen auf die KFA, müssen also mehr Varianzen und Kovarianzen in der Ausgangsmatrix vorhanden sein, als Parameter im Modell geschätzt werden. Die Differenz ist dann die Anzahl der Freiheitsgrade im Modell.

Die Informationseinheiten der VC-Matrix können einfach berechnet werden:

$$b = \frac{p(p + 1)}{2}$$

In diesem Fall ist b die Anzahl der Informationseinheiten (Varianzen und Kovarianzen) und p die Anzahl der Indikatoren. Bezogen auf das Modell A in Abbildung 2 wäre $b = 21$ (6 Varianzen und 15 Kovarianzen) und man müsste 13 Parameter schätzen,

nämlich 6 Ladungen, 6 Fehlervarianzen und 1 Faktorkovarianz. Die Varianz der latenten Variablen wurde auf 1 gesetzt, um eine Metrik zu definieren. Die Anzahl der Freiheitsgrade wäre also $df = 8$ und das Modell ist identifiziert. Dasselbe Ergebnis würde man erhalten, wenn man die Ladung des ersten Indikators je Faktor auf 1 setzt. Dadurch hätte man nur 4 Ladungen zu schätzen, dafür müssten zusätzlich 2 Faktorvarianzen geschätzt werden.

Bei einem identifizierten Modell kann getestet werden, ob das Modell die Daten gut erklärt bzw. ob es im Stande ist, die VC-Matrix der Indikatoren zu reproduzieren. Dafür gibt es neben den klassischen χ^2 -Test verschiedene Fit-Indizes, die später (Kapitel Globale Fit Indizes) im Detail erläutert werden. Zunächst beschäftigen wir uns mit der Parameterschätzung.

3.2. Parameterschätzung

Das Ziel einer KFA ist es, Schätzungen für die einzelnen Parameter (Faktorladungen, Faktorvarianzen, Faktorkorrelationen etc.) des Messmodells zu erhalten, welche die empirische VC-Matrix am besten reproduzieren. Es soll also die Differenz zwischen der modell-implizierten VC-Matrix und der Stichproben VC-Matrix minimiert werden. Dafür gibt es verschiedene Methoden. Wie bereits zuvor erwähnt, wird am häufigsten die Maximum-Likelihood-Methode (ML-Methode) eingesetzt (Jöreskog, 1969). Bei der ML-Schätzung sollen diejenigen Parameter des Modells gefunden werden, die die Wahrscheinlichkeit des Auftretens der beobachteten Daten maximieren. Die Funktion, die bei der ML-Schätzung minimiert werden muss, lautet:

$$F_{ML} = \ln |S| - \ln |\Sigma| + \text{trace}[(S)(\Sigma^{-1})] - p$$

Hierbei ist $|S|$ die Determinante der VC-Matrix der Ausgangsdaten, $|\Sigma|$ die Determinante der vom Modell reproduzierten VC-Matrix, p ist die Anzahl an Indikatoren, mit \ln ist der natürliche Logarithmus gemeint und mit trace die Spur der Matrix. Die Spur ist die Summe der Diagonalelemente der Matrix, in einer VC-Matrix also die Summe der Varianzen. Die Determinante und die Spur sind Skalare und fassen wichtige Matrixeigenschaften zusammen (Brown, 2006).

Das Ziel der ML-Methode ist es, obige Funktion zu minimieren. Passt ein Modell perfekt, so stimmen die beiden VC-Matrizen völlig überein. In diesem Fall sind die Determinanten von S und Σ identisch und die Differenz aus dem Logarithmus dieser Determinanten ergibt 0. Das Produkt der beiden Matrizen $(S)(\Sigma^{-1})$ ergibt eine

Identitätsmatrix. Eine Identitätsmatrix ist eine quadratische Matrix, wobei die Diagonalelemente eins und die anderen Elemente null sind. Die Spur dieser Identitätsmatrix ist gleich der Anzahl der Indikatoren p . Somit ergibt sich bei perfekter Modellpassung $F_{ML} = 0$ (Jaccard & Wan, 1996).

Das Finden der Parameterschätzungen ist bei der KFA zumeist ein iterativer Prozess. Ausgehend von Startwerten, die die meisten Computerprogramme automatisch festlegen, werden die Schätzungen so verändert, dass die Funktion F_{ML} minimiert wird. Dieser iterative Prozess endet, wenn der Wert der Zielfunktion in zwei aufeinanderfolgenden Schritten nicht mehr wesentlich verringert werden kann. F_{ML} ist also der Wert der Fitting-Funktion, die man aufgrund der Schätzungen in der letzten Iteration erhält. Es kann aber auch vorkommen, dass es zu keiner Konvergenz kommt. Dies kann mehrere Ursachen haben. Bei komplexen Modellen kann beispielsweise die Anzahl maximaler Iterationsschritte erreicht worden sein. In diesem Fall muss die maximale Anzahl an Schritten erhöht oder die Analyse mit den zuletzt erhaltenen Parameterschätzungen als Startwert nochmals durchgeführt werden. Ein größeres Problem besteht, wenn der Prozess endet, bevor die maximale Anzahl an Iterationsschritten erreicht wurde und trotzdem keine Konvergenz gegeben ist. Dies ist häufig ein Zeichen für grobe Modellfehlspezifikation (Brown, 2006).

Des Weiteren können bei der Anwendung der ML-Methode sogenannte „*Heywood cases*“ auftreten. Damit sind Schätzungen gemeint, die Werte annehmen, die außerhalb des gültigen Bereichs liegen. Beispiele dafür sind negative Varianzen und vollständig standardisierte Ladungen oder Faktorenkorrelationen die Werte größer als 1 annehmen. Der häufigste Grund für das Auftreten solch ungültiger Lösungen ist ein fehlspezifiziertes Modell. Ebenso führen unteridentifizierte Modelle zu keiner Konvergenz oder zu absurden Lösungen. Um eine gültige Lösung zu erhalten, müssen beide VC-Matrizen positiv definit sein. Außerdem muss die modell-implizierte VC-Matrix Σ regulär sein, da man sonst die Inverse Σ^{-1} nicht bilden kann. Diese benötigt man bei der Berechnung von F_{ML} .

Bei der ML-Methode werden die Beobachtungen als multivariat normalverteilt vorausgesetzt, ansonsten kann es zu verzerrten Schätzungen kommen. ML-Schätzer haben hervorragende asymptotische Eigenschaften, das Verhalten bei kleineren Stichproben kann allerdings problematisch sein. Die Stichprobe sollte also ausreichend groß sein. Die Festlegung einer idealen Größe gestaltet sich jedoch schwierig, da diese je nach Modellkomplexität unterschiedlich sein kann. Im Falle einer Verletzung der Annahme einer multivariaten Normalverteilung ist es besser, andere

Schätzverfahren einzusetzen, wie beispielsweise Maximum Likelihood mit robusten Standardfehlern (Brown, 2006).

Die ML-Methode wird unter anderem deshalb so oft verwendet, weil sie Standardfehler der Parameterschätzungen liefert und somit auch Konfidenzintervalle für die einzelnen Parameterschätzungen angegeben werden können. Außerdem können die χ^2 -Teststatistik und andere Fit-Indizes berechnet und das Modell auf den Grad seiner Passung getestet werden. Diese Kriterien zur Modellbeurteilung werden nun im Detail erläutert.

4. Modellbeurteilung

Modellevaluation ist einer der wichtigsten Aspekte bei Strukturgleichungsmodellen. Somit ist sie auch bei der konfirmatorischen Faktorenanalyse grundlegend. Denn erst wenn das Modell passt, können theoretische und praktische Schlüsse gezogen werden. Es gibt verschiedene Möglichkeiten, um ein Modell auf seine Passung hin zu beurteilen. Grundsätzlich kann man die Kriterien in drei Kategorien einteilen. Erstens kann man den globalen Modell-Fit zur Beurteilung des Gesamtmodells heranziehen. Damit ist eine Signifikanzprüfung mit Hilfe der χ^2 -Teststatistik gemeint. Des Weiteren gibt es verschiedene Arten von Fit-Indizes zur deskriptiven Beurteilung der Modellpassung. Als dritte Variante kann man den lokalen Modell-Fit betrachten. Dabei werden einzelne Parameter überprüft und versucht, eine inhaltlich sinnvolle Interpretation für diese zu finden.

In dieser Arbeit werden die ersten beiden Kriterien (globaler Modell-Fit und Fit-Indizes) zur Modellbeurteilung herangezogen und deren Qualität verglichen. In der Praxis ist es jedenfalls sinnvoll, ein Modell auch nach lokalen Kriterien (Untersuchung einzelner Residuen) zu beurteilen.

Stehen mehrere Modelle zueinander in Konkurrenz und es soll das bestgeeignete Modell ausgewählt werden, gibt es wiederum verschiedene Beurteilungsmöglichkeiten. Diese werden in dieser Arbeit ebenfalls beschrieben und verglichen.

4.1. Globaler Modell Fit

Mit der χ^2 -Statistik kann die Übereinstimmung der modell-implizierten VC-Matrix und der Indikatoren VC-Matrix überprüft werden:

$$H_0: \Sigma = S$$

$$H_1: \Sigma \neq S$$

Es wird getestet, ob die Differenz der beiden Matrizen 0 ergibt. Die Güte des Modells wird also durch den Vergleich der modell-implizierten VC-Matrix und der Indikatoren VC-Matrix bestimmt. Sind die Verteilungsannahmen erfüllt, kann die Teststatistik bei einer ML-Schätzung wie folgt berechnet werden:

$$\chi^2 = F_{ML}(n - 1)$$

Hierbei bezieht sich F_{ML} auf den Wert der Minimierungsfunktion und n auf die Stichprobengröße. Ein statistisch signifikantes Ergebnis bedeutet, dass die beiden VC-Matrizen ungleich sind und die Parameterschätzungen im Modell die empirische VC-Matrix nicht adäquat reproduzieren. Das Modell wird verworfen (Hooper, Coughlan & Mullen, 2008).

Wie man in der Formel zur Berechnung der Statistik erkennen kann, hängt ihr Wert sehr stark von der Stichprobengröße ab. Bei sehr großen Stichproben werden also schon kleine Abweichungen zur Verwerfung der Nullhypothese führen. Andererseits müssen Modellannahmen wie multivariat normalverteilte Beobachtungen bzw. eine ausreichend große Stichprobe erfüllt sein, damit die Statistik χ^2 -verteilt ist. Dies muss in der Praxis aber nicht immer der Fall sein. Das kann dazu führen, dass erhaltene Ergebnisse falsch sind (Schermelleh-Engel, Moosbrugger & Müller, 2003). Die Statistik basiert außerdem auf der sehr strengen Nullhypothese $S = \Sigma$.

Dennoch ist die χ^2 -Statistik der einzige Index, der eine Signifikanztestung möglich macht.

4.2. Globale Fit Indizes

Neben der χ^2 -Statistik haben sich verschiedene Fit-Indizes zur Modellevaluation bei der KFA etabliert. Diese Indizes lassen sich grob in drei Kategorien einteilen:

- Absolute Fit-Indizes
- Relative Fit-Indizes
- Sparsamkeit korrigierte Fit Indizes

Da jeder der drei Indextypen unterschiedliche Informationen über die Modellanpassung enthält, wird empfohlen, zumindest aus jeder Kategorie einen Fit-Index zu betrachten (Brown, 2006). Aus diesem Grund wird nun aus jeder Kategorie ein Fit-Index beschrieben. Es werden jene Indizes gewählt, die in vorangegangenen Simulationsstudien gut abgeschnitten haben (Beauducel & Wittmann, 2005; Fan, Thompson & Wang, 1999; Hu & Bentler, 1999). Diese Indizes kommen ebenso in der hier durchgeführten Simulationsstudie zum Einsatz. Einen Überblick über Fit-Indizes

geben Hooper et al. (2008), McDonald und Ho (2002) bzw. Schermelleh-Engel et al. (2003).

Absolute Fit-Indizes

Diese Fit-Indizes beurteilen ein Modell ohne Bezugnahme auf andere Modelle. Die Passung wird auf einem absoluten Level betrachtet, nämlich ob die Hypothese $S = \Sigma$ haltbar ist. In diesem Sinne ist auch die χ^2 -Statistik ein absoluter Fit-Index.

Ein weiterer Index aus dieser Kategorie ist der *standardized root mean square residual* (SRMR). Der SRMR bestimmt die durchschnittliche Diskrepanz zwischen den Korrelationskoeffizienten der beobachteten und der modell-implizierten Korrelationsmatrix. Er wird also aus einer Residualkorrelationsmatrix bestimmt.

$$SRMR = \sqrt{\frac{2 \sum_{i=1}^p \sum_{j=1}^i \left[\frac{s_{ij} - \widehat{\sigma}_{ij}}{s_{ii} s_{jj}} \right]^2}{p(p+1)}}$$

In dieser Formel steht p für die Anzahl an Indikatoren, s_{ij} für die empirischen Kovarianzen, $\widehat{\sigma}_{ij}$ für die reproduzierten Kovarianzen und s_{ii} bzw. s_{jj} für die empirischen Standardabweichungen. Es wird die Summe der quadrierten Elemente der Residualkorrelationsmatrix berechnet und durch die Anzahl der Elemente der Matrix dividiert.

Der SRMR nimmt Werte zwischen 0 und 1 an, wobei ein Wert von 0 (kein Unterschied in den beiden Korrelationsmatrizen) perfekte Passung bedeutet. Je geringer der SRMR, desto besser die Modellanpassung (Brown, 2006).

Relative Fit-Indizes

Ein Modell wird in Bezug auf ein eingeschränkteres, genestetes „Baseline“-Modell beurteilt. Das „Baseline“-Modell ist zumeist ein Null-Modell, in dem alle Kovarianzen der beobachteten Variablen auf 0 gesetzt werden. Es postuliert also vollkommene Unabhängigkeit der Indikatoren.

Ein weit verbreiteter Index dieser Klasse ist der *comparative fit index* (CFI):

$$CFI = 1 - \frac{\chi_T^2 - df_T}{\chi_0^2 - df_0}$$

Hierbei ist χ_T^2 der χ^2 -Wert des Zielmodells und χ_0^2 jener des Nullmodells. Die Anzahl der Freiheitsgrade des Zielmodells ist durch df_T und die Anzahl der Freiheitsgrade des Nullmodells durch df_0 gegeben.

Der CFI kann Werte zwischen 0 und 1 annehmen. Je größer der CFI, desto besser die Modellanpassung (Bentler, 1990).

Sparsamkeit korrigierte Fit-Indizes

In dieser Gruppe von Fit-Indizes wird bei der Beurteilung die Anzahl der Parameter im Modell berücksichtigt. Es gibt einen Bestrafungsterm für größere Modelle. Erklären beispielsweise zwei Modelle die Daten am absoluten Level gleich gut, schneidet jenes Modell besser ab, welches weniger Parameter dafür benötigt.

Ein Index dieser Klasse ist der *root mean square error of approximation* (RMSEA). Der RMSEA basiert auf dem nichtzentralen Parameter (NKP) einer nichtzentralen χ^2 -Verteilung (Steiger & Lind, 1980). Der NKP ist durch $\chi_T^2 - df_T$ gegeben. Ist die Modellanpassung perfekt, so ist der NKP gleich 0. Um den NKP zu normalisieren wird durch $(n-1)$ dividiert.

$$RMSEA = \sqrt{\max\left\{\frac{\chi_T^2 - df_T}{df_T(n-1)}, 0\right\}}$$

χ_T^2 entspricht wieder dem Wert der Teststatistik, df_T der Anzahl an Freiheitsgraden und n der Stichprobengröße. Der RMSEA entspricht den normalisierten NKP pro Freiheitsgrad. Er ist nicht sensitiv gegenüber der Stichprobengröße, da diese auch in der Berechnung der χ^2 -Teststatistik enthalten ist und somit sowohl im Zähler als auch im Nenner vorkommt.

Der RMSEA ist nach oben hin zwar ungebunden, erreicht aber trotzdem nur selten Werte größer als 1. Ein Wert von 0 bedeutet perfekte Modellpassung.

4.3. Cut-Off-Werte

Fit-Indizes beurteilen die Modellpassung nur deskriptiv. Es können keine Signifikanztests durchgeführt werden, da die zugrunde liegende Verteilung unbekannt ist (Schermelleh-Engel et al., 2003). Daher benötigt man Cut-Off-Werte, um zu entscheiden, ob ein Modell angenommen oder verworfen werden soll.

In dieser Arbeit werden die von Hu und Bentler (1999) empfohlenen Cut-Off-Werte verwendet. Diese ermittelten sie in einer umfangreichen Simulationsstudie. In dieser

waren entweder Faktorkovarianzen oder Faktorladungen fehlspezifiziert und die Stichprobengröße wurde variiert. Modellpassung ist demnach gegeben, wenn:

SRMR < 0,11

RMSEA < 0,08 für $n < 250$

RMSEA < 0,06 für $n \geq 250$

CFI > 0,95

Diese Cut-Off-Werte werden in der Praxis sehr häufig verwendet. Sie sind teilweise als „goldene Regeln“ missinterpretiert worden, da Cut-Off-Werte nicht unabhängig von der Modellkomplexität und der Fehlspezifikation eingesetzt werden sollten. Diese Cut-Off-Werte werden auch in dieser Arbeit herangezogen, um zu überprüfen, ob sie auch bei falsch spezifizierten Faktorenstruktur passend sind.

Ähnliche Cut-Off-Werte ermittelten Beauducel und Wittmann (2005). In ihrer Simulationsstudie lag eine Modellfehlspezifikation aufgrund nicht spezifizierter Nebenladungen vor. Die Stichprobengröße, die Hauptladungen und die Faktorenanzahl wurden variiert.

4.4. Modellvergleich

Die soeben erwähnten Fit-Indizes beziehen sich immer auf ein bestimmtes Modell und dessen Passung. Will man hingegen zwei oder mehrere konkurrierende Modelle direkt miteinander vergleichen und herausfinden, welches besser zu den Daten passt, so gibt es je nach Problemstellung verschiedene Herangehensweisen. Man kann dabei einfacherweise den absoluten Wert der Fit-Indizes vergleichen, es gibt aber noch andere Methoden.

Genestete Modelle

Ein genestetes Modell enthält nur einen Teil der geschätzten Parameter eines anderen (Parent)-Modells.

So seien beispielsweise ein Modell P, das aus einem Faktor und fünf Indikatoren besteht und ein Modell N, das ebenso aus einem Faktor und fünf Indikatoren besteht, gegeben. Modell N hat jedoch die Einschränkung, dass zwei der fünf Indikatoren dieselbe Ladung auf den Faktor haben. Modell P und Modell N haben jeweils 15 Elemente in der empirischen VC-Matrix, nämlich fünf Varianzen und zehn Kovarianzen. Obwohl die Modelle strukturell identisch sind - beide bestehen aus einem Faktor und denselben fünf Indikatoren - unterscheidet sich die Anzahl der geschätzten Parameter. Bei Modell

P müssen zehn Parameter geschätzt werden (fünf Ladungen und fünf Varianzen, die Varianz des Faktors wird auf 1 gesetzt, um eine Metrik zu definieren). Bei Modell N hingegen nur neun, da für zwei Indikatoren die Ladung dieselbe ist und diese nur einmal geschätzt werden muss. Modell N hat somit sechs Freiheitsgrade ($df = 15-9$) und ist genestet unter Modell P, das fünf Freiheitsgrade ($df = 15-10$) besitzt. Man könnte aus dem Modell P das Modell N formen, indem man die die Ladung der zwei betroffenen Indikatoren auf Gleichheit beschränkt (Brown, 2006).

Sind Modelle genestet, so kann die χ^2 -Statistik der beiden Modelle verwendet werden, um sie zu vergleichen. Man nennt diesen Test auch χ^2 -Differenztest oder genesteter χ^2 -Test (Schermelleh-Engel et al., 2003).

$$\chi_{diff}^2 = \chi_N^2 - \chi_P^2$$

$$df_{diff} = df_N - df_P$$

Die Differenz der beiden χ^2 -Werte ist wiederum χ^2 -verteilt. Die Anzahl der Freiheitsgrade entspricht ebenso der Differenz der Freiheitsgrade in den beiden Modellen.

In obigen Beispiel hätte der χ^2 -Differenztest einen Freiheitsgrad ($df = 6-5$). Erhält man ein signifikantes Ergebnis bedeutet das, dass das Parent-Modell die Daten signifikant besser erklärt als das genestete Modell. Andernfalls erklärt das genestete Modell die Daten nicht signifikant schlechter. Wichtig ist, dass das größere Modell die Daten gut erklärt. Ansonsten ist die Anwendung des Differenztests nicht sinnvoll. Der χ^2 -Differenztest besitzt dieselben Vor- und Nachteile (Einfluss der Stichprobengröße) wie der χ^2 -Test (Schermelleh-Engel et al., 2003).

Ebenso können Modelle genestet sein, die sich in der Anzahl an Faktoren unterscheiden. Beispielsweise sind ein 1-Faktor Modell mit acht Indikatoren und ein 2-Faktoren-Modell mit je 4 Indikatoren genestet, wenn die beiden Faktoren des 2-Faktoren-Modells untereinander korrelieren dürfen. Im 1-Faktor Modell müssen 16 Parameter geschätzt werden und es besitzt somit 20 Freiheitsgrade ($df = 36-16$). Im 2-Faktoren-Modell müssen 17 Parameter geschätzt werden, da noch eine Schätzung für die Faktorkorrelation hinzukommt. Somit hat es 19 Freiheitsgrade ($df = 36-17$) und der χ^2 -Differenztest könnte angewandt werden. Das Parent-Modell wäre das 2-Faktoren-Modell. Obwohl die Modelle strukturell nun unterschiedlicher wirken als zuvor, sind sie trotzdem genestet, da man das 1-Faktor Modell aus dem 2-Faktoren-Modell herleiten kann. Dazu müsste man die Faktorkorrelation im 2-Faktoren-Modell auf den Wert 1 festlegen. Somit bräuchte man diesen Parameter nicht schätzen und als Ergebnis würde man das 1-Faktor Modell erhalten, da die beiden Faktoren dann identisch sind.

Dieses Prinzip kann man auf Modelle mit einer höheren Anzahl an Faktoren verallgemeinern. So sind beispielsweise ein 3-Faktoren-Modell und ein 2-Faktoren-Modell ebenfalls genestet, wenn zwei der Faktoren des ersteren Modells zusammengelegt werden (Brown, 2006).

Sind Modelle mit ungleicher Faktorenanzahl nicht genestet, weil beispielsweise die Korrelationen der Faktoren auf 0 gesetzt werden und somit beide Modelle dieselbe Anzahl an Freiheitsgraden besitzen, so bietet sich zum Vergleich das Akaike Informationskriterium (AIK) an.

Akaike Informationskriterium (AIK)

Beim AIK fließen sowohl die Modellpassung als auch die Modellkomplexität in die Beurteilung mit ein. Es gibt verschiedene Berechnungsmethoden für das AIK, aber alle führen zu äquivalenten Ergebnissen.

$$AIK = -2(\loglikelihood) + 2a$$

Dabei bezieht sich loglikelihood auf den Wert der logarithmierten Likelihood-Funktion und a auf die Anzahl der geschätzten Parameter im Modell. Unabhängig von der Berechnungsmethode ist immer jenes Modell zu bevorzugen, das den geringsten Wert aufweist. Mit dem AIK können auch nicht genestete Modelle miteinander verglichen werden. Der Vergleich ist allerdings rein deskriptiver Natur ist, da keine Signifikanztests durchgeführt werden können (Akaike, 1987).

F-Test

Den F-Test zum Vergleich von unterschiedlich großen Modellen schlagen Kubinger, Litzenberger und Mrakotsky vor (2006). Da der Quotient zweier χ^2 -verteilter Zufallsvariablen F-verteilt ist, wird die besagte F-Statistik als Entscheidungsgrundlage für den Modellvergleich herangezogen:

$$F = \frac{\chi_T^2 / df_T}{\chi_B^2 / df_B}$$

Dabei sind χ_T^2 bzw. χ_B^2 die χ^2 -Teststatistiken der beiden zu vergleichenden Modelle und df_T bzw. df_B die Anzahl an Freiheitsgraden. Ist das Ergebnis signifikant, passt das Zielmodell signifikant schlechter zu den Daten als das Vergleichsmodell („Baseline“-Modell). Analog zum χ^2 -Differenztest sollte auch hier das „Baseline“-Modell die Daten ausreichend gut erklären.

Diese F-Statistik beruht allerdings auf χ^2 -verteilten Statistiken, die unter Umständen nicht auf unabhängigen Parameterschätzungen beruhen. Daher ist die F-Statistik nicht zwingend F-verteilt.

III. Simulationsstudie

In diesem Abschnitt wird die durchgeführte Simulationsstudie erläutert und Ergebnisse diskutiert.

5. Ziel der Simulationsstudie

In der Simulationsstudie wurde das Verhalten der χ^2 -Teststatistik und der beschriebenen Fit-Indizes (SRMR, CFI, RMSEA) bei falscher Faktorenstruktur untersucht. Ist die Anzahl an Faktoren fehlspezifiziert, sollten die erwähnten Statistiken dies erkennen. Der Anteil an Fällen, in denen diese Modelle beim χ^2 -Test nicht verworfen wurden, liefert eine Schätzung für das eingegangene Risiko 2. Art. Ebenso wurden richtig spezifizierte Modelle getestet, um das tatsächlich eingegangene Risiko 1. Art zu bestimmen.

Da mit den Fit-Indizes nicht auf Signifikanz getestet werden kann, werden Cut-Off-Werte herangezogen, um eine entsprechende Entscheidung bezüglich der Modellannahme zu treffen. Somit kann überprüft werden, ob die in bereits durchgeführten Simulationsstudien ermittelten Cut-Off-Werte auch bei fehlspezifizierter Faktorenstruktur passend sind.

Sofern es möglich war, wurden Modelle mit unterschiedlicher Faktorenanzahl mit den vorhin dargestellten Methoden (AIK, χ^2 -Differenztest, F-Test) untereinander verglichen. Der χ^2 -Differenztest konnte nur bei Modellen angewandt werden, die sich in der Anzahl der Freiheitsgrade unterschieden.

Dies wurde unter verschiedenen Bedingungen untersucht. So wurde beispielsweise die Höhe der Faktorladungen und die Stichprobengröße variiert.

6. Methoden

Die Simulationen wurden mit dem Statistikprogramm R durchgeführt (R Development Core Team, 2012). Dabei wurde das Paket lavaan verwendet (Rosseel, 2012).

6.1. Populationsmodelle

Der Begriff Populationsmodell steht für das „wahre“ Modell. Aus diesem „richtigen“ Modell werden die multivariat normalverteilten manifesten Variablen generiert. Es

wurden drei verschiedene Populationsmodelle aufgestellt. Alle Populationsmodelle bestehen aus drei Faktoren und 24 manifesten Variablen.

Populationsmodell ohne korrelierte Faktoren, ohne Nebenladungen

Bei diesem Modell laden auf jedem Faktor je 8 Indikatoren. Da die Höhe der Faktorladungen Einfluss auf die Fit-Indizes und die Power des χ^2 -Tests hat, wurde zwischen niedrigen, mittleren, hohen und gemischten Faktorladungen unterschieden (vgl. Heene et al., 2011). Die Faktorladungen wurden für die 24 manifesten Variablen zufällig bestimmt, indem sie aus gleichverteilten Zufallsvariablen gezogen wurden. Für niedrige, mittlere, hohe und gemischte Ladungen wurden je 24 Zufallsvariablen aus den Intervallen [0,3; 0,5], [0,5; 0,7], [0,7; 0,9] und [0,3; 0,9] gezogen. Das Populationsmodell ist für niedrige Faktorladungen in Abbildung 4 schemenhaft dargestellt. Um die Werte für die mittleren bzw. hohen Faktorladungen zu erhalten, muss man zu den gegebenen Hauptfaktorladungen 0,2 bzw. 0,4 addieren. So bleiben die absoluten Differenzen im Faktorladungsprofil dieselben und man kann den Einfluss der Ladungshöhe kontrollieren.

Indikator	Ladungen		
	Faktor 1	Faktor 2	Faktor 3
x1	0,50	0,00	0,00
x2	0,38	0,00	0,00
x3	0,32	0,00	0,00
x4	0,31	0,00	0,00
x5	0,35	0,00	0,00
x6	0,46	0,00	0,00
x7	0,37	0,00	0,00
x8	0,49	0,00	0,00
x9	0,00	0,33	0,00
x10	0,00	0,39	0,00
x11	0,00	0,33	0,00
x12	0,00	0,35	0,00
x13	0,00	0,45	0,00
x14	0,00	0,32	0,00
x15	0,00	0,39	0,00
x16	0,00	0,32	0,00
x17	0,00	0,00	0,41
x18	0,00	0,00	0,30
x19	0,00	0,00	0,50
x20	0,00	0,00	0,36
x21	0,00	0,00	0,43
x22	0,00	0,00	0,36
x23	0,00	0,00	0,50
x24	0,00	0,00	0,48

Abbildung 4: Populationsmodell mit niedrigen Faktorladungen

Im gemischten Faktorladungs-Modell war die Zuteilung der Indikatoren zu den Faktoren dieselbe. Die Ladungen waren dabei wie folgt gegeben:

Faktor 1 (x1 bis x8): 0,89; 0,54; 0,37; 0,34; 0,45; 0,78; 0,50; 0,88

Faktor 2 (x9 bis x16): 0,40; 0,58; 0,40; 0,44; 0,76; 0,36; 0,57; 0,35

Faktor 3 (x17 bis x24): 0,64; 0,31; 0,89; 0,49; 0,68; 0,48; 0,90; 0,84

Populationsmodell mit korrelierten Faktoren, ohne Nebenladungen

Dieses Populationsmodell ist analog zum vorherigen mit dem einzigen Unterschied, dass die Faktoren in folgender Größenordnung korrelieren:

$$\rho_{F1,F2} = 0,5$$

$$\rho_{F1,F3} = 0,4$$

$$\rho_{F2,F3} = 0,3$$

Dieselben Korrelationskoeffizienten wurden bereits in anderen Simulationsstudien verwendet (Hu & Bentler, 1999; Heene et al., 2011).

Populationsmodell ohne korrelierte Faktoren, mit Nebenladungen

Bei diesem Modell wurde zwischen Haupt- und Nebenladungen differenziert (siehe

Indikator	Ladungen		
	Faktor 1	Faktor 2	Faktor 3
x1	0,89	0,20	0,08
x2	0,52	0,02	0,14
x3	0,76	0,09	0,00
x4	0,79	0,11	0,17
x5	0,64	0,18	0,00
x6	0,71	0,03	0,04
x7	0,53	0,20	0,18
x8	0,58	0,19	0,12
x9	0,18	0,63	0,08
x10	0,19	0,85	0,09
x11	0,06	0,38	0,01
x12	0,17	0,72	0,19
x13	0,13	0,40	0,09
x14	0,10	0,66	0,19
x15	0,15	0,60	0,18
x16	0,03	0,84	0,13
x17	0,13	0,02	0,54
x18	0,14	0,10	0,32
x19	0,09	0,08	0,34
x20	0,14	0,18	0,58
x21	0,19	0,09	0,41
x22	0,05	0,17	0,63
x23	0,09	0,15	0,61
x24	0,19	0,16	0,38

Abbildung 5: Populationsmodell mit Nebenladungen

Abbildung 5). Die Hauptladungen wurden aus gleichverteilten Zufallszahlen zwischen 0,3 und 0,9 gezogen, die Nebenladungen aus dem Intervall zwischen 0 und 0,2. Alle manifesten Variablen laden auf allen Faktoren. Die Variablen x1 bis x8 haben ihre Hauptladungen auf Faktor 1, x9 bis x16 auf Faktor 2 und x17 bis x24 auf Faktor 3. Der Rest sind Sekundärladungen.

6.2. Simulationsbedingungen

Alle Modelle wurden für Stichprobengrößen von $n = 150$ und $n = 600$ getestet. Kreuzt man nun alle Bedingungen der ersten beiden Populationsmodelle vollständig miteinander, ergeben sich

2 (korreliert ja/nein) * 4 (Höhe der Faktorladung) * 2 (Stichprobengröße) = 16 Bedingungen. Inklusive des Populationsmodells mit Nebenladungen für beide Stichprobengrößen ergeben sich $16 + 2 = 18$ Simulationsszenarien. In jedem dieser 18 Szenarien wurden drei verschiedene Modelle überprüft, ein 2-Faktoren-Modell (fehlspezifiziert, ein Faktor zu wenig), ein 3-Faktoren-Modell (richtig spezifiziert) und ein 4-Faktoren-Modell (fehlspezifiziert, ein Faktor zu viel). Jede Simulation wurde mit 10.000 Wiederholungen durchgeführt, das heißt insgesamt wurden 540.000 Modelle gefittet.

6.3. Aufgestellte Modelle

Aus den Populationsmodellen wurden die (manifesten) Daten generiert. Es wurden multivariate normalverteilte Beobachtungen simuliert, um richtig aufgestellte und fehlspezifizierte Modelle anhand der ML-basierten χ^2 -Statistik und anhand der Fit-Indizes zu evaluieren.

Das 3-Faktoren-Modell hat immer die exakt gleiche Struktur wie das entsprechende Populationsmodell. Falls es Nebenladungen gibt, sind auch diese richtig angegeben.

Sind im Populationsmodell keine Korrelationen zwischen den Faktoren spezifiziert, so werden auch bei den zu testenden Modellen die Korrelationen unterdrückt. Sind welche spezifiziert, so dürfen die Faktoren in den zu testenden Modellen korrelieren.

Eine Ausnahme bildet hierbei das 4-Faktoren-Modell. Dabei darf der zusätzliche Faktor nicht mit den anderen Faktoren korrelieren, da es sonst zu einem zu hohen Anteil an ungültigen Lösungen kommt. Bei der Parameterschätzung wird dabei für den Korrelationskoeffizienten des dritten und vierten Faktors häufig ein Wert größer als 1 ermittelt.

Im 2-Faktoren-Modell werden zwei Faktoren zu einem zusammengefasst, während im 4-Faktoren-Modell ein Faktor auf zwei separate aufgeteilt wird. In Abbildung 6 sind die Indikatoren bei jenen Faktoren eingetragen, auf denen sie in den fehlspezifizierten Modellen laden.

2-Faktoren-Modell		4-Faktoren-Modell			
Faktor 1	Faktor 2	Faktor 1	Faktor 2	Faktor 3	Faktor 4
x1		x1			
x2		x2			
x3		x3			
x4		x4			
x5		x5			
x6		x6			
x7		x7			
x8		x8			
	x9		x9		
	x10		x10		
	x11		x11		
	x12		x12		
	x13		x13		
	x14		x14		
	x15		x15		
	x16		x16		
x17				x17	
x18				x18	
x19				x19	
x20				x20	
x21					x21
x22					x22
x23					x23
x24					x24

Abbildung 6: Fehlspezifizierte Modelle

Um eine Metrik für die latenten Dimensionen zu definieren, wurde die Varianz jedes Faktors auf 1 gesetzt. In der empirischen Korrelationsmatrix befinden sich $\frac{24 \cdot 25}{2} = 300$ Informationseinheiten, 24 Varianzen und 276 Kovarianzen. Je nach Modell müssen unterschiedlich viele Parameter geschätzt werden. Die Anzahl an Freiheitsgraden ist jedoch stets positiv und ist im Ergebnisteil für die verschiedenen Modelle angeführt. Die Modelle sind also identifiziert.

6.4. Auswertungsmethoden

Die χ^2 -Statistik und die Fit-Indizes wurden für jedes gefittete Modell berechnet. Pro Simulationsbedingung gibt es somit für jedes Modell und jede Statistik 10.000 Werte.

In jeder Simulationsbedingung wurde für jede Fit-Statistik der Anteil an angenommenen 2-Faktoren, 3-Faktoren und 4-Faktoren-Modellen ermittelt.

Der Anteil an angenommenen 2-Faktoren und 4-Faktoren-Modellen liefert jeweils eine Schätzung für das Risiko 2. Art, da diese Modelle fehlspezifiziert sind. Der Anteil an verworfenen 3-Faktoren-Modellen liefert eine Schätzung für das Risiko 1. Art, da dieses Modell angenommen werden sollte. Für den χ^2 -Test betrug das Signifikanzniveau 5 %. Für die Fit-Indizes wurden wie beschrieben die Cut-Off-Werte von Hu und Bentler (1999) verwendet.

Ebenso wurden Mittelwerte und Standardabweichungen der χ^2 -Statistik und der Fit-Indizes je Simulationsbedingung berechnet.

Aufgrund der Fehlspezifizierung der Modelle kann es vorkommen, dass ungültige Lösungen auftreten. Dabei kann es sich, wie im Theorieteil beschrieben, beispielsweise um negative Varianzen oder Faktorladungen größer als 1 handeln. Treten ungültige Lösungen auf, wird das betroffene Modell sofort verworfen und keine Fit-Statistiken berechnet, da diese dann ohnehin verfälscht wären. Das Modell wird deshalb verworfen, da ungültige Lösungen zumeist ein Hinweis auf grobe Modellfehlspezifikation sind (Brown, 2006). Der Anteil an ungültigen Modellen je Simulationsbedingung wird im Ergebnisteil angeführt.

Zum Modellvergleich wurde in jeder Simulationsbedingung für die drei aufgestellten Modelle der Mittelwert des Akaike Informationskriteriums berechnet. Diese Mittelwerte wurden deskriptiv miteinander verglichen.

Falls die Anzahl der Freiheitsgrade im 3-Faktoren-Modell unterschiedlich von denen im 2-Faktoren und 4-Faktoren-Modell war, konnte der χ^2 -Differenztest durchgeführt werden. Dies war beim Populationsmodell mit Nebenladungen der Fall. Das Signifikanzniveau war 5 %. Es wurde ermittelt, zu welchem Anteil das 3-Faktoren-Modell signifikant besser passt als das 2-Faktoren bzw. 4-Faktoren-Modell. Dabei wurden nur jene Modelle miteinander verglichen, in denen das 3-Faktoren-Modell beim χ^2 -Test nicht verworfen wurde.

Ebenso wurde für die F-Statistik in jeder Simulationsbedingung jener Anteil ermittelt, zu welchem das 3-Faktoren-Modell die Daten signifikant besser erklärt als das 2-Faktoren bzw. 4-Faktoren-Modell ($\alpha = 0,05$). Analog zum genesteten Modelltest wurden hier nur

Modelle verglichen, in denen das 3-Faktoren-Modell beim χ^2 -Test angenommen wurde.

7. Ergebnisse

In Tabelle 1 sind die jeweiligen Anteile an ungültigen Lösungen wiedergegeben.

Tabelle 1: Anteil an ungültigen Lösungen je Bedingung und Modell

Modell	Faktorladung							
	Niedrig		Mittel		Hoch		Gemischt	
	n = 150	n = 600	n = 150	n = 600	n = 150	n = 600	n = 150	n = 600
Unkorreliert								
2-Faktoren	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
3-Faktoren	0,001	0,000	0,000	0,000	0,000	0,000	0,000	0,000
4-Faktoren	0,062	0,000	0,000	0,000	0,000	0,000	0,105	0,005
Korreliert								
2-Faktoren	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
3-Faktoren	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
4-Faktoren	0,041	0,000	0,000	0,000	0,000	0,000	0,063	0,001
Nebenladung								
2-Faktoren							0,000	0,000
3-Faktoren							0,000	0,000
4-Faktoren							0,023	0,000

Wie bereits erwähnt, wurden beim 4-Faktoren-Modell im korrelierten Populationsmodell die Korrelationen des vierten Faktors mit den anderen Faktoren auf 0 gesetzt. Sonst würde der Anteil an ungültigen Lösungen in allen acht Bedingungen über 0,52 liegen. Ansonsten hält sich der Anteil an ungültigen Lösungen in Grenzen. Der Maximalwert beträgt 0,105 (4-Faktoren-Modell, n = 150, gemischte Ladungshöhe), zumeist ist er jedoch 0. Unabhängig vom Populationsmodell treten ungültige Lösungen am häufigsten im 4-Faktoren-Modell bei n = 150 und niedriger oder gemischter Ladungshöhe auf. Die ungültigen Lösungen werden bei den weiteren Analysen sofort als unpassend klassifiziert, da sie ein Anzeichen grober Modellfehlspezifikation sind.

7.1. Risiko 1. Art: Das 3-Faktoren-Modell

Um das tatsächlich eingegangene Risiko 1. Art abzuschätzen, wurde überprüft, wie häufig das 3-Faktoren-Modell in jeder Simulationsbedingung abgelehnt wurde (siehe Tabelle 2).

Bei einer Stichprobengröße von $n=150$ wird das nominelle Risiko 1. Art (5 %) beim χ^2 -Test nicht eingehalten. Je nach Bedingung wurden zwischen 20,6 % und 25,4 % der Modelle verworfen. Der RMSEA und SRMR schneiden hier in gewisser Weise besser ab. Allerdings liegt der Anteil an verworfenen Modellen sogar nahezu bei 0. Die Mittelwerte der beiden Fit-Statistiken liegen immer unterhalb ihres Cut-Off-Wertes (RMSEA $<0,08$ und SRMR $<0,11$). Der CFI hat scheinbar Probleme, richtige Modelle anzunehmen, wenn die Faktorladungen niedrig bzw. mäßig sind. Bei niedrigen Ladungen liegt der Anteil an Verwerfungen zwischen 0,589 und 0,613, bei mäßigen Ladungen zwischen 0,139 und 0,161. Bei niedrigen Ladungen beträgt der Mittelwert des CFI 0,92, und liegt somit unter dem Cut-Off-Wert von 0,95.

Ist die Stichprobe größer ($n = 600$), nimmt der χ^2 -Test mehr richtig spezifizierte Modelle an. Der Anteil liegt über alle Bedingungen hinweg zwischen 0,074 und 0,097. Das nominelle Risiko 1. Art von 5 % wird somit trotzdem deutlich nicht eingehalten. Der SRMR und der RMSEA verwerfen kein richtig spezifiziertes Modell. Die Mittelwerte der beiden Indizes liegen weit unterhalb ihrer Cut-Off-Werte. Der CFI verbessert sich bei größerer Stichprobe deutlich. Es werden nur mehr zwischen 0 % und 5 % der Modelle verworfen. Die Mittelwerte erstrecken sich von 0,988 bis 0,999 und liegen somit immer deutlich über den Cut-Off-Wert.

Tabelle 2: Risiko 1. Art bei unterschiedlichen Simulationsbedingungen

Mittelwerte der χ^2 -Statistik und Fit Indizes (Standardabweichungen) und [Anteil Fehler 1. Art] für korrekt spezifizierte 3-Faktoren-Modelle je Simulationsbedingung

Simulationsszenario	Stichprobengröße	
	n = 150	n = 600
Unkorreliert		
Chi-Quadrat ^a		
Niedrig	270,455 (24,142) [0,208]	256,313 (22,604) [0,075]
Mittel	270,985 (24,207) [0,214]	256,418 (22,644) [0,078]
Hoch	271,151 (24,232) [0,215]	256,507 (22,623) [0,075]
Gemischt	271,100 (24,290) [0,218]	256,464 (22,534) [0,076]
SRMR		
Niedrig	0,072 (0,004) [0,001]	0,036 (0,002) [0,000]
Mittel	0,069 (0,006) [0,000]	0,035 (0,003) [0,000]
Hoch	0,066 (0,011) [0,002]	0,033 (0,006) [0,000]
Gemischt	0,070 (0,006) [0,000]	0,035 (0,003) [0,000]
CFI		
Niedrig	0,912 (0,078) [0,613]	0,987 (0,017) [0,050]
Mittel	0,975 (0,023) [0,161]	0,997 (0,005) [0,000]
Hoch	0,991 (0,008) [0,000]	0,999 (0,002) [0,000]
Gemischt	0,981 (0,018) [0,063]	0,997 (0,003) [0,000]
RMSEA		
Niedrig	0,020 (0,014) [0,001]	0,006 (0,006) [0,000]
Mittel	0,020 (0,014) [0,000]	0,006 (0,006) [0,000]
Hoch	0,020 (0,014) [0,000]	0,006 (0,006) [0,000]
Gemischt	0,020 (0,014) [0,000]	0,006 (0,006) [0,000]
Korreliert		
Chi-Quadrat ^b		
Niedrig	267,305 (24,119) [0,206]	253,248 (22,480) [0,074]
Mittel	267,946 (24,206) [0,214]	253,463 (22,477) [0,074]
Hoch	267,972 (24,166) [0,212]	253,400 (22,458) [0,076]
Gemischt	267,964 (24,261) [0,216]	253,376 (22,352) [0,077]
SRMR		
Niedrig	0,068 (0,003) [0,001]	0,034 (0,002) [0,000]
Mittel	0,061 (0,004) [0,000]	0,030 (0,002) [0,000]
Hoch	0,047 (0,004) [0,000]	0,024 (0,002) [0,000]
Gemischt	0,061 (0,004) [0,000]	0,031 (0,002) [0,000]
CFI		
Niedrig	0,920 (0,072) [0,589]	0,988 (0,016) [0,034]
Mittel	0,976 (0,022) [0,139]	0,997 (0,004) [0,000]
Hoch	0,992 (0,008) [0,000]	0,999 (0,001) [0,000]
Gemischt	0,982 (0,017) [0,051]	0,998 (0,003) [0,000]
RMSEA		
Niedrig	0,020 (0,014) [0,001]	0,006 (0,006) [0,000]
Mittel	0,020 (0,014) [0,000]	0,006 (0,006) [0,000]
Hoch	0,020 (0,014) [0,000]	0,006 (0,006) [0,000]
Gemischt	0,020 (0,014) [0,000]	0,006 (0,006) [0,000]
Nebenladungen		
Chi-Quadrat ^c	224,296 (22,146) [0,254]	211,049 (20,694) [0,097]
SRMR	0,041 (0,003) [0,000]	0,020 (0,001) [0,000]
CFI	0,983 (0,014) [0,025]	0,998 (0,003) [0,000]
RMSEA	0,023 (0,014) [0,000]	0,007 (0,007) [0,000]

Anmerkung. ^adf = 252. ^bdf = 249. ^cdf = 204.

7.2. Risiko 2. Art: Das 2-Faktoren-Modell

Der Anteil an angenommenen Modellen im 2-Faktoren-Modell liefert eine Schätzung für das Risiko 2. Art. In Tabelle 3 sind Mittelwerte, Standardabweichungen und Anteile an angenommenen Modellen für jede Statistik dargestellt.

Bei niedrigen Ladungen und einer Stichprobengröße von $n = 150$ nimmt der χ^2 -Test das falsch spezifizierte 2-Faktoren-Modell in manchen Fällen an. Der Anteil an falschen Entscheidungen liegt zwischen 0,061 und 0,222. Die Teststärke beträgt somit im mindesten 0,778. In der Bedingung mit der größeren Stichprobe verwirft der χ^2 -Test alle 2-Faktoren-Modelle.

Der SRMR nimmt einen deutlich größeren Anteil an falsch spezifizierten Modellen an. Unabhängig von der Stichprobengröße werden bei niedrigen bzw. mäßigen Ladungen zwischen 4,6 % und 100 % der Modelle angenommen. Die Mittelwerte des SRMR befinden sich in diesen Bedingungen zumeist unterhalb oder in der Nähe des Cut-Off-Wertes. Im korrelierten Populationsmodell werden Modelle mit gemischten Ladungen zu einem hohen Teil nicht verworfen. Nur bei hoher Faktorladung werden 2-Faktoren-Modelle durch den SRMR richtigerweise verworfen.

Ein ähnliches Bild zeigt sich für den RMSEA. Bei niedrigen bzw. mäßigen Ladungen und $n = 150$ nimmt der RMSEA zwischen 13,5 % und 100 % der falsch spezifizierten Modelle an. Bei einer größeren Stichprobe werden bei niedrigen Ladungen immer noch alle 2-Faktoren-Modelle angenommen. Ist die Faktorladung allerdings hoch, werden keine falsch spezifizierten Modelle angenommen. Die Mittelwerte des RMSEA liegen zwischen 0,034 und 0,152. Sie liegen vor allem bei niedrigen bzw. mäßigen Ladungen unter dem Cut-Off-Wert bzw. sind annähernd identisch mit diesem.

Der CFI liefert beim 2-Faktoren-Modell über alle Bedingungen hinweg die besten Ergebnisse. Es werden nahezu alle Modelle verworfen. Der höchste Anteil an beibehaltenen Modellen liegt bei 0,034. Die Mittelwerte des CFI liegen zwischen 0,622 und 0,837 und sind damit immer weit unterhalb des Cut-Off-Wertes.

Tabelle 3: Risiko 2. Art bei unterschiedlichen Simulationsbedingungen für das 2-Faktoren-Modell

Mittelwerte der χ^2 -Statistik und Fit Indizes (Standardabweichungen) und [Anteil Fehler 2. Art] für falsch spezifizierte Modelle (2-Faktoren) je Simulationsbedingung

Simulationsszenario	Stichprobengröße	
	n = 150	n = 600
Unkorreliert		
Chi-Quadrat ^a		
Niedrig	333,721 (29,400) [0,061]	537,565 (43,101) [0,000]
Mittel	541,861 (46,302) [0,000]	1392,172 (87,788) [0,000]
Hoch	1122,117 (79,998) [0,000]	3712,214 (165,350) [0,000]
Gemischt	677,844 (54,856) [0,000]	1888,571 (102,534) [0,000]
SRMR		
Niedrig	0,084 (0,005) [0,999]	0,060 (0,003) [1,000]
Mittel	0,124 (0,008) [0,046]	0,114 (0,005) [0,180]
Hoch	0,201 (0,009) [0,000]	0,197 (0,005) [0,000]
Gemischt	0,133 (0,008) [0,001]	0,121 (0,004) [0,002]
CFI		
Niedrig	0,649 (0,078) [0,003]	0,660 (0,037) [0,000]
Mittel	0,661 (0,039) [0,000]	0,658 (0,020) [0,000]
Hoch	0,655 (0,024) [0,000]	0,655 (0,013) [0,000]
Gemischt	0,622 (0,038) [0,000]	0,631 (0,018) [0,000]
RMSEA		
Niedrig	0,046 (0,009) [0,999]	0,043 (0,003) [1,000]
Mittel	0,087 (0,007) [0,135]	0,087 (0,003) [0,000]
Hoch	0,152 (0,007) [0,000]	0,151 (0,004) [0,000]
Gemischt	0,106 (0,007) [0,000]	0,104 (0,003) [0,000]
Korreliert		
Chi-Quadrat ^b		
Niedrig	311,035 (28,388) [0,222]	426,627 (39,015) [0,000]
Mittel	466,505 (45,760) [0,000]	1058,450 (84,862) [0,000]
Hoch	1001,626 (80,218) [0,000]	3235,112 (162,868) [0,000]
Gemischt	621,181 (53,375) [0,000]	1668,846 (99,945) [0,000]
SRMR		
Niedrig	0,076 (0,005) [1,000]	0,048 (0,003) [1,000]
Mittel	0,097 (0,010) [0,881]	0,081 (0,006) [1,000]
Hoch	0,169 (0,022) [0,002]	0,169 (0,004) [0,000]
Gemischt	0,116 (0,012) [0,301]	0,104 (0,007) [0,777]
CFI		
Niedrig	0,767 (0,095) [0,034]	0,813 (0,037) [0,000]
Mittel	0,760 (0,046) [0,000]	0,771 (0,022) [0,000]
Hoch	0,709 (0,028) [0,000]	0,709 (0,014) [0,000]
Gemischt	0,686 (0,040) [0,000]	0,695 (0,019) [0,000]
RMSEA		
Niedrig	0,039 (0,011) [1,000]	0,034 (0,004) [1,000]
Mittel	0,075 (0,008) [0,695]	0,073 (0,004) [0,000]
Hoch	0,141 (0,008) [0,000]	0,141 (0,004) [0,000]
Gemischt	0,099 (0,007) [0,005]	0,097 (0,003) [0,000]
Nebenladungen		
Chi-Quadrat ^c	479,702 (43,213) [0,000]	1093,686 (77,021) [0,000]
SRMR	0,147 (0,015) [0,004]	0,137 (0,008) [0,000]
CFI	0,827 (0,029) [0,000]	0,837 (0,013) [0,000]
RMSEA	0,077 (0,007) [0,612]	0,075 (0,003) [0,000]

Anmerkung. ^adf = 252. ^bdf = 251. ^cdf = 252.

7.3. Risiko 2. Art: Das 4-Faktoren-Modell

Analog wurde überprüft, wie viele Modelle im ebenfalls falsch aufgestellten 4-Faktoren-Modell angenommen wurden (siehe Tabelle 4). Grundsätzlich ähneln die Ergebnisse denen vom 2-Faktoren-Modell. Es werden allerdings zumeist noch mehr falsch spezifizierte Modelle angenommen.

Der χ^2 -Test nimmt bei niedrigen Ladungen und $n = 150$ zwischen 20,6 % und 23,7 % der 4-Faktoren-Modelle an. In den anderen Simulationsbedingungen verwirft der χ^2 -Test alle falsch aufgestellten Modelle.

Beim SRMR ist der Anteil angenommener 4-Faktoren-Modelle vor allem in den Bedingungen mit niedrigen bzw. mäßigen Ladungen teilweise erstaunlich hoch. Dies gilt für beide Stichprobengrößen. Der Anteil liegt in diesen Szenarien zwischen 0,062 und 1. Die Mittelwerte des SRMR betragen dabei zwischen 0,055 und 0,125.

Der RMSEA hat beim 4-Faktoren-Modell bei kleiner Stichprobe in allen Bedingungen Schwierigkeiten, das falsch spezifizierte 4-Faktoren-Modell zu erkennen. Bei gemischter Faktorladung liegt der Anteil falsch angenommener Modelle beispielsweise zwischen 0,638 und 0,943. Bei $n = 600$ bleibt das Problem zumindest noch bei niedriger und mäßiger Ladungshöhe bestehen. Die Mittelwerte liegen bei $n = 150$ zwischen 0,037 und 0,086 und bei $n = 600$ zwischen 0,032 und 0,084.

Beim CFI liegt der Anteil an fälschlicherweise angenommenen Modellen nur in zwei Bedingungen über 0. Bei $n = 150$ und niedriger Ladung beträgt der Anteil 0,031 im unkorrelierten bzw. 0,026 im korrelierten Populationsmodell. Die Mittelwerte liegen zwischen 0,762 und 0,898 und damit unterhalb des Cut-Off-Wertes.

Tabelle 4: Risiko 2. Art bei unterschiedlichen Simulationsbedingungen für das 4-Faktoren-Modell

Mittelwerte der χ^2 -Statistik und Fit Indizes (Standardabweichungen) und [Anteil Fehler 2. Art] für falsch spezifizierte Modelle (4-Faktoren) je Simulationsbedingung

Simulationsszenario	Stichprobengröße	
	n = 150	n = 600
Unkorreliert		
Chi-Quadrat ^a		
Niedrig	307,843 (26,547) [0,237]	405,687 (32,425) [0,000]
Mittel	381,955 (29,989) [0,000]	700,372 (42,076) [0,000]
Hoch	529,918 (32,772) [0,000]	1291,168 (50,489) [0,000]
Gemischt	471,557 (32,155) [0,000]	1057,488 (48,216) [0,000]
SRMR		
Niedrig	0,084 (0,005) [0,938]	0,055 (0,003) [1,000]
Mittel	0,113 (0,007) [0,315]	0,095 (0,004) [1,000]
Hoch	0,169 (0,008) [0,000]	0,159 (0,003) [0,000]
Gemischt	0,128 (0,006) [0,002]	0,112 (0,003) [0,224]
CFI		
Niedrig	0,762 (0,094) [0,031]	0,817 (0,033) [0,000]
Mittel	0,848 (0,031) [0,000]	0,866 (0,011) [0,000]
Hoch	0,890 (0,012) [0,000]	0,896 (0,005) [0,000]
Gemischt	0,805 (0,024) [0,000]	0,818 (0,009) [0,000]
RMSEA		
Niedrig	0,037 (0,010) [0,938]	0,032 (0,003) [1,000]
Mittel	0,058 (0,007) [0,999]	0,054 (0,003) [0,979]
Hoch	0,086 (0,005) [0,115]	0,083 (0,002) [0,000]
Gemischt	0,076 (0,006) [0,654]	0,073 (0,002) [0,000]
Korreliert		
Chi-Quadrat ^b		
Niedrig	308,305 (26,920) [0,206]	418,223 (33,215) [0,000]
Mittel	382,407 (30,005) [0,000]	711,751 (42,552) [0,000]
Hoch	528,423 (32,712) [0,000]	1295,357 (50,510) [0,000]
Gemischt	470,666 (32,257) [0,000]	1065,802 (47,895) [0,000]
SRMR		
Niedrig	0,086 (0,006) [0,960]	0,061 (0,004) [1,000]
Mittel	0,125 (0,011) [0,062]	0,113 (0,006) [0,294]
Hoch	0,196 (0,014) [0,000]	0,191 (0,007) [0,000]
Gemischt	0,143 (0,010) [0,000]	0,132 (0,005) [0,000]
CFI		
Niedrig	0,771 (0,087) [0,026]	0,820 (0,030) [0,000]
Mittel	0,852 (0,029) [0,000]	0,869 (0,011) [0,000]
Hoch	0,892 (0,012) [0,000]	0,898 (0,004) [0,000]
Gemischt	0,812 (0,023) [0,000]	0,824 (0,009) [0,000]
RMSEA		
Niedrig	0,039 (0,010) [0,960]	0,033 (0,003) [1,000]
Mittel	0,059 (0,007) [1,000]	0,056 (0,003) [0,942]
Hoch	0,086 (0,005) [0,090]	0,084 (0,002) [0,000]
Gemischt	0,077 (0,006) [0,638]	0,074 (0,002) [0,000]
Nebenladungen		
Chi-Quadrat ^c	421,308 (35,739) [0,000]	857,727 (56,800) [0,000]
SRMR	0,161 (0,019) [0,001]	0,151 (0,010) [0,000]
CFI	0,871 (0,023) [0,000]	0,883 (0,009) [0,000]
RMSEA	0,067 (0,007) [0,943]	0,063 (0,003) [0,107]

Anmerkung. ^adf = 252. ^bdf = 249. ^cdf = 252.

7.4. Modellvergleich

Um die verschiedenen Modelle nun untereinander zu vergleichen, wurden das Akaike Informationskriterium, der χ^2 -Differenztest und der F-Test herangezogen.

In Tabelle 5 sind für jedes Modell und jede Bedingung Mittelwert und Standardabweichung des AIK eingetragen. Je niedriger der Mittelwert, desto besser das Modell.

Tabelle 5: Akaike Informationskriterium

Mittelwerte und (Standardabweichungen) des AIK je Simulationsbedingung

Simulationsszenario	Modell		
	2-Faktoren	3-Faktoren	4-Faktoren
Unkorreliert			
n = 150			
Niedrig	10093,80 (88,06)	10030,54 (86,29)	10068,20 (87,16)
Mittel	9677,47 (95,15)	9406,60 (86,26)	9517,57 (88,07)
Hoch	8587,90 (115,22)	7736,93 (86,28)	7995,70 (89,24)
Gemischt	9538,44 (99,29)	9131,69 (86,32)	9332,41 (88,92)
n = 600			
Niedrig	40333,28 (174,32)	40052,02 (169,63)	40201,38 (171,33)
Mittel	38690,16 (191,05)	37554,41 (169,65)	37998,36 (173,64)
Hoch	34330,89 (237,08)	30875,18 (169,72)	31909,84 (175,84)
Gemischt	38086,48 (197,684)	36454,37 (169,74)	37255,40 (175,38)
Korreliert			
n = 150			
Niedrig	10048,04 (87,641)	10008,31 (86,24)	10049,77 (86,98)
Mittel	9557,23 (94,86)	9362,67 (86,18)	9477,13 (87,97)
Hoch	8406,84 (116,17)	7677,18 (86,31)	7937,63 (89,07)
Gemischt	9431,24 (98,95)	9082,02 (86,21)	9285,33 (88,80)
n = 600			
Niedrig	40124,18 (173,00)	39954,80 (169,711)	40119,77 (171,68)
Mittel	38170,91 (189,28)	37369,92 (169,73)	37828,21 (173,94)
Hoch	33605,50 (235,48)	30627,79 (169,70)	31669,74 (176,12)
Gemischt	37658,61 (196,36)	36247,14 (169,72)	37059,59 (175,47)
Nebenladungen			
n = 150	9154,79 (93,42)	8995,39 (86,77)	9096,54 (89,90)
n = 600	36549,60 (185,51)	35762,96 (170,05)	36313,64 (177,91)

Wie man erkennen kann, ist der AIK-Wert beim 3-Faktoren-Modell immer am niedrigsten. Das 4-Faktoren-Modell schneidet mit einer Ausnahme (Populationsmodell korreliert, niedrige Ladung, n = 150) immer am zweitbesten ab. Am schlechtesten ist demnach das 2-Faktoren-Modell.

Im Populationsmodell mit Nebenladungen war die Anzahl der Freiheitsgrade für die drei Modelle unterschiedlich. So konnte auch der χ^2 -Differenztest durchgeführt werden. Eine Alternative dazu stellt der F-Test dar. Dieser kann unabhängig von der Anzahl der Freiheitsgrade immer durchgeführt werden. Bei der Berechnung des F-Tests wurde

das 3-Faktoren-Modell als das „Baseline“-Modell herangezogen. Es wurde getestet, ob die anderen Modelle die Daten signifikant schlechter erklären als das 3-Faktoren-Modell. Die entsprechenden Anteile an Verwerfungen finden sich in Tabelle 6.

Tabelle 6: Anteil der Modellverwerfungen beim F-Test und χ^2 -Differenztest

Anteil an Verwerfungen des 2-Faktoren und 4-Faktoren-Modells bei Vergleich mit 3-Faktoren-Modell

Simulations- szenario	F-Test		χ^2 -Differenztest	
	2-Faktoren- Modell	4-Faktoren- Modell	2-Faktoren- Modell	4-Faktoren- Modell
Unkorreliert				
n = 150				
Niedrig	0,536	0,091		
Mittel	1,000	0,996		
Hoch	1,000	1,000		
Gemischt	1,000	1,000		
n = 600				
Niedrig	1,000	1,000		
Mittel	1,000	1,000		
Hoch	1,000	1,000		
Gemischt	1,000	1,000		
Korreliert				
n = 150				
Niedrig	0,123	0,113		
Mittel	1,000	0,997		
Hoch	1,000	1,000		
Gemischt	1,000	1,000		
n = 600				
Niedrig	1,000	1,000		
Mittel	1,000	1,000		
Hoch	1,000	1,000		
Gemischt	1,000	1,000		
Nebenladungen				
n = 150	1,000	1,000	1,000	1,000
n = 600	1,000	1,000	1,000	1,000

Der F-Test verwirft nahezu in jeder Bedingung die fehlspezifizierten 2-Faktoren und 4-Faktoren-Modelle vollständig. Nur bei n = 150 und niedriger Ladung wird vor allem das 4-Faktoren-Modell häufig beibehalten.

Im Populationsmodell mit Nebenladungen verwirft sowohl die F-Statistik als auch der χ^2 -Differenztest alle fehlspezifizierten Modelle beim Vergleich mit dem 3-Faktoren-Modell. Es wurden nur jene 3-Faktoren-Modelle herangezogen, die beim χ^2 -Test angenommen wurden. Die Anteile ändern sich aber kaum, wenn man alle 3-Faktoren-Modelle, also auch jene, die beim χ^2 -Test verworfen wurden, beim Vergleich heranzieht.

Nimmt man beim F-Test das 2-Faktoren oder das 4-Faktoren-Modell als das „Baseline“-Modell an, so beträgt der Anteil an verworfenen 3-Faktoren-Modellen in

jeder Bedingung 0, Es wird also kein richtig spezifiziertes 3-Faktoren-Modell beim Vergleich mit den beiden anderen Modellen verworfen.

8. Diskussion

Auffällig ist, dass der χ^2 -Test bei einer Stichprobengröße von $n = 150$ das nominelle Risiko 1. Art von 5 % klar nicht einhält. Dies liegt unter anderem daran, dass die χ^2 -Statistik auf einer ML-Schätzung basiert. ML-Schätzer haben hervorragende asymptotische Eigenschaften, aber das Verhalten bei kleineren Stichproben kann durchaus problematisch sein. Für ein richtig spezifiziertes Modell entspricht der Erwartungswert der χ^2 -Statistik genau der Anzahl an Freiheitsgraden im Modell. Die Teststatistik ist allerdings erst bei ausreichend großer Stichprobe χ^2 -verteilt (Bollen, 1990). In der vorliegenden Studie konnte dieses Problem quantifiziert werden.

Bei $n = 600$ wird das Risiko 1. Art durch den χ^2 -Test noch immer nicht eingehalten, nähert sich diesem aber an. Es liegt jedoch (immer noch) in jeder Bedingung über 7 %. Der Wert der Statistik liegt also näher am Erwartungswert bzw. näher an der Anzahl an Freiheitsgrade.

Der χ^2 -Test nimmt in der Bedingung mit niedrigen Ladungen und kleiner Stichprobe falsch spezifizierte 2-Faktoren und 4-Faktoren-Modelle an. Bei den 4-Faktoren-Modellen liegt der Anteil an Fehler 2. Art etwas über denen beim 2-Faktoren-Modell.

Das Problem beim χ^2 -Test besteht also in einer Ambivalenz von zu kleiner und zu großer Stichprobengröße. Bei einer zu kleinen Stichprobe ist die Statistik nicht χ^2 -verteilt, bei einer zu großen führt schon die geringste Modellabweichung zu einer Verwerfung des Modells. In der Literatur finden sich Empfehlungen für Stichprobengrößen um $n = 150$ (Muthén & Muthén, 2002) für multivariat normalverteilte Daten und korrekt spezifizierte Modelle. In der vorliegenden Studie führte diese Stichprobengröße nicht zu optimalen Ergebnisse bei korrekt spezifizierten Modellen. Solche Empfehlungen dürfen also keinesfalls unabhängig von der Modellkomplexität (Anzahl manifester Variablen, Anzahl Faktoren) betrachtet werden, da diese Einfluss auf die Modellpassung nimmt (Kenny & McCoach, 2003).

Die beiden Fit-Statistiken SRMR und RMSEA verwerfen bei den verwendeten Cut-Off-Werten nahezu kein richtig spezifiziertes Modell, der Anteil an fälschlicherweise verworfenen 3-Faktoren-Modellen liegt fast bei 0. Somit ist das Risiko 1. Art nahezu 0. Andererseits nehmen beide Indizes sehr viele falsch spezifizierte Modelle an und das Risiko 2. Art ist sehr hoch. Insgesamt verwerfen sie also sehr wenige Modelle. Dies

lässt Zweifel aufkommen, ob die für den SRMR und RMSEA verwendeten Cut-Off-Werte bei vorliegender Modellfehlspezifikation sinnvoll sind. Hu und Bentler (1999) haben bei der Ermittlung der Cut-Offs in ihrer Simulationsstudie andere Fehlspezifikationen angenommen. Dabei kamen keine falschen Faktorenstrukturen vor und sie variierten die Höhe der Faktorladungen nicht. Sie verwendeten nur relativ hohe Faktorladungen zwischen 0,7 und 0,8. Eventuell müsste man die Cut-Off-Werte bei vorliegender Fehlspezifikation also etwas strenger festlegen, sodass insgesamt weniger Modelle angenommen werden. Bei den korrekt spezifizierten Modellen beträgt der maximale Mittelwert für den SRMR 0,072 und für den RMSEA 0,023 für $n = 150$ bzw. 0,007 für $n = 600$ (siehe Tabelle 2). Das heißt, man könnte die Cut-Off-Werte durchaus noch nach unten setzen und würde trotzdem viele Modelle annehmen. Damit würden mehr falsch spezifizierte Modelle verworfen werden.

Der CFI verwirft bei $n = 150$ und niedrigen Ladungen viele richtige Modelle (mehr als 58 %). Sind die Ladungen allerdings höher, ist der Anteil an falsch verworfenen Modellen angemessen. Bei den falsch spezifizierten Modellen zeigt der CFI die besten Eigenschaften. In allen Bedingungen nimmt er nie mehr als 3,4 % an falschen Modellen an. In den meisten Szenarien werden alle falschen Modelle verworfen. Das bedeutet somit, dass der Cut-Off-Wert für den CFI als passend erscheint.

Es scheint so, als ob alle Fit-Statistiken Probleme haben, falsche Faktorenstrukturen zu erkennen, wenn die Faktorladungen niedrig sind. Dies mag vielleicht auf den ersten Blick nicht schlimm erscheinen. Peterson (2000) zeigte jedoch in einer Metaanalyse, dass Faktorladungen bei psychologischen Fragebögen im Mittel bei 0,32 liegen, wobei 50 % der Ladungen zwischen 0,23 und 0,37 liegen. Das heißt, dass die in dieser Simulationsstudie als niedrig bezeichneten Ladungen für die psychologische Forschung als durchaus repräsentativ gelten.

Hu und Bentler (1999) schlugen in ihrer Arbeit eine 2-Index Strategie vor. Damit ist gemeint, dass man die Kombination zweier Indizes zur Überprüfung der Modellgültigkeit heranziehen soll. Dies basiert auf der Annahme, dass bestimmte Indizes unabhängig von der Modellkomplexität sensitiver gegenüber bestimmten Fehlspezifikationen sind, als andere. Fan und Sivo (2009) zeigten allerdings, dass diese Annahme nicht immer erfüllt ist. Empfehlenswerter für die Modellevaluation ist es, zumindest je einen Fit-Index aus jeder der drei Klassen (absolut, relativ, Sparsamkeit-korrigiert) heranzuziehen.

Beim direkten Vergleich verschiedener Modelle stellt der F-Test eine Alternative zum χ^2 -Differenztest dar, da dieser unabhängig von der Anzahl der Freiheitsgrade in den Modellen immer einsetzbar ist. Es muss keine Nestung der Modelle vorliegen. Man kann selbst entscheiden, welches Modell man als „Baseline“-Modell heranzieht, gegen

welches die anderen Modelle getestet werden sollen. Da der χ^2 -Differenztest nur beim Populationsmodell mit Nebenladungen eingesetzt werden konnte, ist es schwierig, Vergleiche zwischen den beiden Methoden anzustellen. In dieser Bedingung gab es jedenfalls keinen Unterschied zwischen den beiden Methoden. Es wurden jeweils alle falsch spezifizierten Modelle verworfen. Das AIK dient nur zum deskriptiven Vergleich. Um zu überprüfen, ob die Anzahl an Wiederholungen je Simulationsbedingung zuverlässige Ergebnisse liefert, wurde eine Bedingung sowohl mit dem Standard von 10.000 als auch mit 100.000 Wiederholungen durchgeführt. Die Ergebnisse wichen kaum voneinander ab. Teilweise gab es ab der dritten Nachkommastelle kleine Unterschiede, die aber vernachlässigbar sind.

9. Zusammenfassung

In dieser Arbeit wurde das Verhalten verschiedener Kriterien zur Modellevaluation bei der konfirmatorischen Faktorenanalyse mittels Simulation untersucht. Dies umfasste die ML-basierte χ^2 -Statistik und drei verschiedene Fit-Indizes. Als Fit-Indizes wurden der SRMR (als Repräsentant der absoluten Fit-Indizes), der CFI (als Vertreter der relativen Fit-Indizes) und der RMSEA (aus der Klasse der Sparsamkeit-korrigierten Indizes) herangezogen, da diese von den meisten Forschern präferiert werden. Es wurde das Verhalten dieser Statistiken bei fehlspezifizierter Faktorenstruktur überprüft. Um das Risiko 1. Art einzuschätzen, wurden auch richtig spezifizierte Modelle getestet. Da mit den Fit-Indizes keine Signifikanztestung möglich ist, wurden die Cut-Off-Werte von Hu und Bentler (1999) verwendet, um entscheiden zu können, ob ein Modell angenommen oder verworfen werden soll. Diese Cut-Off-Werte werden von den meisten Praktikern verwendet.

Zum Vergleich konkurrierender Modelle wurde in der Studie das Verhalten des AIK, des χ^2 -Differenztests und des F-Tests untersucht.

Insgesamt gab es 18 verschiedene Simulationsbedingungen. Da die Höhe der Faktorladungen und die Stichprobengröße einen Einfluss auf die Modellpassung haben, wurden diese variiert (Fan et al., 1999; Heene et al., 2011). Jedes 3-Faktoren Populationsmodell bestand aus 24 manifesten Variablen (acht je Faktor). Aus den Populationsmodellen wurden multivariat normalverteilte Daten generiert.

In jeder Bedingung wurde ein fehlspezifiziertes 2-Faktoren-Modell, ein richtig spezifiziertes 3-Faktoren-Modell und ein fehlspezifiziertes 4-Faktoren-Modell aufgestellt. Um zuverlässige Ergebnisse zu erhalten, wurde jede Simulationsbedingung 10.000-mal durchlaufen.

Um das Risiko 1. Art und das Risiko 2. Art zu schätzen, wurde der Anteil an verworfenen 3-Faktoren-Modellen bzw. der Anteil an angenommenen 2-Faktoren/4-Faktoren-Modellen für die χ^2 -Statistik und die Fit-Indizes in jedem Simulationsszenario ermittelt. Zusätzlich wurden Mittelwerte und Standardabweichungen der Statistiken berechnet.

Beim χ^2 -Test stellt sich heraus, dass er das Risiko 1. Art bei einer Stichprobengröße von $n = 150$ nicht einhält. Es werden bis zu 25,4 % der Modelle verworfen. Bei $n = 600$ nähert sich dieser Wert dem nominellen Risiko von 5 % an. Es wird aber immer noch nicht gänzlich eingehalten. Bei niedrigen Ladungen und $n = 150$ nimmt der χ^2 -Test bis zu 23,7 % an falsch spezifizierten Modellen an.

Der SRMR und der RMSEA verwerfen nahezu kein richtig spezifiziertes Modell, andererseits nehmen sie sehr viele falsch spezifizierte Modelle an. Die Anwendung der Cut-Off-Werte scheint bei vorliegender Fehlspezifikation unpassend und müssten nach unten korrigiert werden.

Den besten Eindruck bei den verwendeten Cut-Off-Werten hinterlässt der CFI. Dieser schneidet nur bei der Untersuchung des Risikos 1. Art in der Bedingung mit niedrigen Ladungen und $n = 150$ schlechter ab als die anderen Statistiken. In den anderen Bedingungen ist er überlegen oder zumindest gleichwertig.

Beim Modellvergleich stellt der F-Test eine Alternative zum χ^2 -Differenztest dar. Er kann auch eingesetzt werden, wenn die Anzahl der Freiheitsgrade in den Modellen identisch ist. In den meisten Bedingungen verwirft der F-Test die fehlspezifizierten 2-Faktoren und 4-Faktoren-Modelle beim Vergleich mit dem 3-Faktoren-Modell. In der Bedingung in der beide zum Einsatz kamen, gab es keinen Unterschied in der Testqualität.

IV. Verzeichnisse

10. Literaturverzeichnis

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika*, 52, 317-322.

Beauducel, A., & Wittmann, W. W. (2005). Simulation study on fit indexes in CFA based on data with slightly distorted simple structure. *Structural Equation Modeling*, 12, 41-75, DOI: 10.1207/s15328007sem1201_3.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107, 238-246.

Bollen, K. A. (1990). Overall fit in covariance structure models: Two types of sample size effects. *Psychological Bulletin*, 107, 256-259.

Brown, T. A. (2006). *Confirmatory Factor Analysis for Applied Research*. New York: Guilford.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, 1, 245-276.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, 4, 272-299.

Fan, X., & Sivo, S. A. (2005). Sensitivity of fit indexes to misspecified structural or measurement model components: Rationale of two-index strategy revisited. *Structural Equation Modeling*, 12, 343-367, DOI: 10.1207/s15328007sem1203_1.

Fan, X., Thompson, B., & Wang, L. (1999). Effects of sample size, estimation methods, and model specification on structural equation modeling fit indexes: A Multidisciplinary Journal. *Structural Equation Modeling*, 6, 56-83, DOI: 10.1080/10705519909540119.

- Heene, M., Hilbert, S., Draxler, C., Ziegler, M., & Bühner, M. (2011). Masking Misfit in Confirmatory Factor Analysis by Increasing Unique Variances: A Cautionary Note on the Usefulness of Cutoff Values of Fit Indices. *Psychological Methods*, 16, 319-336, DOI: 10.1037/a0024917.
- Heene, M., Hilbert, S., Freudenthaler, H. H., & Bühner, M. (2012). Sensitivity of SEM Fit Indexes With Respect to Violations of Uncorrelated Errors. *Structural Equation Modeling: A multidisciplinary Journal*, 19:1, 36-50, DOI: 10.1080/10705511.2012.634710.
- Hooper, D., Coughlan, J., & Mullen, M. (2008). Structural Equation Modelling: Guidelines for Determining Model Fit. *Electronic Journal of Business Research Methods*, 6, 53-60.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1-55, DOI: 10.1080/10705519909540118.
- Jaccard, J., & Wan, C. K. (1996). *LISREL approaches to interaction effects in multiple regression*. Thousand Oaks, CA: Sage.
- Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, 34 (2), 183-202.
- Kenny, D. A., & McCoach, D. B. (2003). Effects of the Number of Variables on Measures of Fit in Structural Equation Modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 10, 333-351, DOI: 10.1207/S15328007SEM1003_1.
- Kubinger, K. D., Litzberger, M., & Mrakotsky, C. (2006). Practised intelligence testing based on a modern test conceptualization and its reference to the common intelligence theories. *Learning and Individual Differences*, 16, 175-193, DOI: 10.1016/j.lindif.2005.08.001.
- McDonald, R. P., & Ho, M.-H. R. (2002). Principles and practice in reporting structural equation analyses. *Psychological Methods*, 7, 64-82, DOI: 10.1037//1082-989X.7.1.64.

- Muthén, L. K., & Muthén, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling*, 4, 599-620.
- Peterson, R. A. (2000). A meta analysis of variance accounted for and factor loadings in exploratory factor analysis. *Marketing Letters*, 11, 261-275.
- R Development Core Team. (2012). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing <http://www.R-project.org/>.
- Rosseel, Y. (2012). lavaan: An R Package for Structural Equation. *Journal of Statistical Software*, 48 (2).
- Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the Fit of Structural Equation Models:. *MPR-online*, 8, 23-74.
- Spearman, C. (1904). General intelligence, objectively determined and measured. *American Journal of Psychology*, 15, 201-293.
- Spearman, C. (1927). *The abilities of man*. New York: Macmillan.
- Steiger, J. H., & Lind, J. M. (1980). Statistically based tests for the number of common factors. *Paper presented at the meeting of the Psychometric Society, Iowa City, IA*.
- Thurstone, L. L. (1947). *Multiple-factor analysis*. Chicago: University of Chicago Press.

11. Tabellenverzeichnis

Tabelle 1: Anteil an ungültigen Lösungen je Bedingung und Modell.....	26
Tabelle 2: Risiko 1. Art bei unterschiedlichen Simulationsbedingungen	28
Tabelle 3: Risiko 2. Art bei unterschiedlichen Simulationsbedingungen für das 2-Faktoren-Modell	30
Tabelle 4: Risiko 2. Art bei unterschiedlichen Simulationsbedingungen für das 4-Faktoren-Modell	32
Tabelle 5: Akaike Informationskriterium	33
Tabelle 6: Anteil der Modellverwerfungen beim F-Test und χ^2 -Differenztest	34

12. Abbildungsverzeichnis

Abbildung 1: Beispiel eines Scree-Plots.....	5
Abbildung 2: Pfaddiagramme der konfirmatorischen und explorativen Faktorenanalyse7	
Abbildung 3: Strukturgleichungsmodell	8
Abbildung 4: Populationsmodell mit niedrigen Faktorladungen	21
Abbildung 5: Populationsmodell mit Nebenladungen	22
Abbildung 6: Fehlspezifizierte Modelle	24

13. R-Code

In den weiteren Kapiteln befindet sich der R-Source Code.

13.1. Verwendetes Paket

```
#Library installieren und laden
install.packages("lavaan")
library("lavaan")
class?lavaan
```

13.2. Populationsmodelle

```
#####
#Definition der Populationsmodelle
#####

#-----
#Unkorrelierte Faktoren
#-----

#Populationsmodell mit niedriger Faktorladung (zwischen 0.3 und 0.5)
#set.seed(7)
#24 gleichverteilte Zufallszahlen zwischen 0.3 und 0.5 für
#Faktorladungen werden erzeugt
#fln <- round(runif(24, 0.3, 0.5), 2)
pop.model.n <- '
f1 =~ 0.5*x1 + 0.38*x2 + 0.32*x3 + 0.31*x4 + 0.35*x5 + 0.46*x6 +
0.37*x7 + 0.49*x8
f2 =~ 0.33*x9 + 0.39*x10 + 0.33*x11 + 0.35*x12 + 0.45*x13 + 0.32*x14 +
0.39*x15 + 0.32*x16
f3 =~ 0.41*x17 + 0.3*x18 + 0.5*x19 + 0.36*x20 + 0.43*x21 + 0.36*x22 +
0.5*x23 + 0.48*x24
'

#Populationsmodell mit mäßiger Faktorladung (zwischen 0.5 und 0.7)
#flm <- fln + 0.2
pop.model.m <- '
f1 =~ 0.7*x1 + 0.58*x2 + 0.52*x3 + 0.51*x4 + 0.55*x5 + 0.66*x6 +
0.57*x7 + 0.69*x8
f2 =~ 0.53*x9 + 0.59*x10 + 0.53*x11 + 0.55*x12 + 0.65*x13 + 0.52*x14 +
0.59*x15 + 0.52*x16
f3 =~ 0.61*x17 + 0.5*x18 + 0.7*x19 + 0.56*x20 + 0.63*x21 + 0.56*x22 +
0.7*x23 + 0.68*x24
'

#Populationsmodell mit hoher Faktorladung (zwischen 0.7 und 0.9)
#flh <- fln + 0.4
pop.model.h <- '
f1 =~ 0.9*x1 + 0.78*x2 + 0.72*x3 + 0.71*x4 + 0.75*x5 + 0.86*x6 +
0.77*x7 + 0.89*x8
f2 =~ 0.73*x9 + 0.79*x10 + 0.73*x11 + 0.75*x12 + 0.85*x13 + 0.72*x14 +
0.79*x15 + 0.72*x16
```

```
f3 =~ 0.81*x17 + 0.7*x18 + 0.9*x19 + 0.76*x20 + 0.83*x21 + 0.76*x22 +
0.9*x23 + 0.88*x24
'
```

```
#Populationsmodell mit gemischter Faktorladung
#(zwischen 0.3 und 0.9)
#set.seed(7)
#24 gleichverteilte Zufallszahlen zwischen 0.3 und 0.9 für
#Faktorladungen werden erzeugt
#flgem <- round(runif(24, 0.3, 0.9), 2)
pop.model.gem <- '
f1 =~ 0.89*x1 + 0.54*x2 + 0.37*x3 + 0.34*x4 + 0.45*x5 + 0.78*x6 +
0.50*x7 + 0.88*x8
f2 =~ 0.40*x9 + 0.58*x10 + 0.40*x11 + 0.44*x12 + 0.76*x13 + 0.36*x14 +
0.57*x15 + 0.35*x16
f3 =~ 0.64*x17 + 0.31*x18 + 0.89*x19 + 0.49*x20 + 0.68*x21 + 0.48*x22
+ 0.90*x23 + 0.84*x24
'
```

```
#-----
#Korrelierte Faktoren
#-----
```

```
#Populationsmodell mit niedriger Faktorladung (zwischen 0.3 und 0.5)
#set.seed(7)
#24 gleichverteilte Zufallszahlen zwischen 0.3 und 0.5 für
#Faktorladungen werden erzeugt
#fln <- round(runif(24, 0.3, 0.5), 2)
pop.model.n.korr <- '
f1 =~ 0.5*x1 + 0.38*x2 + 0.32*x3 + 0.31*x4 + 0.35*x5 + 0.46*x6 +
0.37*x7 + 0.49*x8
f2 =~ 0.33*x9 + 0.39*x10 + 0.33*x11 + 0.35*x12 + 0.45*x13 + 0.32*x14 +
0.39*x15 + 0.32*x16
f3 =~ 0.41*x17 + 0.3*x18 + 0.5*x19 + 0.36*x20 + 0.43*x21 + 0.36*x22 +
0.5*x23 + 0.48*x24
f1 ~~ 0.5*f2
f1 ~~ 0.4*f3
f2 ~~ 0.3*f3
'
```

```
#Populationsmodell mit mäßiger Faktorladung (zwischen 0.5 und 0.7)
#flm <- round(runif(24, 0.5, 0.7), 2)
pop.model.m.korr <- '
f1 =~ 0.7*x1 + 0.58*x2 + 0.52*x3 + 0.51*x4 + 0.55*x5 + 0.66*x6 +
0.57*x7 + 0.69*x8
f2 =~ 0.53*x9 + 0.59*x10 + 0.53*x11 + 0.55*x12 + 0.65*x13 + 0.52*x14 +
0.59*x15 + 0.52*x16
f3 =~ 0.61*x17 + 0.5*x18 + 0.7*x19 + 0.56*x20 + 0.63*x21 + 0.56*x22 +
0.7*x23 + 0.68*x24
f1 ~~ 0.5*f2
f1 ~~ 0.4*f3
f2 ~~ 0.3*f3
'
```

```
#Populationsmodell mit hoher Faktorladung (zwischen 0.7 und 0.9)
#flh <- round(runif(24, 0.7, 0.9), 2)
pop.model.h.korr <- '
f1 =~ 0.9*x1 + 0.78*x2 + 0.72*x3 + 0.71*x4 + 0.75*x5 + 0.86*x6 +
0.77*x7 + 0.89*x8
'
```

```

f2 =~ 0.73*x9 + 0.79*x10 + 0.73*x11 + 0.75*x12 + 0.85*x13 + 0.72*x14 +
0.79*x15 + 0.72*x16
f3 =~ 0.81*x17 + 0.7*x18 + 0.9*x19 + 0.76*x20 + 0.83*x21 + 0.76*x22 +
0.9*x23 + 0.88*x24
f1 ~~ 0.5*f2
f1 ~~ 0.4*f3
f2 ~~ 0.3*f3
'

#Populationsmodell mit gemischter Faktorladung (zwischen 0.3 und 0.9)
#set.seed(7)
#24 gleichverteilte Zufallszahlen zwischen 0.7 und 0.9 für
#Faktorladungen werden erzeugt
#flgem <- round(runif(24, 0.3, 0.9), 2)
pop.model.gem.korr <- '
f1 =~ 0.89*x1 + 0.54*x2 + 0.37*x3 + 0.34*x4 + 0.45*x5 + 0.78*x6 +
0.50*x7 + 0.88*x8
f2 =~ 0.40*x9 + 0.58*x10 + 0.40*x11 + 0.44*x12 + 0.76*x13 + 0.36*x14 +
0.57*x15 + 0.35*x16
f3 =~ 0.64*x17 + 0.31*x18 + 0.89*x19 + 0.49*x20 + 0.68*x21 + 0.48*x22
+ 0.90*x23 + 0.84*x24
f1 ~~ 0.5*f2
f1 ~~ 0.4*f3
f2 ~~ 0.3*f3
'

#-----
#Nebenladungen
#-----

#24 gleichverteilte Zufallszahlen zwischen 0.3 und 0.9 für
#Hauptfaktorladungen werden erzeugt
#set.seed(47)
#flhaupt <- round(runif(24, 0.3, 0.9), 2)

#Nebenladungen
#set.seed(42)
#flneben <- round(runif(48, 0, 0.2), 2)

pop.model.cross.a <- '
f1 =~ 0.89*x1 + 0.52*x2 + 0.76*x3 + 0.79*x4 + 0.64*x5 + 0.71*x6 +
0.53*x7 + 0.58*x8 + 0.18*x9 + 0.19*x10 + 0.06*x11 + 0.17*x12 +
0.13*x13 + 0.1*x14 + 0.15*x15 + 0.03*x16 + 0.13*x17 + 0.14*x18 +
0.09*x19 + 0.14*x20 + 0.19*x21 + 0.05*x22 + 0.09*x23 + 0.19*x24
f2 =~ 0.63*x9 + 0.85*x10 + 0.38*x11 + 0.72*x12 + 0.40*x13 + 0.66*x14 +
0.60*x15 + 0.84*x16 + 0.20*x1 + 0.02*x2 + 0.09*x3 + 0.11*x4 + 0.18*x5
+ 0.03*x6 + 0.20*x7 + 0.19*x8 + 0.02*x17 + 0.10*x18 + 0.08*x19 +
0.18*x20 + 0.09*x21 + 0.17*x22 + 0.15*x23 + 0.16*x24
f3 =~ 0.54*x17 + 0.32*x18 + 0.34*x19 + 0.58*x20 + 0.41*x21 + 0.63*x22
+ 0.61*x23 + 0.38*x24 + 0.08*x1 + 0.14*x2 + 0.00*x3 + 0.17*x4 +
0.00*x5 + 0.04*x6 + 0.18*x7 + 0.12*x8 + 0.08*x9 + 0.09*x10 + 0.01*x11
+ 0.19*x12 + 0.09*x13 + 0.19*x14 + 0.18*x15 + 0.13*x16
'

```

13.3. Aufgestellte Modelle

```

#=====
#Erstellung der zu testenden Modelle
#=====

```

```

#-----
#Unkorreliert
#-----

#Zwei Faktormodell (misspezifiziert, ein Faktor zu wenig, Items
#des Faktor 3 laden auf Faktor 1)
ModelA <- '
f1 =~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x17 + x18 + x19 + x20 +
x21 + x22 + x23 + x24
f2 =~ x9 + x10 + x11 + x12 + x13 + x14 + x15 + x16
f1 ~~ 0*f2'

#Drei Faktorenmodell (wahres Modell)
ModelB <- '
f1 =~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8
f2 =~ x9 + x10 + x11 + x12 + x13 + x14 + x15 + x16
f3 =~ x17 + x18 + x19 + x20 + x21 + x22 + x23 + x24
f1 ~~ 0*f2
f1 ~~ 0*f3
f2 ~~ 0*f3'

#Vier Faktorenmodell (misspezifiziert, ein Faktor zu viel,
#keine crossloadings)
#zwei Faktoren aufgeteilt
ModelC <- '
f1 =~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8
f2 =~ x9 + x10 + x11 + x12 + x13 + x14 + x15 + x16
f3 =~ x17 + x18 + x19 + x20
f4 =~ x21 + x22 + x23 + x24
f1 ~~ 0*f2
f1 ~~ 0*f3
f1 ~~ 0*f4
f2 ~~ 0*f3
f2 ~~ 0*f4
f3 ~~ 0*f4
'

#-----
#Korreliert
#-----

#Zwei Faktormodell (misspezifiziert, ein Faktor zu wenig,
#Items des Faktor 3 laden auf Faktor 1)
ModelA.korr <- '
f1 =~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x17 + x18 + x19 + x20 +
x21 + x22 + x23 + x24
f2 =~ x9 + x10 + x11 + x12 + x13 + x14 + x15 + x16
'

#Drei Faktorenmodell (wahres Modell)
ModelB.korr <- '
f1 =~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8
f2 =~ x9 + x10 + x11 + x12 + x13 + x14 + x15 + x16
f3 =~ x17 + x18 + x19 + x20 + x21 + x22 + x23 + x24
'

#Vier Faktorenmodell (misspezifiziert, ein Faktor zu viel)
#zwei Faktoren aufgeteilt
ModelC0.korr <- '
f1 =~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8

```

```

f2 =~ x9 + x10 + x11 + x12 + x13 + x14 + x15 + x16
f3 =~ x17 + x18 + x19 + x20
f4 =~ x21 + x22 + x23 + x24
f3 =~ 0*f4
'

#-----
#Nebenladungen
#-----

#Zwei Faktormodell (misspezifiziert, ein Faktor zu wenig,
#Items des Faktor 3 laden auf Faktor 1)
ModelA.cross <- '
f1 =~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x17 + x18 + x19 + x20
f2 =~ x9 + x10 + x11 + x12 + x13 + x14 + x15 + x16 + x21 + x22 + x23 +
x24
f1 ~~ 0*f2
'

#Drei Faktorenmodell (wahres Modell)
ModelB.cross <- '
f1 =~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x17 + x18 + x19 + x20
f2 =~ x9 + x10 + x11 + x12 + x13 + x14 + x15 + x16 + x21 + x22 + x23 +
x24
f3 =~ x17 + x18 + x19 + x20 + x21 + x22 + x23 + x24
f1 ~~ 0*f2
f1 ~~ 0*f3
f2 ~~ 0*f3
'

#Vier Faktorenmodell (misspezifiziert, ein Faktor zu viel,
#keine crossloadings)
#zwei Faktoren aufgeteilt
ModelC.cross <- '
f1 =~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8
f2 =~ x9 + x10 + x11 + x12 + x13 + x14 + x15 + x16
f3 =~ x17 + x18 + x19 + x20
f4 =~ x21 + x22 + x23 + x24
f1 ~~ 0*f2
f1 ~~ 0*f3
f1 ~~ 0*f4
f2 ~~ 0*f3
f2 ~~ 0*f4
f3 ~~ 0*f4
'

```

13.4. Funktionen

Funktionen: sim, Statistiken, F.test, Nested, savedata

```

#-----
#Eigentliche Simulation
#Funktion sim
#-----

#Funktion sim zur Durchführung der Simulation

```

```

sim<- function(pop.model, Model.A, Model.B, Model.C, n, N) {

#Variablen initialisieren
  fA <- NULL; fB <- NULL; fC <- NULL
#A <- NULL; B <- NULL; C <- NULL

#Anzahl an gespeicherten Fit-Indizes
  nsp <- 8

  #Startwert (einmal zufällig gewählt) festlegen,
  #um Ergebnisse replizieren zu können
  if(n == 150) set.seed(50) else set.seed(52)

start<- proc.time()
for(i in 1:N) {
#Daten generieren
  #Varianzen der latenten Variablen werden auf 1 gesetzt
  #Bei der Modellspezifikation wurden standardisierte
  #Faktorladungen angegeben, deshalb standardized = T
myData<- simulateData(model = pop.model, model.type = "cfa",
std.lv = T, standardized = T, sample.nobs=n)

  #Modelle fitten, Varianzen der latenten Variablen werden auf 1
  #gesetzt
fitA<- cfa(Model.A, data=myData, std.lv = TRUE)
fitB<- cfa(Model.B, data=myData, std.lv = TRUE)
fitC<- cfa(Model.C, data=myData, std.lv = TRUE)

  #Speicherung bestimmter Statistiken (Fit-Indizes und
  #Freiheitsgrade)
  #Falls ML nicht konvergiert gibt es keine Fit Indices,
  #es werden daher NA's erzeugt
  #Falls improper solutions auftreten erhält man verfälschte
  #Indices, es werden daher NA's erzeugt
if(min(eigen(inspect(fitA, "cov.ov"))$values) > 0 & inspect(fitA,
"converged") & min(eigen(inspect(fitA, "cov.lv"))$values) > 0)
  fA <- rbind(fA, fitMeasures(fitA, c("chisq", "df", "pvalue",
"cfi", "aic", "rmsea", "srmr", "agfi"))) else fA <- rbind(fA,
rep(NA,8))
  if(min(eigen(inspect(fitB, "cov.ov"))$values) > 0 & inspect(fitB,
"converged") & min(eigen(inspect(fitB, "cov.lv"))$values) > 0)
  fB <- rbind(fB, fitMeasures(fitB, c("chisq", "df", "pvalue",
"cfi", "aic", "rmsea", "srmr", "agfi"))) else fB <- rbind(fB,
rep(NA,8))
  if(min(eigen(inspect(fitC, "cov.ov"))$values) > 0 & inspect(fitC,
"converged") & min(eigen(inspect(fitC, "cov.lv"))$values) > 0)
  fC <- rbind(fC, fitMeasures(fitC, c("chisq", "df", "pvalue",
"cfi", "aic", "rmsea", "srmr", "agfi"))) else fC <- rbind(fC,
rep(NA,8))
}
  ende <- proc.time()
  time <- ende-start

  #Die Fit-Indices und Chi^2-Statistik der gesamten berechneten
  #Modelle werden zurückgegeben
  return(list(fA = fA, fB = fB, fC = fC, time = time))
}

```



```

#-----
#Funktion zur Berechnung von Anteilen und anderen Statistiken
#-----

#Funktion statistiken zur Ausgabe der wichtigsten Ergebnissen
statistiken <- function(f.A, f.B, f.C, n, N) {

  #Spaltennamen für Matrizen der Fit-Indizes werden vergeben
  colnames(f.A) <- c("chisq", "df", "p", "cfi", "aic", "rmsea",
"srmr", "agfi")
  colnames(f.B) <- c("chisq", "df", "p", "cfi", "aic", "rmsea",
"srmr", "agfi")
  colnames(f.C) <- c("chisq", "df", "p", "cfi", "aic", "rmsea",
"srmr", "agfi")
  f.A <- round(f.A, 3)
  f.B <- round(f.B, 3)
  f.C <- round(f.C, 3)

  #Anzahl an improper Solutions je Modell
  imprA<- mean(is.na(f.A[,1]))
  imprB<- mean(is.na(f.B[,1]))
  imprC<- mean(is.na(f.C[,1]))
  improper<- cbind(imprA, imprB, imprC)
  rownames(improper) <- "anteil"

  #Anteil der angenommenenen Modelle A, B und C durch Chi^2
  #Teststatistik
  #da manchmal improper solutions auftreten, braucht man diese Abfrage
  A.chi <- ifelse(imprA > 0, sum(f.A[-which(is.na(f.A[,1])),][,3] >
0.05)/N, mean(f.A[,3] > 0.05))
  B.chi <- ifelse(imprB > 0, sum(f.B[-which(is.na(f.B[,1])),][,3] >
0.05)/N, mean(f.B[,3] > 0.05))
  C.chi <- ifelse(imprC > 0, sum(f.C[-which(is.na(f.C[,1])),][,3] >
0.05)/N, mean(f.C[,3] > 0.05))
  chi<- rbind(A.chi, B.chi, C.chi)

  #Anteil der angenommenenen Modelle A, B und C durch RMSEA
  #da manchmal improper solutions auftreten, braucht man diese Abfrage
  ifelse(n < 250, cutoff <- 0.08, cutoff <- 0.06)
  A.rmsea <- ifelse(imprA > 0, sum(f.A[-which(is.na(f.A[,1])),][,6] <
cutoff)/N, mean(f.A[,6] < cutoff))
  B.rmsea <- ifelse(imprB > 0, sum(f.B[-which(is.na(f.B[,1])),][,6] <
cutoff)/N, mean(f.B[,6] < cutoff))
  C.rmsea <- ifelse(imprC > 0, sum(f.C[-which(is.na(f.C[,1])),][,6] <
cutoff)/N, mean(f.C[,6] < cutoff))
  rmsea<- rbind(A.rmsea, B.rmsea, C.rmsea)

  #Anteil der angenommenenen Modelle A, B und C durch SRMR
  #da manchmal improper solutions auftreten, braucht man diese Abfrage
  A.srmr <- ifelse(imprA > 0, sum(f.A[-which(is.na(f.A[,1])),][,7] <
0.11)/N, mean(f.A[,7] < 0.11))
  B.srmr <- ifelse(imprB > 0, sum(f.B[-which(is.na(f.B[,1])),][,7] <
0.11)/N, mean(f.B[,7] < 0.11))
  C.srmr <- ifelse(imprC > 0, sum(f.C[-which(is.na(f.C[,1])),][,7] <
0.11)/N, mean(f.C[,7] < 0.11))
  srmr<- rbind(A.srmr, B.srmr, C.srmr)

  #Anteil der angenommenenen Modelle A, B und C durch CFI
  #da manchmal improper solutions auftreten, braucht man diese Abfrage
  A.cfi <- ifelse(imprA > 0, sum(f.A[-which(is.na(f.A[,1])),][,4] >
0.95)/N, mean(f.A[,4] > 0.95))

```

```

    B.cfi <- ifelse(imprB > 0, sum(f.B[-which(is.na(f.B[,1])),][,4] >
0.95)/N, mean(f.B[,4] > 0.95))
    C.cfi <- ifelse(imprC > 0, sum(f.C[-which(is.na(f.C[,1])),][,4] >
0.95)/N, mean(f.C[,4] > 0.95))
cfi<- rbind(A.cfi, B.cfi, C.cfi)

#Anteil der angenommenenen Modelle A, B und C für alle Indizes
anteil.angenommen <- cbind(chi, rmsea, srmr, cfi)
colnames(anteil.angenommen) <- c("chi", "rmsea", "srmr", "cfi")
rownames(anteil.angenommen) <- c("A", "B", "C")

#Mittelwerte und Standardabweichungen aller Indizes je Modell
if(imprA > 0) mwA <- round(apply(f.A[-which(is.na(f.A)),], 2, mean),3)
else mwA <- round(apply(f.A, 2, mean),3)
    if(imprB > 0) mwB <- round(apply(f.B[-which(is.na(f.B)),], 2,
mean),3) else mwB <- round(apply(f.B, 2, mean),3)
    if(imprC > 0) mwC <- round(apply(f.C[-which(is.na(f.C)),], 2,
mean),3) else mwC <- round(apply(f.C, 2, mean),3)
    if(imprA > 0) sdA <- round(apply(f.A[-which(is.na(f.A)),], 2, sd),3)
else sdA <- round(apply(f.A, 2, sd),3)
    if(imprB > 0) sdB <- round(apply(f.B[-which(is.na(f.B)),], 2, sd),3)
else sdB <- round(apply(f.B, 2, sd),3)
    if(imprC > 0) sdC <- round(apply(f.C[-which(is.na(f.C)),], 2, sd),3)
else sdC <- round(apply(f.C, 2, sd),3)
mittelwert <- rbind(mwA, mwB, mwC)
    standardabweichung <- rbind(sdA, sdB, sdC)

    return(list(Improper = improper, Modelle.angenommen =
anteil.angenommen, Mittelwert = mittelwert, SD = standardabweichung))
}

#Funktion F.test, entspricht F-Test nach Kubinger et al.
#Modell M und True Modell werden der Funktion übergeben
#Nur Vergleich mit Modellen, wo True-Modell angenommen wurde
F.test <- function(f.M, f.T, N) {

imprM<- sum(is.na(f.M[,1]))
imprT<- sum(is.na(f.T[,1]))

chiM<- f.M[,1]; dfM <- f.M[,2];
chiT<- f.T[,1]; dfT <- f.T[,2]; pT <- f.T[,3]

    ohne.pT <- which(pT < 0.05)
    ohne.na <- c(which(is.na(chiM)), which(is.na(chiT)))
ohne <- c(ohne.pT, ohne.na)

#F-Test nach Kubinger für Modellvergleich M vs. T
#Berechnung der F-Statistik
#Vergleich von Modellen ohne NA's und wo p-Wert des True Modells
#größer als 0.05
Ftest <- round((chiM[-ohne]/dfM[-ohne])/(chiT[-ohne]/dfT[-ohne]), 2)
#Berechnung des p-values
p <- 1-pf(Ftest, dfM[-ohne], dfT[-ohne])

#Anteil der Verwerfungen von Modell M bei Vergleich mit True-Modell
imp<- sum(is.na(chiM) & pT > 0.05 & !is.na(pT))
verwerf <- (sum(p < 0.05)+imp)/(N-imprT-sum(pT < 0.05 & !is.na(pT)))
    names(verwerf) <- "Anteil Verwerfungen"
    return(verwerf)
}

```

```

#Funktion nested, Vergleich genesteter Modelle
#Zu welchen Anteil erklärt Parent Modell die Daten signifikant besser
#Wird nur verglichen wenn Parent Modell Daten gut erklärt
nested<- function(f.nest, f.parent, N) {

  imprN<- sum(is.na(f.nest[,1]))
  imprP<- sum(is.na(f.parent[,1]))

  chiN<- f.nest[,1]; dfN <- f.nest[,2];
  chiP<- f.parent[,1]; dfP <- f.parent[,2]; pP <- f.parent[,3]

  chidiff<- chiN-chiP
  dfdiff<- dfN-dfP
  p <- 1-pchisq(chidiff, dfdiff)

  ohne<- c(which(pP < 0.05), which(is.na(chiP)))
  p <- p[-ohne]

  imprnurN <- sum(is.na(p))

  #Anteil an signifikant besseren Parent-Modellen
  parentbesser <- (sum(p<0.05 & !is.na(p))+imprnurN)/(N-length(ohne))
  return(parentbesser)
}

#Funktion savedata zur Speicherung der Objekte
savedata<- function(pop.model, n, N, path, Name) {
  dir.create(Name)
  if(n == 150) set.seed(50) else set.seed(52)

  start<- proc.time()
  for(i in 1:N) {
  #Daten generieren
    #Varianzen der latenten Variablen werden auf 1 gesetzt
    #Bei der Modellspezifikation wurden standardisierte
    #Faktorladungen angegeben, deshalb standardized = T
    #myData hat dann ungefähr 20 Gigabyte (inklusive Simulation der
    #100000 Durchgänge)
  myData<- simulateData(model = pop.model, model.type = "cfa", std.lv =
  T, standardized = T, sample.nobs=n)
  save(myData, file=paste(path,Name,"_",i, sep=""))
  }
  ende<- proc.time()
  time<- ende-start
  return(time)
}

```

13.5. Funktionsaufrufe

```

#=====
#Funktionsaufrufe
#=====

#Working Directory festlegen

setwd("./writeable")
getwd()

```

```

#Erstellung der Verzeichnisse
dir.create("Daten")
setwd("./Daten")

dir.create("n150")
dir.create("m150")
dir.create("h150")
dir.create("gem150")
dir.create("n600")
dir.create("m600")
dir.create("h600")
dir.create("gem600")

dir.create("n150korr")
dir.create("m150korr")
dir.create("h150korr")
dir.create("gem150korr")
dir.create("n600korr")
dir.create("m600korr")
dir.create("h600korr")
dir.create("gem600korr")

dir.create("150across")
dir.create("600across")

dir.create("n150.100k")

#Anzahl Wiederholungen/Durchläufe
N <- 10000
#Stichprobengröße
n <- 150L

#Speicherung der Objekte
pfad<- "./n150/"; name <- "n150"; savedata(pop.model.n, n, N, pfad,
name)
pfad<- "./m150/"; name <- "m150"; savedata(pop.model.m, n, N, pfad,
name)
pfad<- "./h150/"; name <- "h150"; savedata(pop.model.h, n, N, pfad,
name)
pfad<- "./gem150/"; name <- "gem150"; savedata(pop.model.gem, n, N,
pfad, name)
pfad<- "./n150korr/"; name <- "n150korr"; savedata(pop.model.n.korr,
n, N, pfad, name)
pfad <- "./m150korr/"; name <- "m150korr"; savedata(pop.model.m.korr,
n, N, pfad, name)
pfad <- "./h150korr/"; name <- "h150korr"; savedata(pop.model.h.korr,
n, N, pfad, name)
pfad <-      "./gem150korr/";      name <-      "gem150korr";
savedata(pop.model.gem.korr, n, N, pfad, name)
pfad<-      "./150crossall/";      name <-      "150crossall";
savedata(pop.model.cross.a, n, N, pfad, name)

n <- 600L
pfad<- "./n600/"; name <- "n600"; savedata(pop.model.n, n, N, pfad,
name)
pfad<- "./m600/"; name <- "m600"; savedata(pop.model.m, n, N, pfad,
name)
pfad<- "./h600/"; name <- "h600"; savedata(pop.model.h, n, N, pfad,
name)

```

```

pfad<- "./gem600/"; name <- "gem600"; savedata(pop.model.gem, n, N,
pfad, name)
pfad<- "./n600korr/"; name <- "n600korr"; savedata(pop.model.n.korr,
n, N, pfad, name)
pfad <- "./m600korr/"; name <- "m600korr"; savedata(pop.model.m.korr,
n, N, pfad, name)
pfad <- "./h600korr/"; name <- "h600korr"; savedata(pop.model.h.korr,
n, N, pfad, name)
pfad <-      "./gem600korr/";      name <-      "gem600korr";
savedata(pop.model.gem.korr, n, N, pfad, name)
pfad<-      "./600crossall/";      name <-      "600crossall";
savedata(pop.model.cross.a, n, N, pfad, name)

#Einmal mit 100000 Wiederholungen
N <- 100000; n <- 150; pfad <- "./n150.100k/"; name <- "n150.100k";
savedata(pop.model.n, n, N, pfad, name)

#Anzahl Wiederholungen/Durchläufe
N <- 10000
#Stichprobengröße
n <- 150L

#Durchführung der Simulation und speichern der Ergebnisse im Objekt
#erg.{Bedingung} für n = 150
#Unkorreliert
erg.n150 <- sim(pop.model.n, ModelA, ModelB, ModelC, n, N)
erg.m150 <- sim(pop.model.m, ModelA, ModelB, ModelC, n, N)
erg.h150 <- sim(pop.model.h, ModelA, ModelB, ModelC, n, N)
erg.gem150 <- sim(pop.model.gem, ModelA, ModelB, ModelC, n, N)

#Korreliert
erg.n150.korr <- sim(pop.model.n.korr, ModelA.korr, ModelB.korr,
ModelC0.korr, n, N)
erg.m150.korr <- sim(pop.model.m.korr, ModelA.korr, ModelB.korr,
ModelC0.korr, n, N)
erg.h150.korr <- sim(pop.model.h.korr, ModelA.korr, ModelB.korr,
ModelC0.korr, n, N)
erg.gem150.korr <- sim(pop.model.gem.korr, ModelA.korr, ModelB.korr,
ModelC0.korr, n, N)

#Nebenladungen
erg.150.cross.a <- sim(pop.model.cross.a, ModelA.cross.a,
ModelB.cross.a, ModelC.cross.a, n, N)

#Durchführung der Simulation und speichern der Ergebnisse im Objekt
#erg.{Bedingung} für n = 600
n <- 600L
#Unkorreliert
erg.n600 <- sim(pop.model.n, ModelA, ModelB, ModelC, n, N)
erg.m600 <- sim(pop.model.m, ModelA, ModelB, ModelC, n, N)
erg.h600 <- sim(pop.model.h, ModelA, ModelB, ModelC, n, N)
erg.gem600 <- sim(pop.model.gem, ModelA, ModelB, ModelC, n, N)

#Korreliert
erg.n600.korr <- sim(pop.model.n.korr, ModelA.korr, ModelB.korr,
ModelC0.korr, n, N)
erg.m600.korr <- sim(pop.model.m.korr, ModelA.korr, ModelB.korr,
ModelC0.korr, n, N)

```

```

erg.h600.korr <- sim(pop.model.h.korr, ModelA.korr, ModelB.korr,
ModelC0.korr, n, N)
erg.gem600.korr <- sim(pop.model.gem.korr, ModelA.korr, ModelB.korr,
ModelC0.korr, n, N)

#Nebenladungen
erg.600.cross.a <- sim(pop.model.cross.a, ModelA.cross.a,
ModelB.cross.a, ModelC.cross.a, n, N)

#Aufruf F-Test nach Kubinger
#Unkorreliert
F.n150 <- cbind(F.test(erg.n150$fA, erg.n150$fB, N),
F.test(erg.n150$fC, erg.n150$fB, N))
colnames(F.n150) <- c("A", "C"); rownames(F.n150) <- c("B angenommen")
F.m150 <- cbind(F.test(erg.m150$fA, erg.m150$fB, N),
F.test(erg.m150$fC, erg.m150$fB, N))
colnames(F.m150) <- c("A", "C"); rownames(F.m150) <- c("B angenommen")
F.h150 <- cbind(F.test(erg.h150$fA, erg.h150$fB, N),
F.test(erg.h150$fC, erg.h150$fB, N))
colnames(F.h150) <- c("A", "C"); rownames(F.h150) <- c("B angenommen")
F.gem150 <- cbind(F.test(erg.gem150$fA, erg.gem150$fB, N),
F.test(erg.gem150$fC, erg.gem150$fB, N))
colnames(F.gem150) <- c("A", "C"); rownames(F.gem150) <- c("B
angenommen")

F.n600 <- cbind(F.test(erg.n600$fA, erg.n600$fB, N),
F.test(erg.n600$fC, erg.n600$fB, N))
colnames(F.n600) <- c("A", "C"); rownames(F.n600) <- c("B angenommen")
F.m600 <- cbind(F.test(erg.m600$fA, erg.m600$fB, N),
F.test(erg.m600$fC, erg.m600$fB, N))
colnames(F.m600) <- c("A", "C"); rownames(F.m600) <- c("B angenommen")
F.h600 <- cbind(F.test(erg.h600$fA, erg.h600$fB, N),
F.test(erg.h600$fC, erg.h600$fB, N))
colnames(F.h600) <- c("A", "C"); rownames(F.h600) <- c("B angenommen")
F.gem600 <- cbind(F.test(erg.gem600$fA, erg.gem600$fB, N),
F.test(erg.gem600$fC, erg.gem600$fB, N))
colnames(F.gem600) <- c("A", "C"); rownames(F.gem600) <- c("B
angenommen")

#Korreliert
F.n150.korr <- cbind(F.test(erg.n150.korr$fA, erg.n150.korr$fB, N),
F.test(erg.n150.korr$fC0, erg.n150.korr$fB, N))
colnames(F.n150.korr) <- c("A", "C0"); rownames(F.n150.korr) <- c("B
angenommen")
F.m150.korr <- cbind(F.test(erg.m150.korr$fA, erg.m150.korr$fB, N),
F.test(erg.m150.korr$fC0, erg.m150.korr$fB, N))
colnames(F.m150.korr) <- c("A", "C0"); rownames(F.m150.korr) <- c("B
angenommen")
F.h150.korr <- cbind(F.test(erg.h150.korr$fA, erg.h150.korr$fB, N),
F.test(erg.h150.korr$fC0, erg.h150.korr$fB, N))
colnames(F.h150.korr) <- c("A", "C0"); rownames(F.h150.korr) <- c("B
angenommen")
F.gem150.korr <- cbind(F.test(erg.gem150.korr$fA, erg.gem150.korr$fB,
N), F.test(erg.gem150.korr$fC0, erg.gem150.korr$fB, N))
colnames(F.gem150.korr) <- c("A", "C0"); rownames(F.gem150.korr) <-
c("B angenommen")

F.n600.korr <- cbind(F.test(erg.n600.korr$fA, erg.n600.korr$fB, N),
F.test(erg.n600.korr$fC0, erg.n600.korr$fB, N))

```

```

colnames(F.n600.korr) <- c("A", "C0"); rownames(F.n600.korr) <- c("B
angenommen")
F.m600.korr <- cbind(F.test(erg.m600.korr$fA, erg.m600.korr$fB, N),
F.test(erg.m600.korr$fC0, erg.m600.korr$fB, N))
colnames(F.m600.korr) <- c("A", "C0"); rownames(F.m600.korr) <- c("B
angenommen")
F.h600.korr <- cbind(F.test(erg.h600.korr$fA, erg.h600.korr$fB, N),
F.test(erg.h600.korr$fC0, erg.h600.korr$fB, N))
colnames(F.h600.korr) <- c("A", "C0"); rownames(F.h600.korr) <- c("B
angenommen")
F.gem600.korr <- cbind(F.test(erg.gem600.korr$fA, erg.gem600.korr$fB,
N), F.test(erg.gem600.korr$fC0, erg.gem600.korr$fB, N))
colnames(F.gem600.korr) <- c("A", "C0"); rownames(F.gem600.korr) <-
c("B angenommen")

#Nebenladungen
F.150.cross.a <- cbind(F.test(erg.150.cross.a$fA, erg.150.cross.a$fB,
N), F.test(erg.150.cross.a$fC, erg.150.cross.a$fB, N))
colnames(F.150.cross.a) <- c("A", "C"); rownames(F.150.cross.a) <-
c("B angenommen")
F.600.cross.a <- cbind(F.test(erg.600.cross.a$fA, erg.600.cross.a$fB,
N), F.test(erg.600.cross.a$fC, erg.600.cross.a$fB, N))
colnames(F.600.cross.a) <- c("A", "C"); rownames(F.600.cross.a) <-
c("B angenommen")

#Aufruf genestete Modelle
N.150.cross.a <-cbind(nested(erg.150.cross.a$fA, erg.150.cross.a$fB,
N), nested(erg.150.cross.a$fC, erg.150.cross.a$fB, N))
colnames(N.150.cross.a) <- c("Bvs.A", "Cvs.B");
rownames(N.150.cross.a) <- c("Parent angenommen")
N.600.cross.a <- cbind(nested(erg.600.cross.a$fA, erg.600.cross.a$fB,
N), nested(erg.600.cross.a$fC, erg.600.cross.a$fB, N))
colnames(N.600.cross.a) <- c("Bvs.A", "Cvs.B");
rownames(N.600.cross.a) <- c("Parent angenommen")

n <- 150
#Berechnung von Fit-Statistiken und Ausgabe für die verschiedenen
#Modelle
#n=150, FL = niedrig, Korr = nein, Variablen = 24
erg.n150$time
statistiken(erg.n150$fA, erg.n150$fB, erg.n150$fC, n, N)
round(F.n150, 3)
#Anteil an verworfenen Modellen bei Vergleich mit True-Modell
#n=150, FL = mäßig, Korr = nein, Variablen = 24
erg.m150$time
statistiken(erg.m150$fA, erg.m150$fB, erg.m150$fC, n, N)
round(F.m150, 3)
#Anteil an verworfenen Modellen bei Vergleich mit True-Modell
#n=150, FL = hoch, Korr = nein, Variablen = 24
erg.h150$time
statistiken(erg.h150$fA, erg.h150$fB, erg.h150$fC, n, N)
round(F.h150, 3)
#Anteil an verworfenen Modellen bei Vergleich mit True-Modell
#n=150, FL = gem, Korr = nein, Variablen = 24
erg.gem150$time
statistiken(erg.gem150$fA, erg.gem150$fB, erg.gem150$fC, n, N)
round(F.gem150, 3)

```

```

#Anteil an verworfenen Modellen bei Vergleich mit True-Modell

#n=150, FL = niedrig, Korr = ja, Variablen = 24
erg.n150.korr$time
statistiken(erg.n150.korr$fA, erg.n150.korr$fB, erg.n150.korr$fC0, n,
N)
round(F.n150.korr, 3)
#Anteil an verworfenen Modellen bei Vergleich mit True-Modell
#n=150, FL = mäßig, Korr = ja, Variablen = 24
erg.m150.korr$time
statistiken(erg.m150.korr$fA, erg.m150.korr$fB, erg.m150.korr$fC0, n,
N)
round(F.m150.korr, 3)
#Anteil an verworfenen Modellen bei Vergleich mit True-Modell
#n=150, FL = hoch, Korr = ja, Variablen = 24
erg.h150.korr$time
statistiken(erg.h150.korr$fA, erg.h150.korr$fB, erg.h150.korr$fC0, n,
N)
round(F.h150.korr, 3)
#Anteil an verworfenen Modellen bei Vergleich mit True-Modell
#n=150, FL = gemischt, Korr = ja, Variablen = 24
erg.gem150.korr$time
statistiken(erg.gem150.korr$fA, erg.gem150.korr$fB,
erg.gem150.korr$fC0, n, N)
round(F.gem150.korr, 3)
#Anteil an verworfenen Modellen bei Vergleich mit True-Modell

#n=150, FL = cross.all, Korr = nein, Variablen = 24
erg.150.cross.a$time
statistiken(erg.150.cross.a$fA, erg.150.cross.a$fB,
erg.150.cross.a$fC, n, N)
round(F.150.cross.a, 3)
#Anteil an verworfenen Modellen bei Vergleich mit True-Modell
round(N.150.cross.a, 3)
#Anteil der Parent Modelle, die Daten signifikant besser erklären

n <- 600
#Berechnung von Fit-Statistiken und Ausgabe für die verschiedenen
#Modelle
#n=600, FL = niedrig, Korr = nein, Variablen = 24
erg.n600$time
statistiken(erg.n600$fA, erg.n600$fB, erg.n600$fC, n, N)
round(F.n600, 3)
#Anteil an verworfenen Modellen bei Vergleich mit True-Modell
#n=600, FL = mäßig, Korr = nein, Variablen = 24
erg.m600$time
statistiken(erg.m600$fA, erg.m600$fB, erg.m600$fC, n, N)
round(F.m600, 3)
#Anteil an verworfenen Modellen bei Vergleich mit True-Modell
#n=600, FL = hoch, Korr = nein, Variablen = 24
erg.h600$time
statistiken(erg.h600$fA, erg.h600$fB, erg.h600$fC, n, N)
round(F.h600, 3)
#Anteil an verworfenen Modellen bei Vergleich mit True-Modell
#n=600, FL = gem, Korr = nein, Variablen = 24
erg.gem600$time
statistiken(erg.gem600$fA, erg.gem600$fB, erg.gem600$fC, n, N)
round(F.gem600, 3)
#Anteil an verworfenen Modellen bei Vergleich mit True-Modell

```



```

#n=600, FL = niedrig, Korr = ja, Variablen = 24
erg.n600.korr$time
statistiken(erg.n600.korr$fA, erg.n600.korr$fB, erg.n600.korr$fC0, n,
N)
round(F.n600.korr, 3)
#Anteil an verworfenen Modellen bei Vergleich mit True-Modell
#n=600, FL = mäßig, Korr = ja, Variablen = 24
erg.m600.korr$time
statistiken(erg.m600.korr$fA, erg.m600.korr$fB, erg.m600.korr$fC0, n,
N)
round(F.m600.korr, 3)
#Anteil an verworfenen Modellen bei Vergleich mit True-Modell
#n=600, FL = hoch, Korr = ja, Variablen = 24
erg.h600.korr$time
statistiken(erg.h600.korr$fA, erg.h600.korr$fB, erg.h600.korr$fC0, n,
N)
round(F.h600.korr, 3)          #Anteil an verworfenen Modellen bei
Vergleich mit True-Modell
#n=600, FL = gemischt, Korr = ja, Variablen = 24
erg.gem600.korr$time
statistiken(erg.gem600.korr$fA,erg.gem600.korr$fB, erg.gem600.korr$fC,
erg.gem600.korr$fC0, erg.gem600.korr$fC5, n, N)
round(F.gem600.korr, 3)
#Anteil an verworfenen Modellen bei Vergleich mit True-Modell

#n=600, FL = cross.all, Korr = nein, Variablen = 24
erg.600.cross.a$time
statistiken(erg.600.cross.a$fA,erg.600.cross.a$fB, erg.600.cross.a$fC,
n, N)
round(F.600.cross.a, 3)
#Anteil an verworfenen Modellen bei Vergleich mit True-Modell
round(N.600.cross.a, 3)
#Anteil der Parent Modelle, die Daten signifikant besser erklären

#MIT 100000 Durchläufen
#Anzahl Wiederholungen/Durchläufe und Stichprobengröße
N <- 100000; n <- 150L
#Durchführung der Simulation und speichern der Ergebnisse im Objekt
#erg.{Bedingung} für n = 150
erg.n150.100k <- sim(pop.model.n, ModelA, ModelB, ModelC, n, N)
F.n150.100k <- cbind(F.test(erg.n150.100k$fA, erg.n150.100k$fB, N),
F.test(erg.n150.100k$fC, erg.n150.100k$fB, N))
colnames(F.n150.100k) <- c("A", "C"); rownames(F.n150.100k) <- c("B
angenommen")
erg.n150.100k$time
statistiken(erg.n150.100k$fA, erg.n150.100k$fB, erg.n150.100k$fC, n,
N)
round(F.n150.100k, 3)

#Speicherung der Ergebnisobjekte
getwd()
dir.create("Ergebnisse")
setwd("./Ergebnisse")

save(erg.n150, file="erg.n150")
save(erg.m150, file="erg.m150")
save(erg.h150, file="erg.h150")
save(erg.gem150, file="erg.gem150")

```

```
save(erg.n600, file="erg.n600")
save(erg.m600, file="erg.m600")
save(erg.h600, file="erg.h600")
save(erg.gem600, file="erg.gem600")

save(erg.n150.korr, file="erg.n150.korr")
save(erg.m150.korr, file="erg.m150.korr")
save(erg.h150.korr, file="erg.h150.korr")
save(erg.gem150.korr, file="erg.gem150.korr")
save(erg.n600.korr, file="erg.n600.korr")
save(erg.m600.korr, file="erg.m600.korr")
save(erg.h600.korr, file="erg.h600.korr")
save(erg.gem600.korr, file="erg.gem600.korr")

save(erg.150.cross.a, file="erg.150.cross.a")
save(erg.600.cross.a, file="erg.600.cross.a")

save(erg.n150.100k, file="erg.n150.100k")
setwd("../")
```


14. Lebenslauf

Michael Themessl-Huber

geboren am 18.12.1986 in Villach

E-Mail: michi.themessl@gmail.com

(Akademische) Ausbildung:

Seit Okt. 2011	Studium der Statistik an der Universität Wien
Seit Dez. 2009	Zweiter Studienabschnitt Diplomstudium Psychologie
Okt. 2007 – Dez. 2009	Erster Studienabschnitt Diplomstudium Psychologie
Okt. 2006	Reifeprüfung an der Höheren Technischen Lehranstalt für EDV-O in Villach

Arbeitserfahrung:

Seit Jan. 2014	Projektmitarbeiter am Institut für medizinische Statistik an der Medizinischen Universität Wien
Seit Okt. 2012	Tutor für das Fach „Lineare Algebra“ an der Universität Wien
Sep. 2012-Okt. 2012	Praktikum im Bereich statistisches Consulting bei SAS
Jul. 2011 – Aug. 2011	Pflichtpraktikum (240h) im neuropsychologischen Bereich in der Gailtal-Klinik Hermagor