

Strukturgleichungsmodelle

Rolf Steyer, Andreas Wolf, Friedrich Funke, Ivailo Partchev

Institut für Psychologie der FSU Jena
Lehrstuhl Methodenlehre und Evaluationsforschung

1. Einleitung

Strukturgleichungsmodelle haben in den letzten Jahrzehnten enorm an Bedeutung gewonnen. Mittlerweile werden sie genauso häufig zitiert wie die Varianzanalyse, mit immer noch steigender Tendenz (vgl. Nachtigall, Kroehne, Funke, & Steyer, 2003). Ihre Attraktivität rührt vor allem

- (a) von ihrer großen Flexibilität, mit der sie
- (b) Beziehungen zwischen latenten Variablen (d. h. psychologischen Konstrukten) beschreiben können,
- (c) deren Messmodelle explizite Bestandteile des Modells sind,
- (d) das mit einer Vielzahl von deskriptiven und inferentiellen Verfahren an empirischen Daten geprüft werden kann.

Mit Punkt (a) ist gemeint, dass viele Annahmen und Voraussetzungen herkömmlicher Verfahren wie z. B. Varianzhomogenität bei der Varianzanalyse, Gleichheit der Kovarianzmatrizen in der multivariaten Varianzanalyse, oder auch Anforderungen an das Skalenniveau der analysierten Variablen entfallen bzw. gelockert werden können. So kann man Strukturgleichungsmodelle auch für dichotome und für ordinale Variablen rechnen, wie sie etwa in Fragebogenuntersuchungen (z. B. als Ratingskalen) häufig vorliegen.

Beziehungen zwischen latenten Variablen (b) beschreiben zu können, ist wohl die wichtigste Eigenschaft von Strukturgleichungsmodellen. Während man mit herkömmlichen Verfahren nur Populationskennwerte wie z. B. Erwartungswerte, Varianzen und Korrelationen von Populationen (d. h. Personengruppen) untersuchen kann, kommen in Strukturgleichungsmodellen auch latente Variablen vor, deren Werte einzelne Individuen in der betrachteten Population charakterisieren. Damit kann man mit diesen Modellen differentialpsychologische Konstrukte direkt modellieren. Aber auch für die Allgemeine Psychologie ergeben sich mit diesen Modellen ganz neue Möglichkeiten, wenn man unter Allgemeiner Psychologie versteht, dass ihre Aussagen sich auf *alle* Individuen und nicht nur auf Gruppennittelwerte beziehen (s. Steyer, 2005.)

Auch der o. g. Punkt (c) ist von großer Bedeutung, da ein Konstrukt letztlich erst innerhalb eines Messmodells präzise genug formuliert wird, um es einer empirischen Prüfung zugänglich zu machen. Angesichts der Vielfalt der Einflüsse auf Messungen (Messfehler, situativ bedingte Effekte wie Tagesform und zircadiane Rhythmen, Antwortstile, soziale Erwünschtheit etc.) müssen die betreffenden Messmodelle oft selbst im Zentrum der Aufmerksamkeit stehen, um die betreffenden Einflüsse zu kontrollieren und auszufiltern.

Schließlich sind auch die unter (d) genannten Verfahren der empirischen Modellprüfung von großer Bedeutung für eine empirisch orientierte Psychologie, da sie viele wertvolle Möglichkeiten bieten, aus Daten Einsichten über die Gültigkeit bzw. die Modifizierung konkreter Modelle zu gewinnen.

In diesem Artikel werden wir Strukturgleichungsmodelle einführen, indem wir mit einem sehr einfachen Pfadmodell zwischen drei manifesten Variablen beginnen. Danach führen wir in einem simplen Einfaktormodell eine latente Variable ein. Dieses Modell wird dann Schritt für Schritt modifiziert, um dann schließlich Strukturgleichungsmodelle in ihrer allgemeinen Form darzustellen.

Wir beginnen mit *Modellen der Klassischen Testtheorie*, bei denen es um die Kontrolle von Messfehlern geht. Danach kommen wir zur Erweiterung dieser Modelle zu *Latent-state-trait-Modellen*, mit denen man neben Messfehlern auch situative Effekte und Effekte der Interaktion zwischen Personen und Situationen kontrollieren und ausfiltern kann. Aufbauend

auf den Latent-state-trait-Modellen können wir dann *Latent-change-Modelle* mit latenten Veränderungsvariablen und *Wachstumskurvenmodelle* behandeln. Die Latent-change-Modelle sind wiederum Grundlage für die Einführung der *Individual-causal-effect-Modelle*, in denen latente Variablen eingeführt sind, deren Werte die individuellen kausalen Effekte eines Treatments sind. Als letzte exemplarisch dargestellte Klasse werden dann *Multitrait-Multimethod-Modelle* behandelt. Danach kommen wir zur Darstellung der allgemeinen Form von *faktorenanalytischen Modellen* und *Strukturgleichungsmodellen* und schließlich zur Behandlung der Schätz- und Testverfahren.

2. Pfadmodelle mit Observablen

Viele Modelle in der Statistik gehen von einer einzigen (univariate Varianz- und Regressionsanalysen) oder mehreren *abhängigen Variablen* (multivariate Varianz-, Regressionsanalysen, Faktoranalysen) aus, deren Abhängigkeit von einer oder mehreren *unabhängigen Variablen* untersucht wird. Dies ist eine relative statische Betrachtungsweise, kann man sich doch auch vorstellen, dass eine Variable sowohl „unabhängig“ in dem Sinn ist, dass sie eine oder mehrere andere erklärt, als auch „abhängig“, d. h. selbst von einer oder mehreren anderen erklärt wird. Ein typisches Beispiel ist das in Abbildung 1 dargestellte *Mediatormodell* (s. z. B. Baron & Kenny, 1986). In diesem Beispiel erklärt die Variable X sowohl die Variable Y , als auch die Variable Z , die ihrerseits aber auch die Variable Y erklärt. Typische Fragen bei dieser Art von Modellen ist, ob es einen Mediationseffekt gibt, ob also ein Teil der Abhängigkeit der Variablen Y (z. B. Leistung) von der Variablen X (z. B. Treatment) durch Z (z. B. Motivation) vermittelt wird, oder ob es überhaupt einen direkten Effekt der Variablen X auf Y gibt, der also nicht durch Z vermittelt wird. Generell wird es damit, wenn auch in einer relativ rudimentären Form, möglich, *Prozesse* darzustellen. Effekte von Variablen können differenzierter betrachtet werden, in dem wir bspw. einen totalen Effekt als Summe von direkten und indirekten Effekten auffassen können.

 Hier Abbildung 1 einfügen

Das in Abbildung 1 dargestellte Pfaddiagramm repräsentiert die folgenden beiden Regressionsgleichungen:

$$E(Z | X) = \gamma_0 + \gamma_1 X \quad [2.1]$$

$$E(Y | X, Z) = \beta_0 + \beta_1 X + \beta_2 Z. \quad [2.2]$$

Dabei bezeichnen γ_1 den *direkten Effekt* von X auf Z , β_1 den *direkten Effekt* von X auf Y und β_2 den *direkten Effekt* von Z auf Y . Nun kann man sich fragen, wie groß der *gesamte Effekt* von X auf Y ist, der sich aus dem direkten Effekt und dem indirekten, über Z vermittelten Effekt zusammensetzen muss. Wie hängt also Y insgesamt von X ab? Die Antwort dazu erhält man, wenn man die Regression von Y auf X betrachtet, die man sich, unter Verwendung der Rechenregeln für Regressionen (s. z. B. Steyer, 2003, Regelbox 6.2) leicht aus den oben aufgeführten Gleichungen herleiten kann.

$$\begin{aligned}
E(Y|X) &= E[E(Y|X, Z)|X] \\
&= E[\beta_0 + \beta_1 X + \beta_2 Z|X] \\
&= E(\beta_0|X) + E(\beta_1 X|X) + E(\beta_2 Z|X) & [2.3] \\
&= \beta_0 + \beta_1 X + \beta_2 E(Z|X) \\
&= \beta_0 + \beta_1 X + \beta_2 (\gamma_0 + \gamma_1 X) \\
&= (\beta_0 + \beta_2 \gamma_0) + (\beta_1 + \beta_2 \gamma_1) X \\
&= \alpha_0 + \alpha_1 X,
\end{aligned}$$

mit dem Achsenabschnitt $\alpha_0 := \beta_0 + \beta_2 \gamma_0$ und dem Steigungskoeffizienten $\alpha_1 := \beta_1 + \beta_2 \gamma_1$. Falls die Gültigkeit der Gleichungen [2.1] und [2.2] vorausgesetzt werden kann, ist die Regression $E(Y|X)$ also wiederum eine lineare Funktion von X mit dem Steigungskoeffizienten $\alpha_1 = \beta_1 + \beta_2 \gamma_1$, den wir als *totalen Effekt* interpretieren können, wobei β_1 der *direkte Effekt* von X auf Y ist und $\beta_2 \gamma_1$ der *indirekte Effekt* von X auf Y , der durch Z vermittelt wird.

Typische Hypothesen in diesem Kontext sind also, dass der indirekte Effekt $\beta_2 \gamma_1 = 0$ ist, d. h. es liegt kein Mediatoreffekt vor, dass der direkte Effekt $\beta_1 = 0$ ist, dass also die gesamte regressive Abhängigkeit durch den Mediator Z vermittelt wird und dass der totale Effekt von X auf Y gleich null ist: $\beta_1 + \beta_2 \gamma_1 = 0$.

Die in den Gleichungen [2.1] und [2.2] vorkommenden Regressionsparameter könnte man im Prinzip mit der zweimaligen Anwendung eines herkömmlichen Programms zur multiplen linearen Regression schätzen. Mit diesem Verfahren könnte man auch die Hypothese $\beta_1 = 0$ (kein direkter Effekt) testen, nicht aber die Hypothese $\beta_2 \gamma_1 = 0$ (kein Mediatoreffekt).

Die hier verwendete kausale Sprache ist natürlich mit größter Vorsicht zu behandeln. Pfadmodelle wurden zwar von Sewall Wright (1921, 1934) als kausale Modelle eingeführt, aber ihre Interpretation als kausale Modelle ist nur in besonderen Fällen erlaubt, z. B. wenn die beteiligten Variablen in besonderen Untersuchungsdesigns, wie z. B. dem randomisierten Experiment mit X als *Treatmentvariable* und Y als *Kriteriumsvariable* und Z als *Mediatorvariable* erhoben werden. Zur Kausalitätsproblematik bei Regressionsmodellen siehe z. B. Steyer (1992, 2003); Steyer, Gabler, von Davier und Nachtigall (2000a); Steyer, Gabler, von Davier, Nachtigall und Buhl (2000b); Steyer, Nachtigall, Wüthrich-Martone und Kraus (2002) sowie die dort zitierte Literatur.

3. Modelle der Klassischen Testtheorie

Das oben dargestellte Mediatormodell ist über direkt beobachtbare Variablen (Observablen) formuliert. In der Regel handeln psychologische Theorien und Hypothesen aber von theoretischen Größen, die nicht direkt beobachtbar sind. Nahezu ausnahmslos sind Observablen mit Messfehlern behaftet und unterscheiden sich u. U. durch weitere, z. B. situativ bedingte, Effekte von den theoretischen Größen, auf die in psychologischen Theorien Bezug genommen wird. Demnach sind in der Regel *latente Variablen* Gegenstand unserer inhaltlichen Theorien und Hypothesen. Sie repräsentieren die Konstrukte, von denen inhaltliche Theorien handeln, sofern sie schon soweit entwickelt sind, dass sie bereits einer empirischen Untersuchung zugänglich sind. Mit Strukturgleichungsmodellen können wir z. B. ein Mediatormodell für *latente* Variablen formulieren. Dies ist von großer praktischer Bedeutung, da man zeigen kann, dass man u. U. stark verfälschte Schätzungen erhält, wenn man anstelle der latenten Variablen die manifesten Variablen verwendet. Mit den Modellen für manifeste Variablen lassen sich keine Hypothesen prüfen, die sich eigentlich auf die entsprechenden latenten Variablen beziehen (s. z. B. Goldberger, 1973).

Wir werden nun eine erste Klasse von Modellen, die Modelle der *Klassischen Testtheorie* der Psychometrie (KTT) betrachten, in denen *latente Variablen* eingeführt werden. Ziele der Anwendung dieser Modelle sind die Messung von stabilen, aber auch von variablen Eigenschaften von Personen (seltener auch von Gruppen oder anderen Beobachtungseinheiten) und die Abschätzung des Ausmaßes der Messfehlerbehaftetheit dieser Messungen. Darüber hinaus kann man dann direkt den Zusammenhang der meßfehlerbereinigten latenten Variablen mit anderen Variablen im Rahmen größerer Strukturgleichungsmodelle untersuchen.

Um die Messfehlerbehaftetheit abschätzen zu können, wird die betrachtete Eigenschaft mindestens zweimal gemessen, um aus den Abweichungen zwischen den Messungen deren Ungenauigkeit und damit auch deren Genauigkeit erschließen zu können. Zunächst führen wir die notwendigen Grundbegriffe, dann entsprechende Modelle ein. Am Beispiel der KTT und ihrer Modelle kann man auch exemplarisch studieren, wie latente Variablen konstruiert werden können. Die gleichen Konstruktionsprinzipien gelten auch für Modelle der Latent-state-trait-Theorie (s. Abschnitt 4), der Theorie individueller und durchschnittlicher kausaler Effekte (s. Abschnitt 6) und der Item-response-Theorie (s. z. B. Steyer & Eid, 2001).

3.1. Grundbegriffe der Klassischen Testtheorie

In der KTT gehen wir vom folgenden Zufallsexperiment (Single-unit-trial) aus: Ziehe eine Person u aus einer Menge von Personen (der Grundgesamtheit oder Population) und beobachte Messwerte, die dieselbe Eigenschaft erfassen sollen. Dabei liegt weder fest, welche Person gezogen wird, noch zu welchem Resultat die Messung führt.

Die Messung einer Personeneigenschaft ist sicherlich der wichtigste Anwendungsbereich. Allgemeiner formuliert geht es um die Messung einer Eigenschaft einer beliebigen Beobachtungseinheit. Dabei kann es sich z. B. um die Messung eines bio-psychologischen Merkmals handeln, beispielsweise um die Konzentration eines bestimmten Hormons im Blut. Solche Messungen sind in aller Regel messfehlerbehaftet, und zwar nicht weniger als die Messungen, die man mit psychologischen Tests vornimmt (s. z. B. Kirschbaum et al., 1990). Aus heutiger Sicht ist die Anwendung der KTT auf Psychologische Tests sogar eher problematisch, da für die Analyse von Items eines Tests nur Modelle der Item-response-Theorie und Strukturgleichungsmodelle für ordinale Daten wirklich angemessen sind. Man beachte auch, dass an dieser Stelle die Begriffe „Eigenschaft“ oder „Merkmal“ noch sehr unspezifisch gebraucht sind. Es kann sich dabei auch um *variable* Zustände handeln.

Für jedes mögliche Ergebnis eines solchen Zufallsexperiments liefert der Messvorgang die Werte y_i der betrachteten *Messwertvariablen* Y_i , die wir im Folgenden auch einfach als „*Messung*“ bezeichnen. Neben diesen Messungen betrachten wir die *Personvariable* U , deren Wert die bei dem o. g. Zufallsexperiment gezogene Person u ist. Mit diesen Grundbegriffen können wir nun die *True-score-Variablen*

$$\tau_i := E(Y_i | U) \quad [3.1]$$

und die *Messfehlervariablen*

$$\varepsilon_i := Y_i - E(Y_i | U) \quad [3.2]$$

als eine spezielle Regression (oder bedingte Erwartung) bzw. als deren Residuum einführen (s. hierzu Steyer, 2003, Kap. 6). Sowohl die True-score- als auch die Fehlervariablen sind Zufallsvariablen, die sowohl untereinander als auch mit den Messungen (den Messwertvariablen) selbst in dem oben dargestellten Single-unit-trial eine gemeinsame Verteilung haben. Ein Wert der True-score-Variablen τ_i ist dann der wahre Wert $E(Y_i | U = u)$ der be-

trachteten Person u bzgl. der Messungen Y_i , d. h. der Erwartungswert der person-spezifischen Verteilung der Messwertvariablen Y_i . Ein Wert $y_i - E(Y_i | U = u)$ der Messfehlervariablen ε_i ist die Differenz zwischen einem Messwert y_i und dem wahren Wert $E(Y_i | U = u)$ dieser Person u (bzgl. der Messung Y_i).

Ziel in Anwendungen ist es nun u. a. das Ausmaß der Messfehler über deren Varianz $Var(\varepsilon_i)$ und die damit verknüpfte *Reliabilität*

$$Rel(Y_i) := \frac{Var[E(Y_i | U)]}{Var(Y_i)} = \frac{Var(\tau_i)}{Var(\tau_i) + Var(\varepsilon_i)} \quad [3.3]$$

abzuschätzen. Mit der KTT durchaus vereinbar ist es aber auch, wenn man *bedingte Fehlervarianzen* $Var(\varepsilon_i | X)$ schätzen möchte, wobei X sowohl die True-score-Variable τ_i als auch eine andere Variable sein kann, von der die Fehlervarianz abhängt.

Hier etwa Box 1

Abbildung 2 zeigt die Dekomposition der Messwertvariablen (s. auch Box 1), wobei alle True-score-Variablen und alle Messfehlervariablen untereinander korrelieren können. Ob diese Korrelationen vorliegen und wie stark sie sind, ist jeweils eine empirische Frage. Unkorreliert sind allerdings die Messfehlervariablen mit den True-score-Variablen, was aus deren Definition als eine Regression und deren Residuum in den Gleichungen [3.1] und [3.2] folgt (s. dazu Steyer, 2003, Regel-Box 6.3). Diese und einige andere Eigenschaften von True-score- und Fehlervariablen wurden früher als Axiome der KTT angesehen und auch heute noch in einigen Lehrbüchern so genannt. Seit Novick (1966) weiß man jedoch, dass diese keine Axiome, sondern Folgerungen aus ihren Definitionen sind.

Hier etwa Abbildung 2

3.2. Das Modell τ -äquivalenter Variablen

Um die o. g. Ziele, die Bestimmung der Fehlervarianz und der Reliabilität, zu erreichen, kann man verschiedene Annahmen über die True-score- und die Fehlervariablen einführen. Diese Annahmen definieren jeweils ein Modell der KTT. Das einfachste Modell besteht aus den Annahmen der τ -Äquivalenz und der Annahme unkorrelierter Fehler. Die Annahme unkorrelierter Fehler wurde früher als Axiom angesehen. Auch hier weiß man seit Novick (1966) mehr: Sie ist kein Axiom, folgt auch nicht aus der Definition des Messfehlers und kann in speziellen empirischen Anwendungen tatsächlich falsch sein (s. z. B. Zimmerman & Williams, 1977).

Die Annahme der τ -Äquivalenz ist $\tau_i = \tau_j$, für $i, j = 1, \dots, m$. Sie ist gleichbedeutend mit der Existenz einer latenten Variablen η , für die gilt:

$$Y_i = \eta + \varepsilon_i, \quad i = 1, \dots, m. \quad [3.4]$$

und

$$Cov(\varepsilon_i, \varepsilon_j) = 0, \quad \text{für } i \neq j, \quad [3.5]$$

$i, j = 1, \dots, m$. Das damit definierte Modell heißt das *Modell τ -äquivalenter Variablen*, da von allen betrachteten Messungen Y_i angenommen wird, dass sie die gleiche True-score-

Variable haben und in diesem Sinn also das Gleiche messen. Diese Annahmen implizieren dann die Kovarianzstruktur

$$\text{Cov}(Y_i, Y_j) = \begin{cases} \text{Var}(\eta), & i \neq j \\ \text{Var}(\eta) + \text{Var}(\varepsilon_i), & i = j \end{cases} \quad i, j = 1, \dots, m. \quad [3.6]$$

sowie die *Identifikationsgleichungen*

$$\text{Var}(\eta) = \text{Cov}(Y_i, Y_j), \quad i, j = 1, \dots, m, \quad i \neq j, \quad [3.7]$$

für die True-score-Varianz,

$$\text{Var}(\varepsilon_i) = \text{Var}(Y_i) - \text{Cov}(Y_i, Y_j), \quad i, j = 1, \dots, m, \quad i \neq j, \quad [3.8]$$

für die Fehlervarianz sowie

$$\text{Rel}(Y_i) := \text{Cov}(Y_i, Y_j) / \text{Var}(Y_i), \quad i = 1, \dots, m, \quad [3.9]$$

für die Reliabilität.

Die letzten drei Formeln zeigen, wie man die drei theoretischen Größen $\text{Var}(\eta)$, $\text{Var}(\varepsilon_i)$ und $\text{Rel}(Y_i)$ aus den Varianzen und Kovarianzen der Messwertvariablen Y_i berechnen kann, die durch die Erhebung einer Stichprobe und der Berechnung der entsprechenden Kennwerte in dieser Stichprobe empirisch schätzbar sind. Das Pfaddiagramm vereinfacht sich für dieses Modell entsprechend. Statt m True-score-Variablen kommt jetzt nur noch eine einzige gemeinsame True-score-Variable, die latente Variable η , vor (s. Abb. 3).

Hier etwa Abbildung 3

Im Fall von $m = 4$ Messungen Y_i kann man die von diesem Modell implizierte Kovarianzstruktur anschaulicher auch durch die folgende Matrix darstellen.

$$\Sigma_{yy} = \begin{pmatrix} \sigma_\eta^2 + \sigma_{\varepsilon_1}^2 & \sigma_\eta^2 & \sigma_\eta^2 & \sigma_\eta^2 \\ \sigma_\eta^2 & \sigma_\eta^2 + \sigma_{\varepsilon_2}^2 & \sigma_\eta^2 & \sigma_\eta^2 \\ \sigma_\eta^2 & \sigma_\eta^2 & \sigma_\eta^2 + \sigma_{\varepsilon_3}^2 & \sigma_\eta^2 \\ \sigma_\eta^2 & \sigma_\eta^2 & \sigma_\eta^2 & \sigma_\eta^2 + \sigma_{\varepsilon_4}^2 \end{pmatrix}. \quad [3.10]$$

Die Kovarianzen zwischen je zwei verschiedenen Messungen Y_i sind in diesem Modell also gleich. Dies ist eine wichtige Eigenschaft, die zur empirischen Überprüfung des Modells herangezogen werden kann. Dabei wird z. B. mittels eines Signifikanztests oder mit verschiedenen Kenngrößen, welche die Güte der Anpassung beschreiben, untersucht, ob die aus einer Stichprobe berechnete empirische Kovarianzmatrix nur zufällig von der vom Modell implizierten Kovarianzmatrix abweicht. Weitere testbare Konsequenzen folgen aus diesem Modell für die Erwartungswerte der Messwertvariablen, die alle gleich sein müssen, wenn die verschiedenen Messungen die gleiche True-score-Variable haben sollen (s. Annahme [3.4]).

3.3. Weitere Modelle der KTT

In vorangegangenen Abschnitt über die KTT haben wir die Grundbegriffe dargestellt und exemplarisch das Modell τ -äquivalenter Variablen behandelt, das als eines der einfachsten faktorenanalytischen Modelle und auch als ein sehr einfaches Strukturgleichungsmodell angesehen werden kann. Ein noch einfacheres Modell, das *Modell paralleler Variablen*, erhalten wir, wenn wir zu den Annahmen der τ -Äquivalenz [3.4] und unkorrelierter Messfehler [3.5] noch die Annahme gleicher Fehlervarianzen

$$\text{Var}(\varepsilon_i) = \sigma_\varepsilon^2, \quad \text{für } i = 1, \dots, m, \quad [3.11]$$

hinzufügen. Bei diesem Modell sind also auch alle diagonalen Komponenten in der Matrix [3.10] identisch. Die vom Modell implizierte Kovarianzstruktur ist also noch einfacher und restriktiver, da nur zwei verschiedene theoretische Parameter vorkommen, die Varianz der True-score-Variablen und die Varianz der Fehlervariablen.

Ein Verallgemeinerung des Modells τ -äquivalenter Variablen erhält man, wenn man neben unkorrelierten Messfehlern [3.5] die Annahme der *essentiellen τ -Äquivalenz* macht: $\tau_i = \tau_j + \lambda_{ij}$, für $i, j = 1, \dots, m$, wobei λ_{ij} eine beliebige Zahl ist. Die Annahme ist äquivalent mit der Annahme der Existenz einer für alle Messwertvariablen Y_i gemeinsamen latenten Variablen η , für die gilt:

$$\tau_i = \eta + \lambda_i, \quad \text{für } i = 1, \dots, m. \quad [3.12]$$

Dabei λ_i ist ein für jede Messwertvariable Y_i spezifischer *Lokationsparameter*. Wie man dieser Gleichung entnehmen kann, ist die latente Variable η eine (beliebige) Translation einer True-score-Variablen. Ein physikalisches Beispiel für ein solches Modell ist die Messung des Gewichts mit verschiedenen Haushaltswaagen, die allerdings verstellt sind, so dass sie für das gleiche Objekt jeweils ein um einen bestimmten Betrag verschobenes Gewicht angeben. Diese Verschiebungen sind jedoch für verschiedene Objekte konstant, so dass man trotz der Verstellungen *Gewichtsdifferenzen* zwischen Objekten problemlos messen kann.

Die von diesem Modell *essentiellen τ -äquivalenter Variablen* implizierte Kovarianzmatrix ist wiederum [3.10], da die additive Konstante λ_i bei der Bildung der Varianzen und Kovarianzen wegfällt. Die Erwartungswerte der Messwertvariablen Y_i müssen nun nicht mehr alle gleich sein. Allerdings müssen sie in verschiedenen Subpopulationen die gleiche Differenz aufweisen (s. Steyer & Eid, 2001).

Im Fall des *Modells τ -kongenerischer Variablen* ist die latente Variable η eine (zunächst beliebige) lineare Funktion jeder True-score-Variablen:

$$\tau_i = \lambda_{i0} + \lambda_{i1} \eta, \quad \text{für } i = 1, \dots, m. \quad [3.13]$$

Dabei ist λ_{i0} wieder ein Lokationsparameter und λ_{i1} heißt die *Ladung* der Variablen Y_i auf der latenten Variablen η . Ein solches Modell braucht man, wenn die Messwertvariablen Y_i zwar das Gleiche aber mit verschiedener Skalierung messen. Ein physikalisches Beispiel ist die Messung der Temperatur in Celsius-, Fahrenheit- und Kelvinskalen.

Mit der Annahme [3.13] ist die latente Variable η zunächst nur eindeutig bis auf eine lineare Transformation bestimmt. Daher kann man den Erwartungswert und die Varianz von η beliebig festlegen. Häufig setzt man daher: $E(\eta) = 0$ und $\text{Var}(\eta) = 1$ oder aber $\lambda_{i0} = 0$ und $\lambda_{i1} = 1$. Diese beiden Arten der Fixierung der Skala der latenten Variablen haben die gleiche Funktion, wie wenn man sich entschließt, die Temperatur in der Celsiusskala anzugeben.

Mit der Festlegung der Varianz der latenten Variablen auf 1 ist die vom Modell τ -kongenerischer Variablen implizierte Kovarianzstruktur bei vier Messwertvariablen:

$$\Sigma_{yy} = \begin{pmatrix} \lambda_{11}^2 + \sigma_{\varepsilon_1}^2 & \lambda_{21}\lambda_{11} & \lambda_{11}\lambda_{31} & \lambda_{11}\lambda_{41} \\ \lambda_{21}\lambda_{11} & \lambda_{21}^2 + \sigma_{\varepsilon_2}^2 & \lambda_{21}\lambda_{31} & \lambda_{21}\lambda_{41} \\ \lambda_{31}\lambda_{11} & \lambda_{31}\lambda_{21} & \lambda_{31}^2 + \sigma_{\varepsilon_3}^2 & \lambda_{31}\lambda_{41} \\ \lambda_{41}\lambda_{11} & \lambda_{41}\lambda_{21} & \lambda_{41}\lambda_{31} & \lambda_{41}^2 + \sigma_{\varepsilon_4}^2 \end{pmatrix}. \quad [3.14]$$

Diese Kovarianzstruktur ist bei vier Messwertvariablen immer noch restriktiv, da die 4 Varianzen und die 6 Kovarianzen der Messwertvariablen aus nur 8 theoretischen Parametern, den 4 Fehlervarianzen und den vier Ladungen, berechenbar sein müssen. Auch aus der Annahme der τ -Kongenerität (s. Annahme [3.13]) folgen testbare Konsequenzen für die Struktur der Erwartungswerte der Messwertvariablen in Subpopulationen. Zur empirischen Prüfung werden hier allerdings vier Subpopulationen benötigt (s. Steyer & Eid, 2001, S. 183).

Festzuhalten ist, dass hier latente Variablen eingeführt werden, indem wir Annahmen (z. B. essentielle τ -Äquivalenz oder τ -Kongenerität) über wohldefinierte Begriffe (die True-score-Variablen) machen, aus denen dann die Existenz und die Eindeutigkeit der latenten Variablen (und damit ihr Skalenniveau), aber auch testbare Konsequenzen (Kovarianzstruktur, Mittelwertstruktur) folgen. Auf ganz analoge Weise kann man übrigens auch die Modelle der Item-response-Theorie einführen, wobei allerdings an die Stelle von True-score-Variablen, die Logits bzw. Probits der Antwortwahrscheinlichkeiten und an die Stelle der Kovarianzstruktur die Strukturen bestimmter Antwort-Häufigkeiten treten. Den Modellen essentiell τ -äquivalenter und τ -kongenerischer Variablen, aber auch dem einfachsten Modell der Item-response-Theorie, dem Rasch-Modell, werden bei Steyer und Eid (2001) jeweils eigene Kapitel gewidmet, in denen man auch mehr zu messtheoretischen und anderen Aspekten nachlesen kann.

3.4. Weiterführende Literatur zur KTT

Viele wichtige Aspekte der KTT können wir hier nicht diskutieren. Dazu sei auf die klassischen Bücher von Gulliksen (1950) sowie Lord und Novick (1968) und viele andere Lehrbücher zur KTT verwiesen. Während Gulliksen noch fünf Axiome nennt, haben Novick (1966) und Zimmerman (1975) gezeigt, dass, bis auf die Unkorreliertheit der Fehler, alle anderen Axiome Gulliksens schon Folgerungen aus der Definition der True-score- und der Fehlervariablen sind. Dies wird auch von Steyer und Eid (2001) herausgearbeitet. Darüber hinaus wird dort die Existenz der latenten Variablen aus den Annahmen über die True-score-Variablen hergeleitet. Dies ist sehr hilfreich, wenn man verstehen will, was latente Variablen eigentlich sind. Zur Darstellung des Anwendungskontextes der KTT ist auch heute noch Lienert (1989) lesenswert, auch wenn man dort über die eigentliche Theorie, d. h. die Definitionen der zentralen Begriffe, die Folgerungen daraus und die Modelle der KTT nichts finden kann. Nachdem wir mit den Modellen der Item-response-Theorie und den Strukturgleichungsmodellen für dichotome und ordinale Variablen statistische Modelle vorliegen haben, die explizit zur Analyse von *Items* psychologischer Tests konstruiert wurden, sind die bei Lienert beschriebenen Vorgehensweisen obsolet geworden.

4. Latent-state-trait-Modelle

Neben Messfehlern, um deren Abschätzung man sich mit den oben skizzierten Modellen der KTT bemüht, kommt in der KTT nur eine einzige weitere Varianzquelle vor, die Personenvariable U . Zwar kann man diese Varianzquelle durchaus auch allgemeiner interpretieren, z. B. als eine Menge von Gruppen, eine Menge von Personen, die sich jeweils in einer Situation befinden, etc., aber dennoch bleibt es, neben den Messfehlern, bei einem einzigen Faktor als Varianzquelle. Bei psychologischen Messungen spielen aber auch situative Effekte und Interaktionen zwischen Personen und Situationen eine nicht zu vernachlässigende Rolle (s. z. B. Mischel, 1977a, 1977b). Um auch diese abschätzen und ausfiltern zu können, wurde die Latent-state-trait-Theorie (LST-Theorie; Steyer, Ferring, & Schmitt, 1992; Steyer, Schmitt, & Eid, 1999) entwickelt. Modelle der LST-Theorie sind recht einfache Erweiterungen der oben dargestellten Modelle der KTT für mehrere Messzeitpunkte. Im Rahmen dieses Artikels über Strukturgleichungsmodelle liefern sie erste Modelle mit *mehreren* latenten Variablen. Zunächst führen wir wieder die notwendigen Grundbegriffe, dann die entsprechenden Modelle ein.

4.1. Grundbegriffe

Ausgangspunkt der LST-Theorie ist der folgende Typ eines Zufallsexperiments: Es wird eine Person aus einer *Population von Personen* gezogen, an der dann zu n Messgelegenheiten Beobachtungen gemacht werden. Zu jeder Messgelegenheit t realisiert sich eine (in der Regel unbekannte biologische, psychologische und soziale) Situation und es wird eine Beobachtung registriert. Dabei können sowohl die Situation als auch die Beobachtung mehrdimensional sein. So könnte z. B. die Situation simultan durch mehrere Aspekte definiert sein und die (zunächst möglicherweise nur qualitative) Beobachtung aus der Registrierung der Antworten auf mehrere Fragen in einem Fragebogen bestehen. Die Beobachtung könnte aber auch aus mehreren Skalenwerten oder Messwerten bestehen, die das gleiche Merkmal (Parallelförmigkeit) oder aber verschiedene Merkmale erheben sollen.

Die i -te reellwertige *Messwertvariable* oder *Messung*, die zur t -ten Messgelegenheit erhoben wird, notieren wir mit Y_{it} . Dabei kann es sich z. B. um physiologische Messungen (Blutdruck, Hautleitfähigkeit, EEG-Amplituden und -Frequenzen, Hormonspiegel, etc.) handeln (s. z. B. Hagemann, Naumann, Thayer, & Bartussek, 2002; Kirschbaum et al., 1990).

Außer auf die Messwertvariablen Y_{it} können wir auch auf die *Personvariable* U und die *Situationsvariablen* S_t zurückgreifen. Dabei gibt ein Wert von U an, welche Person aus der Population gezogen wird und ein Wert von S_t , welche Situation sich zur t -ten Messgelegenheit für die gezogene Person realisiert hat. (Dabei ist es nicht unbedingt notwendig, dass diese Situation in irgendeiner Form beobachtet wird.) Mit diesem begrifflichen Instrumentarium können wir nun die verschiedenen Grundbegriffe der LST-Theorie definieren.

Die *Latent-state-Variable*

$$\tau_{it} := E(Y_{it} | U, S_t) \quad [4.1]$$

ist die Regression (oder bedingte Erwartung) von Y_{it} auf U und S_t . Beide Faktoren, Person und Situation, sowie deren Interaktion determinieren also definitionsgemäß den Zustand einer Person bzgl. der i -ten Messung zur Messgelegenheit t . Der Statebegriff bzgl. einer Messung resultiert also durch die Ausfilterung der *Messfehlervariablen*

$$\varepsilon_{it} := Y_{it} - \tau_{it}, \quad [4.2]$$

d. h. $\tau_{it} = Y_{it} - \varepsilon_{it}$. Die *Latent-trait-Variable*

$$\xi_{it} := E(Y_{it} | U) \quad [4.3]$$

dagegen ist definitionsgemäß die Regression von Y_{it} auf U . Daher determiniert nur die Personvariable U den Trait einer Person bzgl. der i -ten Messung zur Messgelegenheit t . Der Trait bzgl. einer Messung Y_{it} resultiert also durch die Ausfilterung (a) der Messfehlervariablen, (b) situativer Effekte sowie (c) der Effekte der Interaktion zwischen Person und Situation aus der Messwertvariablen Y_{it} . Dies erkennt man, wenn man das *Latent-state-Residuum*

$$\zeta_{it} := \tau_{it} - \xi_{it} \quad [4.4]$$

betrachtet. Man kann nämlich zeigen, dass nicht nur $\xi_{it} = Y_{it} - \zeta_{it} - \varepsilon_{it}$ gilt, sondern auch, dass das Latent-state-Residuum ζ_{it} aus situativen Effekten $E(Y_{it} | S_t)$ und der Interaktion zwischen Person und Situation $[E(Y_{it} | U, S_t) - E(Y_{it} | U) - E(Y_{it} | S_t)]$ besteht. Es gilt nämlich:

$$\begin{aligned} \zeta_{it} &:= \tau_{it} - \xi_{it} = E(Y_{it} | U, S_t) - E(Y_{it} | U) \\ &= E(Y_{it} | S_t) + [E(Y_{it} | U, S_t) - E(Y_{it} | U) - E(Y_{it} | S_t)]. \end{aligned} \quad [4.5]$$

Eine Umordnung der Definitionsgleichungen für die Latent-state- bzw. die Latent-trait-Variablen führt zu den beiden *Dekompositionen der Variablen* $Y_{it} = \tau_{it} + \varepsilon_{it}$ und $\tau_{it} = \xi_{it} + \zeta_{it}$. Demnach wird also eine Messwertvariable Y_{it} additiv in eine Latent-state-Variable τ_{it} und eine Messfehlervariable ε_{it} und die Latent-state-Variable ihrerseits additiv in die Latent-trait-Variable ξ_{it} und das Latent-state-Residuum ζ_{it} zerlegt.

Zur Interpretation der durch Gleichung [4.3] definierten Latent-trait-Variablen ist zu beachten, dass die so definierten Traits keineswegs unveränderlich sind. Sie sind zwar in der Regel zeitlich stabiler, als die entsprechenden Latent-state-Variablen, aber sie können sich durchaus über die Zeit verändern. Lern-, Entwicklungs- und Reifungsprozessen stehen diesem Traitbegriff keineswegs entgegen. Im Gegenteil, diese werden erst durch diesen Traitbegriff und die dazu passenden Modelle angemessenen empirischen Untersuchungen zugänglich gemacht (s. dazu Steyer, Krambeer, & Hannover, 2004).

 Hier etwa Box 2 einfügen

Interessanterweise gehen in die obigen Zerlegungen keinerlei Annahmen ein, die sich in einer Anwendung als falsch erweisen könnten. Vielmehr handelt es sich um mathematische Folgerungen aus *der Definition* der vier Arten theoretischer Variablen τ_{it} , ε_{it} , ξ_{it} und ζ_{it} . Noch interessanter ist, dass auch die entsprechenden *Dekompositionen der Varianzen* $Var(Y_{it}) = Var(\tau_{it}) + Var(\varepsilon_{it})$ und $Var(\tau_{it}) = Var(\xi_{it}) + Var(\zeta_{it})$ auf keinerlei Annahmen beruhen, sondern ebenfalls ausschließlich aus den obigen Definitionen und den allgemeingültigen Eigenschaften des Regressionsbegriffs (s. Steyer, 2003, Kap. 6) abgeleitet werden können.

Auf dieser allgemeingültigen Additivität der Varianzen beruhen die Definitionen der *Reliabilität*, *Konsistenz* und *Messgelegenheitsspezifität*, die wichtigsten Kenngrößen der LST-Theorie. Die *Reliabilität*

$$Rel(Y_{it}) := Var(\tau_{it}) / Var(Y_{it}) \quad [4.6]$$

ist ein Maß für die Genauigkeit des durch Y_{it} repräsentierten Messverfahrens in dem oben beschriebenen Zufallsexperiment. Die *Konsistenz*

$$Con(Y_{it}) := Var(\xi_{it}) / Var(Y_{it}) \quad [4.7]$$

beschreibt das Ausmaß, in dem die Messung Y_{it} durch den Personvariable bedingt ist, wohingegen die *Messgelegenheitsspezifität*

$$Spe(Y_{it}) := Var(\zeta_{it}) / Var(Y_{it}) \quad [4.8]$$

angibt, inwieweit Situation und Interaktion zwischen Person und Situation die Messung determinieren. Konsistenz und Messgelegenheitsspezifität addieren sich zur Reliabilität auf. Darüber hinaus kann man natürlich auch den relativen Anteil der traitbedingten Varianz $Var(\xi_{it})$ und den relativen Anteil der Situations- und Interaktionsvarianz $Var(\zeta_{it})$ an der gesamten *systematischen* Varianz $Var(\tau_{it})$ betrachten, die sich dann zu 1 aufaddieren.

 Hier Abbildung 4 einfügen

Abbildung 4 zeigt die in der LST-Theorie vorgenommenen Dekompositionen der Messwertvariablen Y_{it} (durch Vierecke dargestellt) in die verschiedenen LST-theoretischen Komponenten (dargestellt durch Kreise). Die Messfehlervariablen und die Latent-state-Residuen sind nicht durch Kreise gekennzeichnet, zählen aber ebenfalls zu den LST-theoretischen, nicht direkt beobachtbaren Variablen (s. auch Gähde, Jagodzinski, & Steyer, 1992).

In dieser Abbildung wird auch deutlich, dass zu vielen theoretischen Variablen viel zu wenige Messwertvariablen gegenüberstehen. Ohne die Hinzufügung von Annahmen, die dieses Missverhältnis drastisch ändert, besteht keine Möglichkeit, aus den Verteilungen der Messungen (oder deren Kennwerte wie Mittelwerte, Varianzen, Kovarianzen und Korrelationen) etwas über die theoretischen Variablen (z. B. deren Mittelwerte, Varianzen, Kovarianzen und Korrelationen) ableiten zu können. Dabei beachte man, dass zunächst alle Latent-state-Variablen, alle Latent-state-Residuen, alle Latent-trait-Variablen und alle Messfehlervariablen miteinander korrelieren können. Erst in den unten dargestellten Modellen werden viele dieser Variablen als unkorreliert angenommen. Aus den Definitionen der LST-theoretischen Variablen τ_{it} , ε_{it} , ξ_{it} und ζ_{it} folgen jedoch schon für alle Indexpaare (i, t) und (j, s) die folgenden Nullkorrelationen:

$$Cov(\varepsilon_{it}, \tau_{js}) = Cov(\varepsilon_{it}, \xi_{js}) = Cov(\varepsilon_{it}, \zeta_{js}) = Cov(\zeta_{it}, \xi_{js}) = 0. \quad [4.9]$$

Demnach sind also die Messfehlervariablen mit den State- und Trait-Variablen sowie mit den Latent-state-Residuen unkorreliert. Außerdem sind auch die Latent-state-Residuen mit den Latent-trait-Variablen unkorreliert. Dabei beachte man, dass es sich hier wieder nicht um Annahmen handelt, die in irgendeiner Anwendung falsch sein könnten, sondern um logische Folgerungen aus den Definitionen der LST-theoretischen Variablen. (Beweise hierzu, die auf den Eigenschaften von Regressionen und ihrer Residuen beruhen, findet man z. B. in Steyer & Schmitt, 1990).

4.2. Das Multistate-Modell

Modelle der LST-Theorie entstehen nun durch die Einführung von Annahmen, die man mithilfe der oben dargestellten Grundbegriffe formulieren kann. Repräsentieren bspw. Y_{it} und Y_{jt} zu jeder Messgelegenheit t zwei Messungen des gleichen Zustands, so ist die Annahme

$$\tau_{it} = \tau_{jt} =: \eta_t, \quad i, j = 1, \dots, m, \quad t = 1, \dots, n, \quad [4.10]$$

dass ihre Latent-state-Variablen identisch sind, plausibel. Diese Annahme können wir alternativ auch wie folgt schreiben:

$$Y_{it} = \eta_t + \varepsilon_{it}, \quad i = 1, \dots, m, \quad t = 1, \dots, n, \quad [4.11]$$

Die gemeinsame latente Zustandsvariable zur Messgelegenheit t bezeichnen wir jeweils mit η_t . Eine weitere Vereinfachung ergibt sich, wenn die Messfehlervariablen ε_{it} als unkorreliert angenommen werden können

$$\text{Cov}(\varepsilon_{it}, \varepsilon_{js}) = 0, \quad (i, t) \neq (j, s). \quad [4.12]$$

Wie bei der KTT kann die Annahme [4.10] bzw. [4.11] durch die entsprechenden Annahmen der essentiellen Äquivalenz bzgl. η_t bzw. der Kongenerität bzgl. η_t abgeschwächt werden.

 Hier etwa Abbildung 5

Das mit den beiden Annahmen [4.10] und [4.12] (bzw. den gerade genannten Abschwächungen) definierte Modell bezeichnen wir als das *Multistate-Modell* der LST-Theorie. In Abbildung 5 ist das zugehörige Pfaddiagramm dieses Modells angegeben. Die implizierte Kovarianzstruktur ist wie folgt:

$$\text{Var}(Y_{it}) = \sigma_{\eta_t}^2 + \sigma_{\varepsilon_{it}}^2, \quad [4.13]$$

und

$$\text{Cov}(Y_{it}, Y_{js}) = \begin{cases} \sigma_{\eta_t}^2, & \text{falls } i \neq j \text{ und } t = s \\ \sigma_{\eta_t \eta_s}, & \text{falls } t \neq s. \end{cases} \quad [4.14]$$

Bei $n = 2$ Messgelegenheiten mit jeweils $m = 2$ Messwertvariablen hat die von diesem Modell implizierte Varianz-Kovarianzmatrix Σ_{yy} der Messwertvariablen Y_{it} also die Form:

$$\Sigma_{yy} = \begin{pmatrix} \sigma_{\eta_1}^2 + \sigma_{\varepsilon_{11}}^2 & \sigma_{\eta_1}^2 & \sigma_{\eta_1 \eta_2} & \sigma_{\eta_1 \eta_2} \\ \sigma_{\eta_1}^2 & \sigma_{\eta_1}^2 + \sigma_{\varepsilon_{21}}^2 & \sigma_{\eta_1 \eta_2} & \sigma_{\eta_1 \eta_2} \\ \sigma_{\eta_1 \eta_2} & \sigma_{\eta_1 \eta_2} & \sigma_{\eta_2}^2 + \sigma_{\varepsilon_{12}}^2 & \sigma_{\eta_2}^2 \\ \sigma_{\eta_1 \eta_2} & \sigma_{\eta_1 \eta_2} & \sigma_{\eta_2}^2 & \sigma_{\eta_2}^2 + \sigma_{\varepsilon_{22}}^2 \end{pmatrix}. \quad [4.15]$$

Anhand dieser Matrix oder auch der Gleichungen [4.13] und [4.23] kann man sich nun leicht überlegen, wie man die theoretischen Parameter dieses LST-Modells aus den empirisch

schätzbaren Varianzen und Kovarianzen der Messwertvariablen Y_{it} bestimmen kann. So ist z. B. die Varianz der Latent-state-Variablen η_t gleich der Kovarianz zwischen Y_{it} und Y_{jt} , $i \neq j$. Die Kovarianz der Latent-state-Variablen η_t und η_s zweier verschiedener Messzeitpunkte t und s ist gleich der Kovarianz zwischen Y_{it} und Y_{js} , $t \neq s$.

4.3. Das Multistate-singletrait-Modell

Führen wir außer den Annahmen [4.10] und [4.12] noch die Annahme ein, dass die latenten Eigenschaften sowohl zwischen den Messwertvariablen als auch über den betrachteten Messzeitraum hinweg identisch sind,

$$\xi_{it} = \xi_{js} =: \xi, \quad i, j = 1, \dots, m, \quad s, t = 1, \dots, n, \quad [4.16]$$

so können wir die Latent-trait-Variablen ξ_{it} durch eine einzige Latent-trait-Variable ξ ersetzen. Diese Annahmen implizieren die Gleichungen

$$Y_{it} = \eta_t + \varepsilon_{it}, \quad i = 1, \dots, m, \quad [4.17]$$

$$\eta_t = \xi + \zeta_t, \quad t = 1, \dots, n, \quad [4.18]$$

wobei ε_{it} die in Gleichung [4.2] eingeführten Messfehlervariablen und ζ_t die in [4.4] definierten Latent-state-Residuen sind, die wir hier allerdings wegen Annahme [4.10] nur noch mit einem einzigen Index schreiben müssen. Anstelle der Annahme [4.16] könnten wir auch hier wieder die schwächeren Annahmen der essentiellen Äquivalenz oder der Kongenerität bzgl. ξ einführen.

Setzen wir die Gleichung [4.18] in die Gleichung [4.17] ein, erhalten wir

$$Y_{it} = \xi + \zeta_t + \varepsilon_{it}, \quad i = 1, \dots, m, \quad t = 1, \dots, n. \quad [4.19]$$

Während die Gleichungen [4.17] und [4.18] ein faktorenanalytisches Modell *zweiter Ordnung* darstellen, da die Latent-state-Variablen in weitere Faktoren, die Latent-trait-Variable und die Latent-state-Residuen zerlegt werden, liegt mit der Gleichung [4.19] eine Darstellung des gleichen Modells als faktorenanalytisches Modell *erster Ordnung* vor. In dieser Darstellung werden scheinbar nur die Messwertvariablen in Faktoren zerlegt (s. auch die Abbildungen 6a und 6b).

Weitere Vereinfachungen ergeben sich, wenn wir annehmen, dass die Situationen, in denen die Tests (oder Messungen) erhoben werden, zwischen den Messgelegenheiten unabhängig sind, oder dass wenigstens gilt:

$$\text{Cov}(\zeta_t, \zeta_s) = 0, \quad t \neq s. \quad [4.20]$$

Die Unkorreliertheit

$$\text{Cov}(\xi, \zeta_t) = 0, \quad [4.21]$$

der Latent-trait-Variablen mit den Latent-state-Residuen dagegen folgt bereits aus den Gleichungen [4.9] und [4.16]. In den Abbildungen 6a und b sind zwei verschiedene Pfaddiagramme dieses Modells angegeben, das wir als *Singletrait-multistate-Modell* bezeichnen.

Hier Abbildungen 6a und 6b.

Die von diesem Modell implizierte Kovarianzstruktur ist wie folgt:

$$\text{Var}(Y_{it}) = \sigma_{\xi}^2 + \sigma_{\zeta_t}^2 + \sigma_{\varepsilon_{it}}^2, \quad [4.22]$$

und

$$\text{Cov}(Y_{it}, Y_{js}) = \begin{cases} \sigma_{\xi}^2 + \sigma_{\zeta_t}^2, & \text{falls } i \neq j \text{ und } t = s \\ \sigma_{\xi}^2, & \text{falls } t \neq s. \end{cases} \quad [4.23]$$

Bei $n = 2$ Messgelegenheiten zu denen jeweils $m = 2$ Messungen vorgenommen werden, hat die von diesem Modell implizierte Varianz-Kovarianzmatrix Σ_{yy} der Messwertvariablen also die Form:

$$\Sigma_{yy} = \begin{pmatrix} \sigma_{\xi}^2 + \sigma_{\zeta_1}^2 + \sigma_{\varepsilon_{11}}^2 & \sigma_{\xi}^2 + \sigma_{\zeta_1}^2 & \sigma_{\xi}^2 & \sigma_{\xi}^2 \\ \sigma_{\xi}^2 + \sigma_{\zeta_1}^2 & \sigma_{\xi}^2 + \sigma_{\zeta_1}^2 + \sigma_{\varepsilon_{21}}^2 & \sigma_{\xi}^2 & \sigma_{\xi}^2 \\ \sigma_{\xi}^2 & \sigma_{\xi}^2 & \sigma_{\xi}^2 + \sigma_{\zeta_2}^2 + \sigma_{\varepsilon_{12}}^2 & \sigma_{\xi}^2 + \sigma_{\zeta_2}^2 \\ \sigma_{\xi}^2 & \sigma_{\xi}^2 & \sigma_{\xi}^2 + \sigma_{\zeta_2}^2 & \sigma_{\xi}^2 + \sigma_{\zeta_2}^2 + \sigma_{\varepsilon_{22}}^2 \end{pmatrix}. \quad [4.24]$$

Anhand dieser Matrix oder auch der Gleichungen [4.22] und [4.23] kann man sich nun leicht überlegen, wie man die theoretischen Parameter dieses LST-Modells aus den empirisch schätzbaren Varianzen und Kovarianzen der Messungen Y_{it} bestimmen kann. So ist z. B. die Varianz der Latent-trait-Variablen gleich den vier Kovarianzen der Messwertvariablen Y_{i1} und Y_{j2} .

Die hier vorgestellten Singletrait-multistate-Modelle sind eine relativ realistische Formulierung der Hypothese, dass keine Traitveränderung stattfindet. Sie lassen Messfehler sowie situativ und interaktiv bedingte Fluktuationen zu, nicht aber eine Traitveränderung. In der Entwicklungspsychologie bedeutet dies, dass in dem betrachteten Zeitraum bzgl. des betrachteten Merkmals keine Entwicklung, kein Lernen oder eine sonstige entwicklungspsychologisch relevante Veränderung stattfindet. Hat man vier Messgelegenheiten vorliegen und zwischen den Messgelegenheiten zwei und drei findet eine Intervention statt, so würde man mit dem Singletrait-multistate-Modell formulieren, dass es auch keine differentiellen Effekte der Intervention gibt. Diese Modelle werden daher bei der Analyse individueller kausaler Effekte (s. Abschnitt 6) eine wichtige Rolle spielen, insbesondere dann, wenn sie noch um Methodenfaktoren erweitert werden, die es ermöglichen, dass man innerhalb jeder Messgelegenheit nicht mehr unbedingt von zwei (oder mehr) perfekt unidimensionalen Messungen ausgehen muss. In diesem Zusammenhang sind die von Eid (2000) sowie Eid, Lischetzke, Trierweiler und Nußbeck (2003) entwickelten Modelle zur Behandlung von Methodenfaktoren wichtig, die es erlauben, die systematischen Unterschiede zwischen den Messungen adäquat zu modellieren.

4.4. Weiterführende Literatur

In verschiedenen Arbeiten (s. z. B. Steyer et al., 1992; Eid, 1995; Steyer, Eid, & Schwenkmezger, 1996) wurden alternative, weniger restriktive Annahmen dargestellt, die ebenfalls zu Modellen führen, die erlauben, die LST-theoretischen Kenngrößen zu bestimmen und das resultierende Modell zu testen, indem man die Implikationen für die Struktur der Kovarianzmatrix der Messwertvariablen auf Übereinstimmung mit deren empirischer Kovarianzmatrix untersucht. Verallgemeinerungen erhält man sofort, wenn man die Ladungen der Messungen auf den Latent-state-Variablen nicht mehr auf 1 fixiert [s. Gln. [4.11], [7.4], [4.17], und [4.19]]. Darüber hinaus gibt es natürlich auch Modelle, in denen Latent-trait-Variablen vorkommen, die sich über die Zeit verändern (s. z. B. Eid & Hoffmann, 1998; Steyer et al., 2004). Allerdings braucht man für solche Modelle mindestens zwei relativ weit auseinander liegende Erhebungsphasen, innerhalb derer mindestens zweimal eine Latent-state-Variable mit mindestens je zwei parallelen Messungen erhoben werden. Die beiden Erhebungsphasen sollten deswegen relativ weit auseinander liegen, damit eine wahre Traitveränderung auch möglich ist. Was „relativ weit“ heißt, hängt vom betrachteten Konstrukt und den sonstigen Bedingungen ab, insbesondere davon, wie schnell dabei eine tatsächliche Traitveränderung erwartet werden kann. In bestimmten Phasen kann eine Entwicklung (d. h. eine wahre Traitveränderung) mehr oder auch weniger schnell vonstatten gehen.

Steyer et al. (1999) geben einen Überblick über verschiedene Forschungsfragen, die mit LST-Modellen untersucht werden können und über vorliegende Anwendungen in verschiedenen Gebieten der Psychologie. Die Beziehung zu Wachstumskurvenmodellen wird von Tisak und Tisak (2000) herausgearbeitet. Eid (1995) stellt eine Erweiterung der LST-Theorie für Probit-Modelle, d. h. für Strukturgleichungsmodelle mit dichotomen und ordinalen Variablen dar. Andere Erweiterungen wurden kürzlich von Schermelleh-Engel, Keith, Moosbrugger, & Hodapp (2004) sowie von Cole, Martin, & Steiger (in press) vorgestellt.

5. Latent-change-Modelle und Wachstumskurven-Modelle

Die im letzten Abschnitt dargestellten Multistate-Modelle bieten einen guten Ausgangspunkt zur Einführung von Latent-change-Modellen und von Wachstumskurvenmodellen. Oft ist von Interesse, ob sich Personen in der Veränderung bestimmter Eigenschaften voneinander unterscheiden und wenn ja, womit diese inter-individuellen Unterschiede in der intra-individuellen Veränderung zusammenhängen. In der Pädagogischen Psychologie geht es dabei um die Frage der Unterschiede in der Lernfähigkeit bzw. Lerngeschwindigkeit, in der Klinischen Psychologie um die Frage der differentiellen Indikation von Therapien, in der Entwicklungspsychologie um die Erklärung der Unterschiede im Ausmaß entwicklungsbedingter Veränderungen. Obwohl all die genannten inhaltlichen Fragestellungen sich eigentlich auf wahre *Traitveränderung* beziehen, wollen wir uns hier der Einfachheit halber auf die Frage der *Stateveränderung* beschränken. Die hier besprochenen Prinzipien lassen sich jedoch leicht auf Fragen der Traitveränderung verallgemeinern (s. z. B. Steyer et al., 2004).

5.1. Latent-change-Modelle

Mit den Gleichungen [4.11] und [4.12] haben wir das Multistate-Modell definiert. Für den zweiten und die weiteren Zeitpunkte lässt sich die Gleichung [4.11] aber auch wie folgt schreiben:

$$Y_{it} = \eta_t + \varepsilon_{it} = \eta_{t-1} + (\eta_t - \eta_{t-1}) + \varepsilon_{it} \quad i=1, \dots, m, \quad t=2, \dots, n. \quad [5.1]$$

Damit ist die latente Veränderungsvariable $\eta_t - \eta_{t-1}$ eingeführt (s. auch Abb. 7), deren Varianz von Interesse sein kann, da sie die Frage beantwortet, ob es inter-individuelle Unterschiede in der intra-individuellen Veränderung gibt. Darüber hinaus kann die Veränderungsvariable mit anderen Veränderungsvariablen korreliert sein (Steyer, Eid, & Schwenkmezger, 1997) oder als unabhängige oder auch als abhängige Variable in einer Regressionsgleichung betrachtet werden (Steyer, Partchev, & Shanahan, 2000).

 Hier Abbildung 7

Alternativ zu Gleichung [5.1] kann das Multistate-Modell auch durch die folgenden Gleichungen ergänzt werden:

$$\eta_t = \eta_{t-1} + (\eta_t - \eta_{t-1}), \quad t=2, \dots, n. \quad [5.2]$$

Diese Gleichung ist zwar mathematisch trivial, führt aber ebenso zur Einführung der latenten Veränderungsvariablen (s. Abbildung 7 b), deren Varianz und deren Zusammenhang mit anderen Variablen man dann explizit in einem umfassenderen Strukturgleichungsmodell analysieren kann.

Die Gleichung [5.2] bezieht sich zwar im hier vorliegenden Kontext explizit auf *Latent-state-Variablen*, aber man kann eine entsprechende Gleichung natürlich auch über *Latent-trait-Variablen* formulieren und damit die von Messfehlern und situativen Schwankungen bereinigten Traitveränderungen betrachten, sofern man in einem Modell mindestens zwei Latent-trait-Variablen vorliegen hat (s. z. B. Steyer et al., 2004).

Die Analyse von Veränderungsvariablen ist oft auch dadurch motiviert, dass man glaubt, mit Veränderungsvariablen eher kausale Effekte erfassen zu können, als mit der Analyse von Zustandsvariablen. Die dahinter stehende Intuition ist wohl, dass eine Veränderung auf eine zwischenzeitlich durchgeführte Intervention zurückführbar sein muss. Warum sollte sonst eine Veränderung stattfinden? Auch wenn wir wissen, dass eine Veränderung neben einer Intervention auch andere Ursachen (z. B. Reifung, Lernen, Transfereffekte, kritische Lebensereignisse) haben kann (s. z. B. Campbell & Stanley, 1963; Cook & Campbell, 1979; Shadish, Cook, & Campbell, 2002), ist der Grundgedanke nicht völlig verkehrt. Steyer (2005) stellt Annahmen, Designs und Modelle dar, in denen die Latent-change-Variable $\eta_t - \eta_{t-1}$ als Individual-causal-effect-Variable interpretiert werden kann.

5.2. Wachstumskurven-Modelle

Auf ganz ähnliche Weise wie mit Gleichung [5.2] lassen sich nun auch Wachstumskurvenmodelle einführen. Anstelle der latenten Veränderungsvariablen $\eta_t - \eta_{t-1}$ werden dabei latente Wachstumskurvenkomponenten π_0, π_1, \dots eingeführt, bezüglich derer sich Personen unterscheiden können, d. h. jede Person kann einen unterschiedlichen Wert auf den latenten Variablen π_0, π_1, \dots haben. Liegen bspw. $n = 3$ Messzeitpunkte vor, so lassen sich durch

$$\eta_t = \pi_0 + \lambda_t \cdot \pi_1 + \lambda(t)^2 \cdot \pi_2, \quad t=1, \dots, n, \quad [5.3]$$

drei solcher Komponenten π_0, π_1 und π_2 einführen, wobei die Koeffizienten λ_t von der Zeit t abhängen. Dass man auf diese Weise tatsächlich neue latente Variablen einführen kann, die eine sinnvolle Interpretation haben, kann man sich am besten anhand eines speziellen Falls

klarmachen. Definieren wir $\lambda_1 := -1$, $\lambda_2 := 0$, $\lambda_3 := +1$, so erhalten wir aus [5.3] die drei Gleichungen:

$$\eta_1 = \pi_0 + (-1) \cdot \pi_1 + 1 \cdot \pi_2, \quad [5.4]$$

$$\eta_2 = \pi_0 + 0 \cdot \pi_1 + 0 \cdot \pi_2 = \pi_0, \quad [5.5]$$

$$\eta_3 = \pi_0 + 1 \cdot \pi_1 + 1 \cdot \pi_2. \quad [5.6]$$

Wie man nun leicht erkennt, handelt es sich hier um drei Gleichungen mit den drei Unbekannten π_0 , π_1 und π_2 . Für je drei Werte der Zustandsvariablen η_1 , η_2 und η_3 lassen sich daher drei Werte der Komponenten π_0 , π_1 und π_2 ausrechnen. Eine Person mit den Werten der Zustandsvariablen $\eta_1 = 5$, $\eta_2 = 7$ und $\eta_3 = 9$ bspw. hat dann die Komponenten $\pi_0 = 7$, $\pi_1 = 2$, und $\pi_2 = 0$. Tatsächlich können wir π_0 als *Niveauekomponente*, π_1 als *lineare* und π_2 als *quadratische Wachstumskurvenkomponente* interpretieren. Dabei beachte man, dass diese Interpretationen aber von der Wahl der Zeitkodierung λ_t abhängen. Bei der Zeitkodierung $\lambda_1 := 0$, $\lambda_2 := 1$, $\lambda_3 := 2$ bspw. ist π_1 *nicht* als lineare Wachstumskurvenkomponente zu interpretieren Biesanz, Deeb-Sossa, Papadakis, Bollen, & Curran, 2004. In Abbildung 8 ist das Pfaddiagramm des durch die Gleichungen [5.4] bis [5.6] definierten Wachstumskurvenmodells dargestellt, wobei wir unterstellen, dass zu jedem Zeitpunkt t zwei parallele Messungen der Latent-state-Variablen η_t vorliegen.

Hier Abbildung 8

Was die implizierte Kovarianzstruktur angeht, ist festzustellen, dass sich diese gegenüber dem Multistate-Modell in keiner Weise ändert. Mit der Einführung der latenten Veränderungsvariablen $\eta_t - \eta_{t-1}$ oder der Wachstumskurvenkomponenten π_0 , π_1 und π_2 werden keinerlei neue Restriktionen der Varianzen und Kovarianzen der manifesten Variablen eingeführt oder gelockert. In beiden Fällen handelt es sich lediglich um eine Reparametrisierung des Multistate-Modells. Daher werden sich auch alle Kennwerte der Modellanpassung nicht verändern.

6. Multitrait-multimethod-Modelle

Bei den in Abschnitt 4 dargestellten Multistate-Modellen gehen wir davon aus, dass zu jeder von mehreren Messgelegenheiten mindestens zwei Messwertvariablen die gleiche Latent-state-Variable haben und sich nur durch die jeweiligen Messfehlervariablen voneinander unterscheiden (s. die Gleichung [4.11]). Dies ist eine relativ strenge Annahme, die nur selten perfekt erfüllt ist. Selbst wenn die Messinstrumente mit dem Ziel konstruiert werden, dass sie die gleiche latente Zustandsvariable messen, wird dies nicht immer erreicht. Zwei Parallelformen eines psychologischen Tests bspw. haben neben der intendierten gemeinsamen Komponente in der Regel auch spezifische Komponenten, die sie *nicht* miteinander teilen, auch dann nicht, wenn man sich bemüht, alle Regeln der Testkonstruktion nach der Item-response-Theorie (IRT) einzuhalten und es sich bei den beiden Messungen um Schätzungen des Personenparameters eines IRT-Modells handelt. Die dabei geforderte perfekte Eindimensionalität kommt in der Praxis wohl nur selten vor. Dies kommt z. B. dadurch zustande, dass die beiden Testformen durch verschiedene *Aufgaben* oder *Items* gebildet werden, auf die verschiedene Personen unterschiedlich reagieren. In formaler Hinsicht kann jede der Parallelformen als eine

eigene (mehr oder weniger verfälschte) *Methode* zur Messung der gleichen latenten Zustandsvariablen angesehen werden.

Ein anderes Beispiel sind zwei oder mehr Beobachter, die das Verhalten eines Probanden hinsichtlich eines bestimmten Kriteriums beurteilen sollen. In diesem Fall kann man die Beobachter als *Messinstrumente* oder *Messmethoden* betrachten, deren Urteile die Messungen liefern.

Manchmal liegen auch tatsächlich völlig unterschiedliche Messmethoden vor, die aber dennoch die gleiche latente Variable messen sollen. Ein typisches Beispiel ist die Beurteilung von Eigenschaften eines Probanden durch *Selbstbeurteilung*, *Fremdbeurteilung* durch Peers und *Verhaltensbeobachtung* durch geschultes Personal. Auch hier soll mit drei verschiedenen *Messmethoden* die gleiche latente Zustandsvariable gemessen werden.

Ein ganz ähnliches Design liegt vor, wenn man nicht zu verschiedenen Messzeitpunkten die gleiche latente Zustandsvariable erheben, sondern zu einem einzigen Messzeitpunkt *mehrere Eigenschaften* gleichzeitig, messen will, und das wiederum mit *mehreren Messmethoden* in dem gerade beschriebenen Sinn. Fassen wir diesen Traitbegriff in diesem Kontext nicht im Sinne der LST-Theorie, sondern als Synonym für „latente Variable“ auf, und lassen dabei offen, ob es sich um Latent-state- oder um Latent-trait-Variablen im oben definierten Sinn handelt, dann können wir die damit beschriebenen *Multitrait-multimethod-Modelle* wiederum als spezielle faktoranalytische und damit auch als spezielle Strukturgleichungsmodelle behandeln.

6.1. Das Multitrait-multimethod-Modell

Ausgangspunkt von Multitrait-multimethod-Modellen sind wieder beobachtbare Messungen Y_{it} , die mit der Messmethode i den Trait (allgemeiner: die latente Variable) ξ_t messen sollen. Bezeichnen wir mit ζ_i den jeweiligen Methodenfaktor und mit ε_{it} die Messfehlervariablen, und gehen wir wieder von mittelwertzentrierten Messungen [mit $E(Y_{it}) = 0$] aus, dann lassen sich die Modellgleichungen für die Messwertvariablen Y_{it} beim einfachsten Multitrait-multimethod-Modell wie folgt schreiben:

$$Y_{it} = \xi_t + \zeta_i + \varepsilon_{it}, \quad i = 1, \dots, m, \quad t = 1, \dots, n. \quad [6.1]$$

Die Werte jedes Methodenfaktors ζ_i lassen sich als Interaktionseffekte zwischen Personen und Methoden interpretieren. Hinzu kommen die Annahmen, dass Traits und Methodenfaktoren unkorreliert sind,

$$\text{Cov}(\xi_t, \zeta_i) = 0, \quad i = 1, \dots, m, \quad t = 1, \dots, n, \quad [6.2]$$

dass die Messfehlervariablen mit den Traits und den Methodenfaktoren unkorreliert sind,

$$\text{Cov}(\xi_t, \varepsilon_{it}) = \text{Cov}(\zeta_i, \varepsilon_{it}) = 0, \quad i = 1, \dots, m, \quad t = 1, \dots, n, \quad [6.3]$$

und dass die Methodenfaktoren untereinander unkorreliert sind,

$$\text{Cov}(\zeta_i, \zeta_j) = 0, \quad i, j = 1, \dots, m, \quad i \neq j. \quad [6.4]$$

Dabei beachte man, dass es durchaus auch andere MTMM-Modelle gibt, in denen die Methodenfaktoren miteinander korrelieren können (s. z. B. Widaman, 1993, oder Eid, Lischetzke, Trierweiler, & Nußbeck, 2003). Allerdings braucht man mindestens zwei Methodenfaktoren, die miteinander unkorreliert sind, andernfalls sind Identifikationsprobleme zu befürchten.

In Abbildung 12 ist das Pfaddiagramm dieses *Multitrait-multimethod-Modells* angegeben. Dabei beachte man, dass wir aus Gründen der Einfachheit die Ladungen aller Faktoren auf 1 gesetzt haben. Auf allgemeinere Modelle kommen wir unten zurück.

Die vom oben formulierten Modell implizierte Kovarianzstruktur ist wie folgt:

$$\text{Var}(Y_{it}) = \sigma_{\xi_t}^2 + \sigma_{\zeta_i}^2 + \sigma_{\varepsilon_{it}}^2, \quad [6.5]$$

und

$$\text{Cov}(Y_{it}, Y_{js}) = \begin{cases} \sigma_{\xi_t}^2, & \text{falls } i \neq j \text{ und } t = s \\ \sigma_{\xi_t \xi_s}, & \text{falls } i \neq j \text{ und } t \neq s \\ \sigma_{\xi_t \xi_s} + \sigma_{\zeta_i}^2, & \text{falls } i = j \text{ und } t \neq s. \end{cases} \quad [6.6]$$

Bei $n = 2$ Messgelegenheiten, zu denen jeweils $n = 2$ Messungen vorgenommen werden, hat die von diesem Modell implizierte Varianz-Kovarianzmatrix Σ_{yy} der Messwertvariablen Y_{it} also die Form:

$$\Sigma_{yy} = \begin{pmatrix} \sigma_{\xi_1}^2 + \sigma_{\zeta_1}^2 + \sigma_{\varepsilon_{11}}^2 & \sigma_{\xi_1}^2 & \sigma_{\xi_1 \xi_2} + \sigma_{\zeta_1}^2 & \sigma_{\xi_1 \xi_2} \\ \sigma_{\xi_1}^2 & \sigma_{\xi_1}^2 + \sigma_{\zeta_2}^2 + \sigma_{\varepsilon_{21}}^2 & \sigma_{\xi_1 \xi_2} & \sigma_{\xi_1 \xi_2} + \sigma_{\zeta_2}^2 \\ \sigma_{\xi_1 \xi_2} + \sigma_{\zeta_1}^2 & \sigma_{\xi_1 \xi_2} & \sigma_{\xi_2}^2 + \sigma_{\zeta_1}^2 + \sigma_{\varepsilon_{12}}^2 & \sigma_{\xi_2}^2 \\ \sigma_{\xi_1 \xi_2} & \sigma_{\xi_1 \xi_2} + \sigma_{\zeta_2}^2 & \sigma_{\xi_2}^2 & \sigma_{\xi_2}^2 + \sigma_{\zeta_2}^2 + \sigma_{\varepsilon_{22}}^2 \end{pmatrix}. \quad [6.7]$$

Anhand dieser Matrix oder auch der Gleichungen [6.5] und [6.6] kann man sich nun leicht überlegen, wie man die theoretischen Parameter dieses sehr einfachen MTMM-Modells aus den empirisch schätzbaren Varianzen und Kovarianzen der Messungen Y_{it} bestimmen kann. So ist z. B. die Varianz des Traits ξ_1 gleich der Kovarianz der zwischen Y_{11} und Y_{21} . Die Kovarianz zwischen den beiden Traitvariablen ξ_1 und ξ_2 ist gleich der Kovarianz zwischen Y_{11} und Y_{22} . Die Varianz des Methodenfaktors ζ_1 ist gleich der Kovarianz zwischen Y_{11} und Y_{12} minus der Kovarianz zwischen Y_{11} und Y_{22} etc.

6.2. Weiterführende Literatur

In den bisherigen Modellen haben wir aus didaktischen Gründen explizit nur Ladungen betrachtet, die gleich 1 waren. Im Allgemeinen muss man aber davon ausgehen, dass die Ladungen durchaus unterschiedlich sein können. Insbesondere wenn ganz unterschiedliche Messmethoden mit ganz unterschiedlichen Skalierungen als Methoden zur Messung der gleichen latenten Variablen zum Einsatz kommen, werden die Ladungen in aller Regel sehr unterschiedlich sein. In diesem Fall wäre dann übrigens auch ein Modell mit 2 Traits und 2 Methodenfaktoren nicht identifiziert, da dies dann nicht nur 9, sondern 13 unbekannte theoretische Parameter hätte, die man aus 10 Gleichungen für die Varianzen und Kovarianzen der Messwertvariablen zu berechnen hätte. Dies ist aber unmöglich, sofern man nicht z. B. die Zahl der unbekannt Parameter durch Gleichsetzung auf höchstens 10 reduziert.

Verschiedene Arten von MTMM-Modellen werden von Kenny und Kashy (1992), Marsh (1989), Marsh und Grayson (1995), Widaman (1985) sowie von Wothke (1996) diskutiert. Grayson und Marsh (1994) betrachten Probleme der Identifizierbarkeit bei MTMM-Modellen.

Besonders empfohlen seien auch das Modell von Eid (2000) und seine Verallgemeinerung (2003).

7. Faktorenanalytische Modelle

In den vorangegangenen Abschnitten haben wir bereits *spezielle* faktorenanalytische Modelle kennen gelernt. Faktorenanalytische Modelle kann man als spezielle Messmodelle auffassen, in denen aus *manifesten Variablen* auf dahinter liegende *latente Variablen* oder *Faktoren* geschlossen werden kann, insbesondere was deren Varianzen, Kovarianzen und Korrelationen angeht. In speziellen Fällen, etwa bei längsschnittlichen und/oder Mehrgruppenmodellen, kann man aber auch auf deren Erwartungswerte schließen. Für diagnostische Zwecke ist es auch möglich und sinnvoll, die Werte von Personen auf den latenten Variablen, die *Faktorwerte*, zu schätzen. Für Zusammenhangs- und Abhängigkeitsanalysen dagegen werden die Schätzungen der Faktorwerte nicht benötigt, da man diese Zusammenhänge und Abhängigkeiten besser und genauer im Rahmen von Strukturgleichungsmodellen untersuchen kann, die als eine Verallgemeinerung faktorenanalytischer Modelle angesehen werden können (s. Abschnitt 9).

Faktorenanalytische Modelle bestehen aus einer Modellgleichung

$$\mathbf{y} = \boldsymbol{\lambda}_0 + \mathbf{\Lambda} \boldsymbol{\eta} + \boldsymbol{\varepsilon} \quad [7.1]$$

für die Messwertvariablen des Spaltenvektors $\mathbf{y} = (Y_1 \dots Y_m)'$, sowie aus Annahmen über die Kovarianzmatrix $\boldsymbol{\Sigma}_{\boldsymbol{\eta}\boldsymbol{\eta}}$ der *latenten Variablen* oder *Faktoren* des Vektors $\boldsymbol{\eta} = (\eta_1 \dots \eta_m)'$ und über die Kovarianzmatrix $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}}$ der *Messfehlervariablen*, die in $\boldsymbol{\varepsilon} = (\varepsilon_1 \dots \varepsilon_m)'$ zusammengefasst sind. Oft geht man von zentrierten manifesten und latenten Variablen aus, die definitionsgemäß alle den Erwartungswert 0 haben. In diesem Fall entfällt der Konstantenvektor $\boldsymbol{\lambda}_0$, da dann gilt: $\boldsymbol{\lambda}_0 = \mathbf{0}$. Die Matrix $\mathbf{\Lambda}$ heißt *Ladungsmatrix* und ihre Komponenten, die Ladungen λ_{ij} , lassen sich als Regressionskoeffizienten in einer multiplen linearen Regression der Messwertvariablen Y_i auf die Faktoren η_1, \dots, η_m interpretieren (s. dazu Steyer, 2003, Kap. 9). In speziellen Modellen kann die Gleichung [7.1] als eine Matrixdarstellung von m Regressionsgleichungen interpretiert werden, so etwa bei Modellen der KTT und bei LST-Modellen, aber auch immer dann, wenn die Messwertvariablen als m -variater normalverteilt angenommen werden. In anderen Fällen ist Gleichung [7.1] eine Matrixdarstellung von m linearen Quasi-Regressionen (s. Steyer, 2003, Kap. 14). Die multiple lineare Regression ist daher grundlegend auch für das Verständnis der Faktorenanalyse. Der Determinationskoeffizient $R_{y_i|\boldsymbol{\eta}}^2$ heißt im Rahmen des faktorenanalytischen Modells *Kommunalität*.

Generell wird angenommen, dass die Messfehler und Faktoren unkorreliert sind:

$$\text{Cov}(\boldsymbol{\eta}, \boldsymbol{\varepsilon}) = \mathbf{0}. \quad [7.2]$$

Für die Kovarianzmatrix der Messwertvariablen ergibt sich daraus

$$\boldsymbol{\Sigma}_{\mathbf{y}\mathbf{y}} = \mathbf{\Lambda} \boldsymbol{\Sigma}_{\boldsymbol{\eta}\boldsymbol{\eta}} \mathbf{\Lambda}' + \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}}. \quad [7.3]$$

Eine weitere Vereinfachung ergibt sich, wenn man die *Unkorreliertheit der Messfehler* annimmt, dass also die Kovarianzmatrix $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}}$ der Messfehler eine Diagonalmatrix ist.

Beispiele für solche Modelle haben wir in den vorangegangenen Abschnitten schon kennen gelernt. Der Einfachheit halber haben wir bei fast allen bisher explizit dargestellten Modellen angenommen, dass alle Ladungen entweder gleich 0 oder 1 sind. Alle diese Modelle lassen

sich aber dahingehend verallgemeinern, dass die Ladungen auch andere Zahlen sein können. Allerdings ist dann immer zu untersuchen, ob die unbekanntes theoretischen Parameter noch aus den Varianzen und Kovarianzen der Messwertvariablen (sowie aus deren Erwartungswerten, wenn im Modell auch Lokationsparameter vorkommen) eindeutig bestimmt werden (identifiziert) können. Beim MTMM-Modell hatten wir schon darauf hingewiesen, dass ein Modell mit ungleichen Ladungen bei nur 2 Traits und 2 Methoden nicht identifiziert wäre.

Bei den bisher behandelten faktorenanalytischen Modellen handelte es sich um Beispiele für *konfirmatorische Modelle*. Die Ladungen und auch einige Kovarianzen zwischen den latenten Variablen sind im Modell schon festgelegt. In diesem Kontext spricht man daher auch von *fixierten Parametern*. In anderen konfirmatorischen Modellen ist es auch bei den Ladungen so, dass einige festgelegt sind, andere aber nicht. Für diese nicht festgelegten oder *freien Parameter* kann man auch Gleichheits-, lineare und nichtlineare Restriktionen setzen. Die freien Parameter können dann in diesen Modellen geschätzt werden. Voraussetzung dabei ist, dass sie *identifizierbar* sind, d. h. in dem betreffenden Modell eindeutig aus den Erwartungswerten, Varianzen und Kovarianzen der Messwertvariablen berechenbar sind.

In *exploratorischen Modellen* ist zunächst weder etwas über die Kovarianz der Faktoren, noch über die Ladungsmatrix bekannt. Erst über bestimmte Extraktions- (z. B. Maximierung der Varianz der Faktoren bei gleichzeitiger Orthogonalität) bzw. Rotationskriterien (z. B. Varimax, Oblimin oder Promax) werden die Ladungen und Korrelationen zwischen den Faktoren identifizierbar (d. h. eindeutig bestimmbar), und können dann auch geschätzt werden. (Dabei geht man in der Regel von der z -Standardisierung der manifesten und latenten Variablen aus, die dann alle Erwartungswert 0 und Varianz 1 haben.) Bei konfirmatorischen Modellen dagegen entfällt das Rotationsproblem. Dort sind die Faktoren schon durch die vorgegebene Struktur der Ladungsmatrix eindeutig bestimmt, sofern das Modell überhaupt identifiziert ist.

In den folgenden Abschnitten werden wir die bereits behandelten Beispiele als spezielle faktorenanalytische Modelle in Matrizenform darstellen.

7.1. Darstellung des Modells τ -äquivalenter Variablen in Matrizenform

Hier handelt es sich nun um ein erstes, sehr einfaches faktorenanalytisches und auch um ein sehr einfaches Strukturgleichungsmodell, wie wir später sehen werden. Wie oben beschrieben, lässt sich ein faktorenanalytisches Modell immer durch die Modellgleichung [7.1] für die Variablen sowie die Modellgleichung [7.3] für deren Kovarianzmatrix darstellen.

Im vorliegenden Beispiel [s. die Modellgleichung [3.4]] gilt $\lambda_0 = \mathbf{0}$ und die *Ladungsmatrix* $\Lambda = (1 \ 1 \ \dots \ 1)'$ ist ein Spaltenvektor von Einsen. Der Vektor $\eta = (\eta)$ besteht hier nur aus einer einzigen latenten Variablen und in $\epsilon = (\epsilon_1 \ \dots \ \epsilon_m)'$ sind die Messfehlervariablen zusammengefasst. Die Kovarianzstruktur der Messwertvariablen, die wir oben in den Gleichungen [3.6] und [3.7], sowie für den Fall von $m = 4$ Messwertvariablen durch die Gleichung [3.10] dargestellt haben, hat die Form der Gleichung [7.3]. Im vorliegenden Spezialfall hat die Kovarianzmatrix $\Sigma_{\eta\eta}$ der latenten Variablen nur die einzige Komponente, σ_η^2 und $\Sigma_{\epsilon\epsilon}$ ist die (4×4) -Diagonalmatrix mit den Fehlervarianzen $\sigma_{\epsilon_i}^2$ in der Diagonalen.

Andere Modelle gehen nicht von $\tau_i = \tau_j =: \eta$ (s. Gl. [3.4]) aus, sondern von $\tau_i = \eta + \lambda_i$ (*Modell essentiell τ -äquivalenter Variablen*) bzw. von $\tau_i = \lambda_{j1}\eta + \lambda_{j0}$, (*Modell τ -kongenerischer Variablen*), wobei die λ -Koeffizienten reelle Zahlen sind. In diesen Modellen wird auf eine andere Weise präzisiert, was es heißt, dass die Messwertvariablen Y_i das Gleiche messen. Die oben aufgeführten Gleichungen für die vom Modell implizierte Kovarianzstruktur und für die Identifikation der theoretischen Parameter ändern sich nicht für das Modell essentiell τ -äquivalenter Variablen, wohl aber für das Modell τ -kongenerischer Variablen (s. z. B. Steyer & Eid, 2001, Kap. 13).

7.2. Das Multistate-Modell in Matrizenform

Auch hier handelt es sich wieder um ein relativ einfaches faktorenanalytisches Modell, d. h. es gilt wieder die Modellgleichung [7.1] für die Variablen und die Modellgleichung [7.3] für die Kovarianzmatrix der Messwertvariablen. Im Fall von $m = 2$ Messwertvariablen, die zu jeder von $n = 2$ Messgelegenheiten erhoben werden, sind $\mathbf{y} = (Y_{11} Y_{21} Y_{12} Y_{22})'$ der Spaltenvektor der *Observablen*, $\boldsymbol{\lambda}_0 = \mathbf{0}$ ein Konstantenvektor, und die *Ladungsmatrix* $\mathbf{\Lambda}$ hat $m \times n = 2 \times 2 = 4$ Zeilen und $n = 2$ Spalten, genau so viele, wie Messzeitpunkte vorliegen. Für die Komponenten λ_{it} von $\mathbf{\Lambda}$ gilt:

$$\mathbf{\Lambda} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix}. \quad [7.4]$$

Der Vektor $\boldsymbol{\eta} = (\eta_1 \eta_2)'$ besteht nun aus zwei latenten Variablen und in $\boldsymbol{\varepsilon} = (\varepsilon_{11} \varepsilon_{21} \varepsilon_{12} \varepsilon_{22})'$ sind wieder die Messfehlervariablen zusammengefasst. Die Kovarianzstruktur der Messwertvariablen, die wir oben in den Gleichungen [4.13] und [4.14], sowie für den Fall von $m = 2$ Messwertvariablen, die zu jeder von $n = 2$ Messzeitpunkten erhoben werden, durch die Gleichung [4.15] dargestellt haben, hat die Form der Gleichung [7.3]. Im vorliegenden Spezialfall hat die Kovarianzmatrix

$$\boldsymbol{\Sigma}_{\boldsymbol{\eta}\boldsymbol{\eta}} = \begin{pmatrix} \sigma_{\eta_1}^2 & \sigma_{\eta_1\eta_2} \\ \sigma_{\eta_2\eta_1} & \sigma_{\eta_2}^2 \end{pmatrix} \quad [7.5]$$

der latenten Variablen vier Komponenten, wobei natürlich $\sigma_{\eta_1\eta_2} = \sigma_{\eta_2\eta_1}$ und

$$\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}} = \begin{pmatrix} \sigma_{\varepsilon_{11}}^2 & 0 & 0 & 0 \\ 0 & \sigma_{\varepsilon_{21}}^2 & 0 & 0 \\ 0 & 0 & \sigma_{\varepsilon_{12}}^2 & 0 \\ 0 & 0 & 0 & \sigma_{\varepsilon_{22}}^2 \end{pmatrix} \quad [7.6]$$

ist eine (4×4) -Diagonalmatrix mit den Fehlervarianzen $\sigma_{\varepsilon_{it}}^2$ in der Hauptdiagonalen.

7.3. Das Multistate-Singletrait-Modell in Matrizenform

Betrachten wir die Gleichung [4.19], so erkennt man, dass es sich auch hier wieder um ein faktorenanalytisches Modell handelt, d. h. es gilt wieder die Modellgleichung [7.1] für die Variablen sowie die Modellgleichung [7.3] für die Kovarianzmatrix der Messwertvariablen. Im Fall von $m = 2$ Messwertvariablen, die zu jeder von $n = 2$ Messgelegenheiten erhoben werden, sind $\mathbf{y} = (Y_{11} Y_{21} Y_{12} Y_{22})'$ der Spaltenvektor der Messwertvariablen, $\boldsymbol{\lambda}_0 = \mathbf{0}$ ein Konstantenvektor, und die *Ladungsmatrix* $\mathbf{\Lambda}$ hat $m \times n = 2 \times 2 = 4$ Zeilen und 3 Spalten, genau so viele, wie Messzeitpunkte vorliegen. Für die Komponenten λ_{it} von $\mathbf{\Lambda}$ gilt:

$$\mathbf{\Lambda} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix}. \quad [7.7]$$

Der Vektor $\boldsymbol{\eta} = (\xi \zeta_1 \zeta_2)'$ besteht nun aus drei latenten Variablen und in $\boldsymbol{\varepsilon} = (\varepsilon_{11} \varepsilon_{21} \varepsilon_{12} \varepsilon_{22})'$ sind wieder die Messfehlervariablen zusammengefasst. Die Kovarianzstruktur der Messwertvariablen, die wir oben in den Gleichungen [4.22] und [4.23], sowie für den Fall von $m = 2$ Messwertvariablen, die zu jeder von $n = 2$ Messzeitpunkten erhoben werden, durch die Gleichung [4.24] dargestellt haben, hat die Form der Gleichung [7.3]. Im vorliegenden Spezialfall hat die Kovarianzmatrix die Form

$$\boldsymbol{\Sigma}_{\boldsymbol{\eta}\boldsymbol{\eta}} = \begin{pmatrix} \sigma_{\xi}^2 & 0 & 0 \\ 0 & \sigma_{\zeta_1}^2 & 0 \\ 0 & 0 & \sigma_{\zeta_2}^2 \end{pmatrix} \quad [7.8]$$

und

$$\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}} = \begin{pmatrix} \sigma_{\varepsilon_{11}}^2 & 0 & 0 & 0 \\ 0 & \sigma_{\varepsilon_{21}}^2 & 0 & 0 \\ 0 & 0 & \sigma_{\varepsilon_{12}}^2 & 0 \\ 0 & 0 & 0 & \sigma_{\varepsilon_{22}}^2 \end{pmatrix}. \quad [7.9]$$

ist wieder die (4×4) -Diagonalmatrix mit den Fehlervarianzen $\sigma_{\varepsilon_{it}}^2$ in der Diagonalen.

7.4. Das Multitrait-multimethod-Modell in Matrizenform

Auch hier handelt es sich wieder um ein faktorenanalytisches Modell, d. h. es gilt wieder die Modellgleichung [7.1] für die Variablen sowie die Modellgleichung [7.3] für die Kovarianzmatrix der Messwertvariablen. Im Fall von $m = 2$ Messmethoden, mit denen jeweils $n = 2$ Traits erhoben werden, sind $\mathbf{y} = (Y_{11} Y_{21} Y_{12} Y_{22})'$ der Spaltenvektor der Messwertvariablen, $\boldsymbol{\lambda}_0 = \mathbf{0}$ ein Konstantenvektor, und die *Ladungsmatrix* $\mathbf{\Lambda}$ hat $m \times n = 2 \times 2 = 4$ Zeilen und 4 Spalten, genau so viele, wie latente Variablen vorliegen. Für die Komponenten von $\mathbf{\Lambda}$ gilt:

$$\mathbf{\Lambda} = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}. \quad [7.10]$$

Der Vektor $\boldsymbol{\eta} = (\xi \zeta_2 \zeta_1 \zeta_2)'$ besteht nun aus vier latenten Variablen, den beiden Traits und den beiden Methodenfaktoren, und in $\boldsymbol{\varepsilon} = (\varepsilon_{11} \varepsilon_{21} \varepsilon_{12} \varepsilon_{22})'$ sind wieder die Messfehlervariablen zusammengefasst. Die Kovarianzstruktur der Messwertvariablen, die wir oben in den Gleichungen [6.5] und [6.6], sowie für den Fall von $m = 2$ Messmethoden und $n = 2$ Traits,

durch die Gleichung [6.7] dargestellt haben, hat die Form der Gleichung [7.3]. Im vorliegenden Spezialfall hat die Kovarianzmatrix der latenten Variablen die Form

$$\Sigma_{\eta\eta} = \begin{pmatrix} \sigma_{\xi_1}^2 & \sigma_{\xi_1\xi_2} & 0 & 0 \\ \sigma_{\xi_1\xi_2} & \sigma_{\xi_2}^2 & 0 & 0 \\ 0 & 0 & \sigma_{\zeta_1}^2 & 0 \\ 0 & 0 & 0 & \sigma_{\zeta_2}^2 \end{pmatrix} \quad [7.11]$$

und

$$\Sigma_{\varepsilon\varepsilon} = \begin{pmatrix} \sigma_{\varepsilon_{11}}^2 & 0 & 0 & 0 \\ 0 & \sigma_{\varepsilon_{21}}^2 & 0 & 0 \\ 0 & 0 & \sigma_{\varepsilon_{12}}^2 & 0 \\ 0 & 0 & 0 & \sigma_{\varepsilon_{22}}^2 \end{pmatrix}. \quad [7.12]$$

ist wieder die (4×4) -Diagonalmatrix mit den Fehlervarianzen $\sigma_{\varepsilon_{it}}^2$ in der Diagonalen.

7.5. Weiterführende Literatur

Ein klassisches Buch zur Faktorenanalyse ist Lawley und Maxwell (1971). Bei Anderson und Rubin (1956) findet man u. a. wichtige Hinweise zur Identifizierbarkeit von Modellen. Klassische Artikel zur konfirmatorischen Faktorenanalyse sind Jöreskog (1969, 1971). Zur Unterscheidung zwischen Faktoren- und Hauptkomponentenanalyse siehe Jöreskog (1979). Zur Faktorenanalyse qualitativer Variablen sei auf Mislevy (1986), Muthén, (1978) sowie auf Muthén und Christoffersson (1981) hingewiesen. Ein deutschsprachiges Lehrbuch zur Hauptkomponentenanalyse und verwandte Verfahren ist Röhr (1993). Weitere Lehrbücher zu verschiedenen Verfahren der Faktoren- und Hauptkomponentenanalyse sind Überla (1971) und Revenstorf (1980). PC-Programme zur exploratorischen Faktorenanalyse enthalten alle gängigen Software-Pakete zur Statistik. Konfirmatorische faktorenanalytische Modelle können mit den gängigen Programmen zur Analyse von Strukturgleichungsmodellen gerechnet werden.

8. Strukturgleichungsmodelle

Strukturgleichungsmodelle können als eine Erweiterung faktorenanalytischer Modelle angesehen werden, die es u. a. auch ermöglichen, multiple lineare Regressionen zwischen den latenten Variablen zu betrachten. Es gibt verschiedene Notationssysteme zur Beschreibung von Strukturgleichungsmodellen. Das bekannteste wurde von Jöreskog (1970, 1973) eingeführt. Beschränken wir uns auf zentrierte Variablen (mit Erwartungswerten gleich null), besteht es aus den drei Modellgleichungen

$$\mathbf{y} = \Lambda_y \boldsymbol{\eta} + \boldsymbol{\varepsilon}, \quad [8.1]$$

$$\mathbf{x} = \Lambda_x \boldsymbol{\xi} + \boldsymbol{\delta}, \quad [8.2]$$

$$\boldsymbol{\eta} = \mathbf{B} \boldsymbol{\eta} + \boldsymbol{\Gamma} \boldsymbol{\xi} + \boldsymbol{\zeta}, \quad [8.3]$$

den Unkorreliertheitsannahmen

$$\text{Cov}(\boldsymbol{\eta}, \boldsymbol{\varepsilon}) = \mathbf{0}, \quad \text{Cov}(\boldsymbol{\xi}, \boldsymbol{\delta}) = \mathbf{0}, \quad \text{Cov}(\boldsymbol{\xi}, \boldsymbol{\zeta}) = \mathbf{0} \quad [8.4]$$

und

$$\text{Cov}(\boldsymbol{\varepsilon}, \boldsymbol{\delta}) = \mathbf{0}, \quad \text{Cov}(\boldsymbol{\varepsilon}, \boldsymbol{\zeta}) = \mathbf{0}, \quad \text{Cov}(\boldsymbol{\delta}, \boldsymbol{\zeta}) = \mathbf{0}, \quad [8.5]$$

sowie der Annahme, dass die Matrix $\mathbf{I} - \mathbf{B}$ regulär ist, wobei \mathbf{I} die Einheitsmatrix darstellt.

Bei den beiden Gleichungen [8.1] und [8.2] handelt es sich um Messmodelle für die *endogenen* latenten Variablen im Vektor $\boldsymbol{\eta}$ bzw. für die *exogenen* latenten Variablen im Vektor $\boldsymbol{\xi}$. Die Strukturgleichung [8.3] dagegen erlaubt die Abhängigkeiten der latenten Variablen untereinander zu spezifizieren. Die *exogenen* latenten Variablen im Vektor $\boldsymbol{\xi}$ können untereinander korrelieren, werden aber ansonsten im Modell ausschließlich als unabhängige Variablen behandelt. Die *endogenen* latenten Variablen im Vektor $\boldsymbol{\eta}$ dagegen können gleichzeitig abhängig und unabhängig sein. Strukturgleichungen eignen sich daher auch in besonderer Weise zur Darstellung kausaler Systeme, da man mit ihnen u. a. kausale Ketten darstellen sowie zwischen direkten, indirekten und totalen Effekten unterscheiden kann. Nicht jedes Strukturgleichungsmodell erlaubt jedoch kausale Interpretationen, genauso wenig wie jedes Regressionsmodell kausale Abhängigkeiten darstellt. Bei Strukturgleichungsmodellen, wie bei Regressionsmodellen, entscheiden zusätzliche Annahmen darüber, ob die damit beschriebenen Abhängigkeiten kausal interpretiert werden können (s. z. B. Abschnitt 6, Steyer, 2003, Kap. 15 bis 17 sowie Steyer, 1992 und Steyer, Gabler, von Davier, Nachtigall, & Buhl, 2000b, Steyer, Gabler, von Davier, & Nachtigall, 2000a, Steyer, Nachtigall, Wüthrich-Martone, & Kraus, 2002).

In Anwendungen sind die Komponenten der Matrizen $\boldsymbol{\Lambda}_y$, $\boldsymbol{\Lambda}_x$, \mathbf{B} und $\boldsymbol{\Gamma}$, die Kovarianzmatrizen der Messfehlervariablen $\boldsymbol{\varepsilon}$ und $\boldsymbol{\delta}$, sowie die Kovarianzmatrix der latenten Variablen im Vektor $\boldsymbol{\xi}$ und die Kovarianzmatrix der Residualvariablen im Vektor $\boldsymbol{\zeta}$ zu spezifizieren, und zwar wiederum durch die Angabe, welche Parameter auf welchen Werten fixiert, welche dagegen frei zu schätzen und welche, z. B. durch Gleichheitsrestriktionen, restringiert zu schätzen sind. Diese Annahmen implizieren, genau wie das oben dargestellte Modell der Faktorenanalyse, eine bestimmte Struktur der Kovarianzmatrix der Messwertvariablen in \mathbf{x} und \mathbf{y} , die in Anwendungen mit der jeweils vorliegenden empirischen Kovarianzmatrix dieser Messwertvariablen verglichen werden, um die Passung des Modells zu überprüfen.

8.1. Parameterschätzung

Nach der Spezifizierung des theoretischen Modells und dem Vorliegen der empirischen Kovarianzmatrix müssen im nächsten Schritt die Modellparameter geschätzt und die Passung des Modells für die beobachteten Daten überprüft werden.

Für die Parameterschätzung gibt es verschiedene Methoden, die sich in den Verteilungsannahmen und der zu minimierenden Abweichungsfunktion unterscheiden. Generell unterscheidet man dabei *Maximum-likelihood-* und *Least-squares-basierte* Schätzmethoden. Im Folgenden werden beide Klassen von Schätzmethoden mit ihren jeweiligen Anwendungsbereichen sowie Vor- und Nachteilen vorgestellt (für eine ausführliche Beschreibung der Schätzmethoden siehe Bollen, 1989; Kaplan, 2000; einen guten Überblick geben auch Schermelleh-Engel, Moosbrugger, & Müller, 2003).

Das *Maximum-likelihood-Verfahren* (ML) ist die am häufigsten verwendete Schätz-
methode für die Parameter eines Strukturgleichungsmodells. ML schätzt den Vektor der
Modellparameter $\boldsymbol{\theta}$, indem er die Wahrscheinlichkeit L maximiert, dass die empirische
Kovarianzmatrix S aus einer Population gezogen wurde, für welche die vom Modell
implizierte Kovarianzmatrix $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ gilt. Anstelle von L kann man auch die Log-Likelihood-
Funktion

$$\ln L = -0.5 (N - 1) (\ln |\boldsymbol{\Sigma}(\boldsymbol{\theta})| + \text{tr}[S \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}]) + c, \quad [8.6],$$

bzw. die folgende Funktion (unter Verwendung der empirisch geschätzten Modellparameter)

$$F_{\text{ML}}[S, \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})] = \ln |\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})| - \ln |S| + \text{tr}[S \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})^{-1}] - p, \quad [8.7],$$

minimieren (s. z. B. Bollen, 1989, S. 135) mit $\ln L$ als natürlichen Logarithmus der Likeli-
hoodfunktion, der Stichprobengröße N , dem Parametervektor $\boldsymbol{\theta}$, der vom Modell implizierten
Kovarianzmatrix $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ und seiner Determinante $|\boldsymbol{\Sigma}(\boldsymbol{\theta})|$, der Spur der Matrix tr , einer Kon-
stanten c (ein Parameter der zugrunde liegenden Wishart-Verteilung, der in einer gegebenen
Stichprobe konstant ist), dem Wert der Fitting-Funktion F_{ML} und der Anzahl der beobachteten
Variablen p .

Der Maximum-likelihood-Algorithmus setzt die multivariate Normalverteilung der im
Modell vorkommenden manifesten Variablen und die Invertierbarkeit der Kovarianzmatrizen
 S und $\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$ voraus. Bei gegebenen Voraussetzungen, einer korrekten Spezifizierung des
Modells sowie einer ausreichend großen Stichprobengröße liefert der ML-Algorithmus
asymptotisch unverfälschte, konsistente und effiziente Schätzungen der Parameter und ihrer
Standardfehler (Bollen, 1989). Ein klarer Vorteil des ML-Schätzers ist die Möglichkeit der
direkten Testbarkeit des Gesamtmodells bei überidentifizierten Modellen, d. h. Modellen, bei
denen die Anzahl der unabhängigen empirischen Parameter (alle Varianzen und Kovarianzen
der Messwertvariablen), größer ist als die Zahl der unabhängigen, im Modell zu schätzenden
Parameter ($df > 0$). Die Modellbewertung mithilfe der ML-Funktion ist im Abschnitt 9.2.
beschrieben.

Ein entscheidender Nachteil des ML-Algorithmus ist die Annahme der multivariaten Nor-
malverteilung aller Messwertvariablen. Die Verletzung dieser Annahme führt zu potentiell
verzerrten Ergebnissen bei der Modellbewertung mithilfe des ML-Schätzers im Sinne eines
steigenden α -Fehlers. Zwar scheint der ML-Schätzer recht robust bei weniger extremen Ab-
weichungen von der Normalverteilung (z. B. Boomsma & Hoogland, 2001; Curran, West, &
Finch, 1996), dennoch wird bei Verletzung der Verteilungsannahmen die *Satorra-Bentler*
Korrektur des ML-Schätzers empfohlen (Satorra & Bentler, 1994). Dieser Algorithmus be-
rücksichtigt die Momente 3. und 4. Ordnung (Schiefe und Kurtosis) und führt auch bei
extremen Verletzungen der Verteilungsannahmen zu einer unverfälschten Teststatistik mit
guten statistischen Eigenschaften im Vergleich sowohl zum unkorrigierten ML-Schätzer als
auch zu *Least Square* basierten Schätzern (Chou & Bentler, 1995; Hu & Bentler, 1995;
Hoogland, 1999).

Alternativ zur ML Schätzung finden verschiedenen *Least Square* (LS) basierte Schätz-
methoden im Rahmen von Strukturgleichungsmodellen Verwendung, besonders bei vorlie-
genden kontinuierlichen aber nicht-normalverteilten Daten (auch wenn Simulationsstudien
darauf hinweisen, dass auch in diesen Fällen ML-Schätzer mit oder ohne Korrekturen bessere
Ergebnisse im Sinne von unverfälschten und effizienten Schätzern geben; siehe Boomsma &
Hoogland, 2001, und Olsson, Foss, Troye, & Howell, 2000). LS-basierte Schätzer sind
asymptotisch verteilungsfreie Schätzer, die sowohl bei nicht-normaler Verteilung der Obser-
variablen als auch bei ordinalen und dichotomen Variablen im Modell angewendet werden kön-

nen. Das allgemeinste Schätzverfahren ist das *Weighted-Least-Squares* Verfahren (WLS), welches die Funktion

$$F_{\text{WLS}} = [s - \sigma(\hat{\theta})]' \mathbf{W}^{-1} [s - \sigma(\hat{\theta})] \quad [8.8]$$

minimiert, mit s als Vektor nicht-redundanter Elemente der empirischen Kovarianzmatrix, $\sigma(\hat{\theta})$ als Vektor unabhängiger Elemente der vom Modell implizierten Kovarianzmatrix, dem Parametervektor $\hat{\theta}$ der Größe $(t \times 1)$ und der $(k \times k)$ Gewichtsmatrix \mathbf{W}^{-1} mit $k = p(p+1)/2$ und p als Anzahl der Messwertvariablen. Vorteile des WLS-Schätzers sind die minimalen Annahmen über die Verteilung der Messwertvariablen bzw. die Unverfälschtheit des Schätzers bei nicht-normalen Daten und ausreichend großer Stichprobe (Hoogland & Boomsma, 1998; West, Finch, & Curran, 1995) und die Möglichkeit der direkten Analyse von Korrelationsmatrizen (wobei die Matrix \mathbf{W} die Kovarianzen der Korrelationen enthalten muss). Der WLS-Algorithmus ist auch unter dem Synonym ADF (*asymptotically distribution free*) bekannt.

Ein großer Nachteil des WLS-Schätzers ist die erforderliche Stichprobengröße bei zunehmender Zahl von Messwertvariablen. Bereits bei 10 Messwertvariablen enthält die asymptotische Kovarianzmatrix eine Matrix vom Typ 55×55 mit 1540 nicht-redundanten Elementen (s. Schermelleh-Engel et al., 2003), deren Schätzung eine vielfach größere Stichprobe erfordert. Daher wird bei komplexen Modellen mit kleinerer Stichprobengröße und weniger starker Verletzungen der Verteilungsannahmen im Allgemeinen der ML-Schätzer empfohlen.

Als Spezialfälle des WLS-Schätzers seien außerdem die *General Least Square* (GLS) und *Unweighted Least Square* (ULS) Schätzer erwähnt. Unter Annahme der multivariaten Normalverteilung, vereinfacht sich der WLS-Schätzer, da die höheren Momente in der Matrix \mathbf{W} nun nicht mehr berücksichtigt werden. Die Abweichungsfunktion des GLS-Schätzer hat dann die Form

$$F_{\text{GLS}} = 0.5 \text{tr} \{ [S - \Sigma(\hat{\theta})] S^{-1} \}^2 \quad [8.9]$$

mit tr als Spur der Matrix, der empirischen Kovarianzmatrix S und deren Inverse S^{-1} , der vom Modell implizierten Kovarianzmatrix $\Sigma(\hat{\theta})$ und dem $(t \times 1)$ Vektor der Modellparameter $\hat{\theta}$. Der GLS-Schätzer ist asymptotisch äquivalent zum ML-Schätzer. Da beide Schätzer dieselben Voraussetzungen hinsichtlich der Verteilung der Messwertvariablen haben, wird der ML-Schätzer aufgrund der besseren Effizienz – besonders in kleinen Stichproben – meist bevorzugt (Chou & Bentler, 1995; Hu & Bentler, 1995; Hoogland, 1999).

Eine weitere Vereinfachung Least Square basierter Schätzer ergibt sich beim *Unweighted Least Square* (ULS) Verfahren, dessen Abweichungsfunktion

$$F_{\text{ULS}} = 0.5 \text{tr} [S - \Sigma(\hat{\theta})]^2 \quad [8.10]$$

ist, wobei im Vergleich zum GLS-Schätzer die Inverse der empirischen Kovarianzmatrix S^{-1} durch die Identitätsmatrix \mathbf{I} ersetzt wird. Der ULS-Schätzer führt im Vergleich zu ML und WLS zu konsistenten Schätzern der Modellparameter $\hat{\theta}$, hat aber den Nachteil einer weniger effizienten Schätzung der Modellparameter und der fehlenden Skalenfreiheit bzw. Skaleninvarianz der Schätzer (Bollen, 1989).

Neben den vorgestellten gibt es weitere Schätzverfahren, z. B. *Ordinary Least Squares* (OLS), *Two-Stage Least Squares* (2SLS) und *Three-Stage Least Squares* (3SLS). Einen Überblick über diese Methoden findet sich z. B. bei Johnston (1984) und Fox (1984). Aufgrund der zunehmenden Bedeutung im Rahmen der Analyse von Strukturgleichungsmodellen mit fehlenden Werten (*Missing Data*) sei abschließend die Modellschätzung mittels *Full Information*

Maximum Likelihood (FIML; Finkbeiner, 1979) erwähnt. Dieser auch als *Raw Maximum Likelihood* bekannte Ansatz ist eine Realisation des allgemeinen ML-Schätzverfahrens, wobei die zu minimierende Wahrscheinlichkeitsfunktion – unter Voraussetzung der multivariaten Normalverteilung – für jeden Fall (*casewise*) spezifiziert wird. Die Wahrscheinlichkeitsfunktion

$$\ln L_i = K_i - 0.5\{\ln |\Sigma_i| - 0.5(\mathbf{X}_i - \boldsymbol{\mu}_i)' \Sigma_i^{-1} (\mathbf{X}_i - \boldsymbol{\mu}_i)\} \quad [8.11],$$

benötigt daher den Vektor \mathbf{X}_i der Daten für alle Personen i und den zu \mathbf{X}_i gehörenden Vektor der geschätzten Mittelwerte $\boldsymbol{\mu}_i$ für die gesamte Stichprobe basierend auf den beobachteten Daten aller Personen i (K_i ist eine von der Anzahl vorhandener Variablen abhängige Konstante). Die einzelnen Gleichungen werden zu einer logarithmierten Wahrscheinlichkeit $\ln L$ zusammengefasst (entspricht der Summe der einzelnen $\ln L_i$), welche dann minimiert wird. Einzelheiten zu diesem Schätzverfahren finden sich bei Neale (2000), Jamshidian und Bentler (1999) sowie Wothke (2000).

8.2. Modellbewertung

Nach der Konvergenz des Schätzalgorithmus und der damit verbundenen Schätzung der Modellparameter wird die Passung des Modells (Fit) überprüft. Die Beurteilung der Passung eines Strukturgleichungsmodells kann auf zwei abstrakten Ebenen geschehen, der a) Beurteilung der globalen Modellpassung (*overall fit*) mit nachfolgender Evaluation einzelner Parameterschätzungen (*component fit*) des spezifizierten Modells sowie b) dem Vergleich verschiedener Modelle (*model comparison*), was die Beurteilung des Modells unter Einbeziehung der Restriktivität der Modellspezifikation (*parsimonious fit*) und den Vergleich zu anderen, allgemeineren Modellen (*incremental* bzw. *relative fit*) einschließt.

Ausgangspunkt für die Beurteilung des Gesamtfits eines Modells ist die Abweichung der vom Modell implizierten Kovarianzmatrix von der empirischen Kovarianzmatrix der beobachteten Daten. Diese Abweichung wird mit Hilfe der für die Modellschätzung verwendeten Verlustfunktion, z. B. F_{ML} , minimiert.

Die Prüfgröße für die Beurteilung des Gesamtfits ist das mit der Stichprobengröße verrechnete Minimum der Likelihood-Funktion F_{ML} , das approximativ χ^2 -verteilt ist. Voraussetzung ist dabei die Gültigkeit des Modells, für das die Parameter geschätzt wurden (Neale, Boker, Xie, & Maes, 1999). Diese Prüfgröße,

$$(N-1)F_{ML}[\mathbf{S}, \Sigma(\hat{\boldsymbol{\theta}})] \square \chi^2(df), \quad [8.12],$$

mit den Freiheitsgraden $df = (s - t)$ ist die Grundlage für den Signifikanztest zur Überprüfung der Passung der beobachteten Daten zum theoretisch angenommenen Modell. Dabei sind s die Zahl nicht-redundanter Elemente der empirischen Kovarianzmatrix \mathbf{S} und t die Anzahl aller unabhängigen geschätzten Modellparameter. Die Nullhypothese des Tests auf Gesamtfits ist dabei, dass die vom Modell implizierte Kovarianzmatrix $\Sigma(\boldsymbol{\theta})$ gleich der Kovarianzmatrix Σ der gemeinsamen Verteilung der Observablen ist, d. h. $H_0: \Sigma = \Sigma(\boldsymbol{\theta})$. Gilt die Nullhypothese, kann die *wahre* Kovarianzmatrix der Observablen aus dem Vektor $\boldsymbol{\theta}$ der *wahren* Parameter perfekt berechnet werden. Zur Beurteilung dieser Nullhypothese stehen natürlich nur die empirische Kovarianzmatrix und die geschätzten Parameter zur Verfügung.

Man beachte die Testlogik für die Beurteilung des Gesamtfits des Modells: Wird die Nullhypothese beibehalten, so passen die beobachteten Daten zu dem theoretisch spezifizierten Modell, während eine Verwerfung der Nullhypothese als eine Nichtpassung von Daten

und Modell interpretiert wird. Dabei ist zu bemerken, dass die Beibehaltung der Nullhypothese lediglich die Passung von empirischen Daten und Modell bedeutet. Alternative Modelle können die Daten aber u. U. ebenso gut beschreiben (Lee & Hershberger, 1990).

Die χ^2 -Statistik hat einige Eigenschaften, die bei der Beurteilung des Modells beachtet werden müssen. So ist die Prüfgröße abhängig von

- a) der Stichprobengröße,
- b) der Anzahl der Modellparameter,
- c) der korrekten Spezifikation des theoretischen Modells und
- d) der Gültigkeit der Annahmen bezüglich der Verteilung der Observablen.

Das vielleicht größte Problem der χ^2 -Statistik ist die Abhängigkeit von der Stichprobengröße, die aufgrund der Testlogik bei der Bewertung von Strukturgleichungsmodellen besondere Bedeutung erlangt. Bei gleicher Zahl von Modellfreiheitsgraden und steigender Stichprobengröße wird der χ^2 -Wert größer. Das hat zur Folge, dass plausible Modelle mit großen Stichproben auf der Basis dieser Teststatistik verworfen werden, obwohl die Abweichung der empirischen von der modellimplizierten Kovarianzmatrix vernachlässigbar klein ist. Auf der anderen Seite können Modelle mit einer kleinen Stichprobe trotz großer Abweichung beider Kovarianzmatrizen einen nicht-signifikanten p -Wert erreichen, d. h. dass das Modell nicht verworfen wird. Simulationen zum Einfluss der Stichprobengröße auf die χ^2 -Statistik zeigen, dass dieser Effekt bei gut passenden Modellen weniger stark auftritt, während Modelle, bei denen das Modell die Daten schlecht erklärt, überproportional hart bestraft werden (Wolf & Kroehne, 2002). Zur praktischen Handhabung des Problems der Stichprobengröße empfehlen Jöreskog und Sörbom (1993) statt der Modellbeurteilung mittels χ^2 -Test den Vergleich des χ^2 -Werts mit dem Erwartungswert von χ^2 in der Stichprobenverteilung, d. h. den Freiheitsgraden [$E(\chi^2) = df$]. Als Kriterium für den Modellfit sollte der Quotient χ^2 / df nicht größer als 2 (guter Fit) bzw. 3 (akzeptabler Fit) sein.

Ein weiterer Nachteil der χ^2 -Statistik ist, dass die Prüfgröße bei zunehmender Zahl der Parameter im Modell kleiner wird. Das bedeutet, dass komplexe Modelle – bei gleicher Stichprobengröße – einen besseren Gesamtfit haben als entsprechend einfachere Modelle (Müller, 1996). Das gewinnt besondere Bedeutung im Zusammenhang mit der Restriktivität von Modellen. So verändert die künstliche numerische Fixierung von Parametern auf den frei geschätzten Wert den χ^2 -Wert nicht, während gleichzeitig die Zahl der Freiheitsgrade im Modell steigt. Zur Vermeidung des damit verbundenen Verlusts des konfirmatorischen Charakters des Modelltests sollten alle Modellrestriktionen inhaltlich begründet sein. Wie bereits im letzten Abschnitt erwähnt, setzt die Verwendung des χ^2 -Tests außerdem die multivariate Normalverteilung der Observablen voraus, die jedoch in vielen empirischen Untersuchungen nicht gegeben ist, wobei die χ^2 -Statistik aber – bis auf Fälle von extremen Verletzungen dieser Annahme – hinreichend robust ist.

Neben der χ^2 -Statistik, die jedoch aufgrund der oben getroffenen Einschränkungen hinsichtlich der Effekte von Stichprobengröße, Anzahl der Modellparameter und Verteilungsannahmen kein alleiniges Kriterium für die Modellbewertung darstellen sollte, existieren eine Reihe deskriptiver Kriterien für die Passung eines Modells. Zentrale deskriptive Indizes bei Strukturgleichungsmodellen sind neben dem RMSEA, der RMR (Jöreskog & Sörbom, 1981) bzw. SRMR (Bentler, 1995) zur Analyse der vom Modell nicht-aufgeklärten Varianzanteile der empirischen Kovarianzmatrix S .

Ein häufig verwendeter und obligatorisch zu berichtender Index zur Beurteilung der Passung von Strukturgleichungsmodellen ist der *Root Mean Square Error of Approximation* (RMSEA; Steiger, 1990; MacCallum, 1986). Ausgangspunkt hierfür ist der auf der Populationsebene definierten *Root Mean Square Index* (MacCallum, 1986)

$$\varepsilon = \sqrt{F[\Sigma, \Sigma(\theta)] / df}, \quad [8.13],$$

wobei $F[\mathbf{\Sigma}, \mathbf{\Sigma}(\boldsymbol{\theta})]$ ein Maß der Nichtpassung des Modells auf Populationsebene darstellt, welches das Ausmaß der Nichtpassung an den Freiheitsgraden normiert. Ersetzt man die theoretischen Populationskennwerte durch ihre empirischen Stichprobenkennwerte, erhält man

$$\text{RMSEA} = \hat{\varepsilon} = \sqrt{\max\left(\left(F[\mathbf{S}, \mathbf{\Sigma}(\hat{\boldsymbol{\theta})}] - df\right) / df, 0\right)}, \quad [8.14],$$

mit der empirischen Kovarianzmatrix \mathbf{S} , der vom Modell implizierten Kovarianzmatrix $\mathbf{\Sigma}(\boldsymbol{\theta})$ und df als Zahl der Freiheitsgrade des Modells, wobei im Term $(F[\mathbf{S}, \mathbf{\Sigma}(\hat{\boldsymbol{\theta})}] - df)$ für die Nichtzentralität der Stichprobenverteilung korrigiert wird. Aufgrund der unbekanntem Verteilung dieses, wie auch anderer, Fit-Indizes, können dabei keine kritische Werte angegeben werden (Hu & Bentler, 1995; Raykov, 2000). Vielmehr sollte beim RMSEA der *test of close fit* (Browne & Cudeck, 1993) und dessen Konfidenzintervall Anwendung finden. Werte des RMSEA unter 0.05 werden allgemein als "guter Fit" und Werte zwischen 0.05 und 0.08 als "adäquater Fit" eines Modells interpretiert. Darüber hinaus sollte das 90%-Konfidenzintervall des RMSEA, welches im Rahmen der Irrtumswahrscheinlichkeit den wahren RMSEA dieses Modells in der Population beinhaltet, für einen "guten Fit" den Wert 0 enthalten. Im Gegensatz zur χ^2 -Statistik ist das RMSEA-Kriterium relativ robust gegenüber der Stichprobengröße und der Anzahl der Modellparameter. Das gilt aber nicht für die Größe des Konfidenzintervalls des RMSEA, welches sensitiv für Veränderungen beider Einflussgrößen ist.

In angewandten Untersuchungen sind Forscher häufig in der Situation, zwischen mehreren alternativen Modellen auswählen zu müssen. Die Auswahl der Teststatistik für einen direkten Modellvergleich hängt dabei vor allem davon ab, ob die alternativen Modelle *genestete* Modelle sind, d. h. ob das liberalere Modell A (mit mehr frei zu schätzenden Parametern und weniger Freiheitsgraden) durch Restriktion (Fixierung oder Gleichsetzung) mindestens eines Modellparameters in das alternative Modell B überführt werden kann. Anders ausgedrückt sind die frei zu schätzenden Parameter des Modells B eine Teilmenge der frei zu schätzenden Parameter des Modells A. Unter der Voraussetzung der Gültigkeit der Nullhypothese in beiden Modellen, d. h. der Passung beider Modelle, sind sowohl die Teststatistik der einzelnen Modelle als auch die Differenz der Prüfgrößen beider Modelle χ^2 -verteilt (Steiger, Shapiro, & Browne, 1985), wobei die Zahl der Freiheitsgrade im Falle des Modellvergleichs gleich der Differenz der Freiheitsgrade beider Modelle ist. Unter Voraussetzung der Annahmen der χ^2 -Statistik, können die Modelle daher mit

$$\chi_{\text{diff}}^2(df_{\text{diff}}) = \chi_{\text{B}}^2(df_{\text{B}}) - \chi_{\text{A}}^2(df_{\text{A}}) \quad [8.15],$$

mit $df_{\text{diff}} = df_{\text{B}} - df_{\text{A}}$ verglichen werden (Bollen, 1989). Die Nullhypothese ist, dass sich der Gesamtfitt des restriktiveren Modells B – gemessen an der größeren Zahl der Freiheitsgrade – nicht verschlechtert. Daher wird bei einem nicht-signifikanten χ^2 -Differenzentest das restriktivere Modell, hier Modell B, aus Gründen der Sparsamkeit beibehalten. Der χ^2 -Differenzentest hat dabei dieselben Vor- bzw. Nachteile wie der χ^2 -Test für den Gesamtfitt eines einzelnen Modells.

Einschränkend soll darauf hingewiesen werden, dass die Differenz zweier Satorra-Bentler korrigierter χ^2 -Werte (zur robusten Modellbeurteilung) *nicht* χ^2 -verteilt ist (Satorra, 2000). Unter Verwendung einer einfachen Korrekturformel (Satorra & Bentler, 2001) lassen sich aber auch entsprechend korrigierte Modelle vergleichen (<http://www.abdn.ac.uk/~psy086/dept/psychom.htm>). Entscheidend für den Vergleich zweier Modelle mittels (korrigiertem) χ^2 -Differenzentest ist, dass diese Teststatistik – bis auf einige Ausnahmen – nur gültig ist, wenn das Basismodell richtig spezifiziert wurde. Beim Vergleich zweier misspezifizierter

Modelle ist die Verteilung der Prüfgröße des χ^2 -Differenzentests in der Regel unbekannt, und damit sind inferenzstatistische Aussagen nicht möglich.

Sollen zwei nicht-genestete Modelle miteinander verglichen werden, so kann der χ^2 -Differenzentest keine Verwendung finden, da über die Verteilungseigenschaften der so ermittelten Prüfgröße keine Aussagen getroffen werden können. Hierzu existieren lediglich deskriptive Statistiken wie das *Akaike Information Criterion* (AIC; Akaike, 1974, 1987), das mit

$$\text{AIC} = \chi^2 + 2p = (N - 1)F[\mathbf{S}, \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})] + 2p, \quad [8.16],$$

definiert ist. Dabei wird die Passung der Modelle (χ^2 -Wert) unter Berücksichtigung der Komplexität der konkurrierenden Modelle (*parsimony*) – über die Anzahl der geschätzten Modellparameter p – verglichen. Kritische Punkte hierbei sind die fehlende inferenzstatistische Absicherung der Befunde sowie Effekte unterschiedlicher Stichprobengröße, wie sie bereits für den Gesamtfit von Modellen diskutiert wurden.

Weitere deskriptive Indizes existieren für a) den Modellvergleich mit allgemeineren Modellen wie dem Unabhängigkeitsmodell (*incremental fit*), z. B. der *NFI* (Bentler & Bonett, 1980) und der *GFI* bzw. *AGFI* (Jöreskog & Sörbom, 1989) sowie b) für den Modellvergleich unter Berücksichtigung der Komplexität der alternativen Modelle (*parsimonious fit*), z. B. der *PNFI* (James, Mulaik, & Brett, 1982) und der *PGFI* (Mulaik et al., 1989). Nähere Ausführungen zu diesen Fitindizes finden sich beispielsweise bei Schermelleh-Engel et al., 2003.

Neben der Beurteilung eines Gesamtmodells bzw. dem Vergleich zweier konkurrierender Modelle können im Rahmen von Strukturgleichungsmodellen auch einzelne Modellparameter (Ladungen, Varianzen, Kovarianzen) inferenzstatistisch getestet werden. Auch bei einem Modell mit gutem oder akzeptablem Gesamtfit, können einzelne Modellparameter bedeutungslos sein, d. h. nicht zur Verbesserung des Modells beitragen. Nach der Prüfung aller Parameter auf Plausibilität (z. B. positive Varianzen, Korrelationen nicht größer 1, Vorzeichen), sollten diese mittels der *t-Statistik* gegen 0 (oder eine andere Nullhypothese) geprüft werden. Bei einer großen Stichprobe nähert sich die *t-Statistik* der *z-Statistik* an. Außerdem kann die Bewertung einzelner Parameter auch als Spezialfall eines Modellvergleichs betrachtet werden, bei dem im restriktiveren Modell der Parameter auf Null gesetzt wird. Somit wird der Vergleich der Modelle mittels χ^2 -Statistik möglich. Allgemein können alle Parameterrestriktionen eines Modells auch mittels eines Lagrange Multipliers (LM) oder der Wald-Statistik (W) getestet werden, wobei ersterer die mögliche Aufhebung von Restriktionen aus dem restringierten Modell und die Wald-Statistik das mögliche Setzen von Restriktionen im liberalen Modell testet (Bollen, 1989).

9. Ausblick

Im vorliegenden Artikel wurden Spezialfälle, aber auch Grundlagen und Einsatzmöglichkeiten von Strukturgleichungsmodellen aufgezeigt. Mit Hilfe der Strukturgleichungsmodellierung gelingt die Integration von faktoranalytischen und regressionsanalytischen Verfahren: Faktoranalytische Messmodelle und Modelle zur Analyse komplexer regressiver Zusammenhängestrukturen können damit integriert und in einem Schritt geschätzt und bewertet werden. Die Darstellung der Breite der sich daraus ergebenden Einsatzmöglichkeiten konnte in diesem Übersichtsartikel nur ansatzweise erfolgen, soll nun aber durch einige weitere Punkte ergänzt werden.

Die wissenschaftliche Bedeutung und Tragweite der Strukturgleichungsmodellierung geht über die technischen Vorteile einer flexiblen und mächtigen Klasse statistischer Analyseverfahren weit hinaus. Durch die mittlerweile gelungene Etablierung und leichtere Zugänglich-

keit dieser Verfahren infolge benutzerfreundlicher Software ist die Methode mehr als lediglich ein Werkzeug zur Überprüfung inhaltlicher Fragestellungen. Vielmehr kommt der Strukturgleichungsmodellierung im Prozess der wissenschaftlichen Erkenntnis eine theoriegenerierende und –formalisierende Position zu. Viele inhaltliche Theorien nutzen die diagrammatische Sprache von SEM zur Darstellung komplexer Zusammenhänge (vgl. z. B. die Fishbein-Ajzen-Theorie geplanten Verhaltens).

9.1. Weitere Anwendungsbereiche von Strukturgleichungsmodellen

Eine bedeutende Erweiterung der bisher besprochenen Modelle ist die Berücksichtigung der *Mittelwerte der latenten Variablen*. Bei der Einführung der vollständigen Modellgleichungen in ihrer allgemeinen Form wurden additive Konstanten nicht berücksichtigt, da in der Regel die Zusammenhangsstruktur der manifesten und latenten Variablen zentraler Bestandteil von inhaltlichen Theorien ist. Mittelwerte latenter Variablen können jedoch ebenfalls explizit in Strukturgleichungsmodellen modelliert werden. Bei längsschnittlichen Modellen lassen sich dann nicht nur die Varianzen von latenten Variablen beschreiben, sondern auch Veränderungen in den Mittelwerten der latenten Variablen.

Neben der bisher zugrunde liegenden Modellierung einer Stichprobe können Strukturgleichungsmodelle auch für die *Analyse mehrerer Gruppen* verallgemeinert werden (s. z. B. Byrne, 1989). Diese Gruppen können sowohl natürliche Substichproben (Frauen vs. Männer, Experimental- vs. Kontrollgruppe) als auch artifiziell gebildete Gruppen sein (verschiedene Ausprägung eines kontinuierlichen Merkmals nach Perzentilsplit). Dieser Ansatz bietet bspw. eine relativ einfache Möglichkeit des Tests von Moderatorhypothesen, wie man sie als Interaktionseffekte aus varianzanalytischen Zusammenhängen kennt. Somit lassen sich alle varianzanalytischen Verfahren bis hin zur Messwiederholungs-MANOVA – hinreichend große Stichproben vorausgesetzt – auch im Rahmen von SEM modellieren. Darüber hinaus können in SEM-Mehrgruppenanalysen traditionelle Probleme der Varianzanalyse (Varianzhomogenität, latente Variablen) mit der Strukturgleichungsmodellierung überwunden werden. Die Flexibilität der Modelle vergrößert sich in der Mehrgruppenanalyse weiterhin dadurch, dass alle Parameter eines Mess- und Strukturmodells zwischen den Stichproben variieren können. Allerdings lassen sich aber auch Restriktionen (z. B. Gleichsetzungen) über die Parameter aus verschiedenen Gruppen einführen, die man dann wiederum testen kann.

Eine dritte wichtige Erweiterung von Strukturgleichungsmodellen wurde mit der Einführung von *Modellen für dichotome und ordinale Variablen* vorgenommen (s. z. B. Muthen, 1984). Diese Modelle verbinden alle Vorteile von Strukturgleichungsmodellen mit der Möglichkeit, die Messmodelle auf Itemebene anzusetzen. Dabei kann es sich um dichotome oder auch um ordinale Items handeln, wie sie etwa bei Ratingskalen vorliegen, bei denen das Ausmaß der Zustimmung zu einer Aussage (z. B. über die eigene Befindlichkeit, Persönlichkeit oder Einstellung) durch die Auswahl einer von mehreren Antwortkategorien anzugeben ist. Die Wahrscheinlichkeit einer Person, in einer bestimmten Kategorie zu antworten, hängt dann von einer oder mehreren latenten Variablen ab, deren Abhängigkeit untereinander mit normalen Strukturgleichungen modelliert werden kann. Sind diese latenten Variablen wieder normalverteilt, so kann man diese Art von Modellen auch in einem zweistufigen Verfahren analysieren: zuerst werden die polychorischen Korrelationen der Items (z. B. mit PRELIS 2; Jöreskog & Sörbom, 1996c) errechnet, danach werden diese mit Strukturgleichungen modelliert (z. B. mit LISREL 8; Jöreskog & Sörbom, 1996a). Kombiniert man die drei genannten Erweiterungen von Strukturgleichungsmodellen, die Analyse von Mittelwerten latenter Variablen, die Analyse mehrerer Gruppen und die Analyse von ordinalen Variablen, so kann man die Flexibilität und den riesigen Bereich der Anwendbarkeit dieser Modelle erahnen.

9.2. Aktuelle Entwicklungsrichtungen von SEM

In den letzten Jahren etablierte sich die Strukturgleichungsmodellierung kontinuierlich in den Sozial- und Verhaltenswissenschaften. Die Weiterentwicklungen beschäftigten sich in erster Linie mit der Optimierung bestehender und Ausarbeitung neuer Schätzverfahren, mit den Problemen nicht normalverteilter, nicht kontinuierlicher sowie systematisch oder unsystematisch fehlender Daten. Neben diesen neuen Schätzalgorithmen wurden verschiedene neue Fit-Indizes etabliert.

Eine neue Generation der Strukturgleichungsmodellierung entsteht derzeit durch die Einführung von Mischverteilungsmodellen. Bei diesen Modellen gibt es latente Klassen von Personen innerhalb derer jeweils ein Strukturgleichungsmodell gilt, allerdings mit Parametern, die sich zwischen den latenten Klassen unterscheiden können.

Die Integration von Mehrebenenmodellen in Strukturgleichungsmodelle erlaubt die Analyse von hierarchisch strukturierten Daten (bspw. bei der längsschnittlichen Untersuchung von Schülern in Schulklassen) unter Einbeziehung von *random intercepts* und *random slopes*, mit der theoretischen Möglichkeit individuell variierender Erhebungsintervalle und zeitabhängiger Kovariaten. Viele dieser Verfahren finden ihre praktische Umsetzung in der von Linda und Bengt Muthén entwickelten Software MPlus, die seit 2004 in der Version 3 vorliegt (<http://www.statmodel.com/>).

9.3. Literatur und Informationsquellen über SEM

Als Standardwerk zu Strukturgleichungsmodellen ist Bollen (1989, aktualisierte Auflage 2004) zu empfehlen. Hier wird eine anspruchsvolle und dennoch verständliche Einführung in die gesamte Breite des Verfahrens geboten. Weitere Einführungstexte sind teilweise unabhängig von der zugrunde liegenden Software (z. B. Kline, 1998; Hoyle, 1995; Schumacker & Lomax, 1996) oder aber sie beziehen sich explizit auf ein Programm: LISREL (Kelloway, 1998; Byrne, 1989, 1998; Hayduk, 1996; Jöreskog & Sörbom, 1996a, 1996b, 1996c, Jöreskog, Sörbom, du Toit, & du Toit, 2001), AMOS (Byrne, 2001), MPlus (Muthen & Muthen, 1999) oder EQS (Bentler, 1995; Byrne, 1994). Spezielle Fragen wie etwa Schätzalgorithmen, nicht-lineare Effekte, Interaktionseffekte oder Multi-Level-Ansätze werden in Sammelwerken behandelt (z. B. Bollen & Long, 1993; Kaplan, 2000; Marcoulides & Schumacker, 1996; Schumacker & Marcoulides, 1998a, 1998b). Hinweise zu Konventionen der „*state-of-the-art*“-Berichterstattung geben Anne Boomsma (2000) oder Hoyle und Panter (1995).

Die neuesten Entwicklungen verfolgen interessierte Leser am besten in der Zeitschrift „Structural Equation Modeling - A Multidisciplinary Journal“ oder im sehr aktiven Diskussionsforum SEMNET (<http://www.gsu.edu/~mkteer/semnet.html>).

Box 1. Grundbegriffe der Klassischen Testtheorie

Die Menge der möglichen Ergebnisse des Zufallsexperiments

$$\Omega = \Omega_U \times \Omega_O$$

Messwertvariablen

$$Y_i: \Omega \rightarrow \mathbb{R}$$

Person-Projektion

$$U: \Omega \rightarrow \Omega_U$$

Latente Variablen

$$\tau_i := E(Y_i | U)$$

$$\varepsilon_i := Y_i - \tau_i$$

Dekomposition der Variablen

$$Y_i = \tau_i + \varepsilon_i$$

Dekomposition der Varianzen

$$\text{Var}(Y_i) = \text{Var}(\tau_i) + \text{Var}(\varepsilon_i)$$

Wichtige Kenngröße

$$\text{Rel}(Y_i) := \text{Var}(\tau_i) / \text{Var}(Y_i)$$

Box 2. Grundbegriffe der Latent-state-trait-Theorie

Die Menge der möglichen Ergebnisse des Zufallsexperiments

$$\Omega = \Omega_U \times \Omega_{S_1} \times \dots \times \Omega_{S_r} \times \dots \times \Omega_{S_n} \times \Omega_{O_1} \times \dots \times \Omega_{O_r} \times \dots \times \Omega_{O_n}$$

Messwertvariablen

$$Y_{it} : \Omega \rightarrow \mathbb{R}$$

Projektionen

$$U : \Omega \rightarrow \Omega_U$$

$$S_i : \Omega \rightarrow \Omega_{S_i}$$

Latente Variablen

$$\tau_{it} := E(Y_{it} | U, S_i)$$

$$\varepsilon_{it} := Y_{it} - \tau_{it}$$

$$\xi_{it} := E(Y_{it} | U)$$

$$\zeta_{it} := \tau_{it} - \xi_{it}$$

Dekomposition der Variablen

$$Y_{it} = \tau_{it} + \varepsilon_{it}$$

$$\tau_{it} = \xi_{it} + \zeta_{it}$$

Dekomposition der Varianzen

$$\text{Var}(Y_{it}) = \text{Var}(\tau_{it}) + \text{Var}(\varepsilon_{it})$$

$$\text{Var}(\tau_{it}) = \text{Var}(\xi_{it}) + \text{Var}(\zeta_{it})$$

Wichtige Kenngrößen

$$\text{Rel}(Y_{it}) := \text{Var}(\tau_{it}) / \text{Var}(Y_{it})$$

$$\text{Con}(Y_{it}) := \text{Var}(\xi_{it}) / \text{Var}(Y_{it})$$

$$\text{Spe}(Y_{it}) := \text{Var}(\zeta_{it}) / \text{Var}(Y_{it})$$

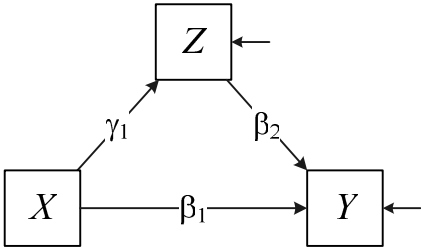


Abbildung 1. Pfaddiagramm eines Mediatormodells für drei Variablen

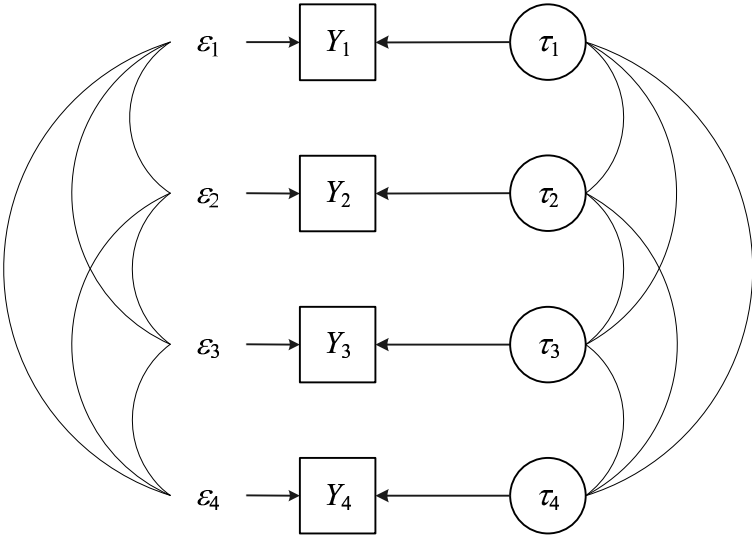


Abbildung 2. Pfaddiagramm zur Dekomposition der Messwertvariablen in True-score- und Messfehlervariablen.

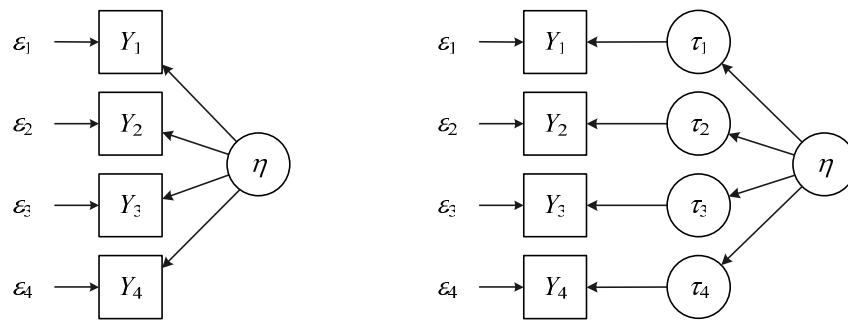


Abbildung 3. Pfaddiagramm für das Modell τ -äquivalenter Variablen. In der rechts dargestellten Abbildung haben die True-score-Variablen keine Residuen.

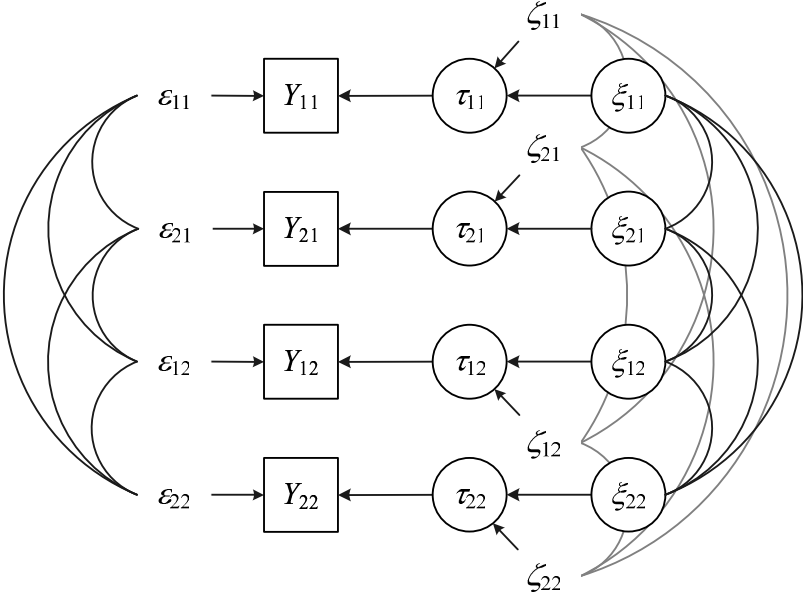


Abbildung 4. Pfaddiagramm zur Zerlegung der Messwertvariablen in Messfehler- und Latent-state-Variablen, die wiederum in Latent-trait-Variablen und Latent-state-Residuen zerlegt werden.

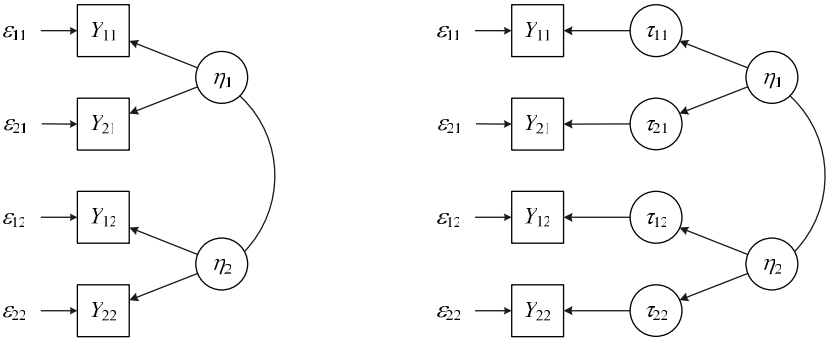


Abbildung 5. Pfaddiagramm des Multistate-Modells. Die Messfehlervariablen werden als unkorreliert angenommen, mit und ohne explizite Darstellung der Latentstate-Variablen.

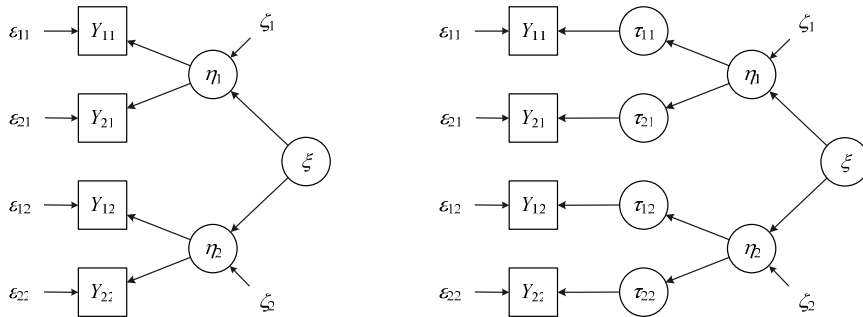


Abbildung 6. Pfaddiagramm des Single-trait-multistate-Modells. Die Messfehlervariablen werden als unkorreliert angenommen, ebenso die Latent-state-Residuen., mit und ohne explizite Darstellung der Latent-state-Variablen.

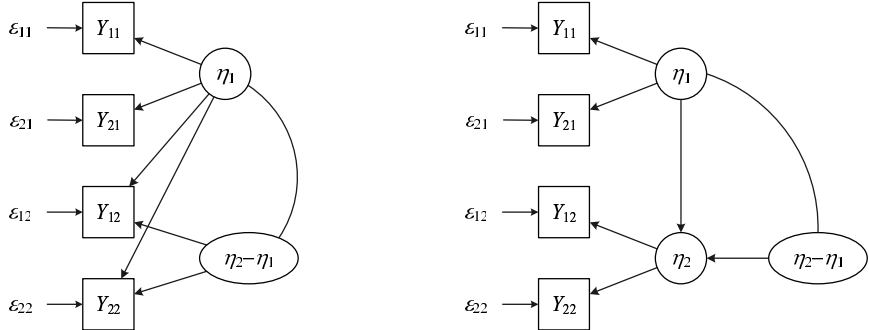


Abbildung 7. Zwei Darstellungen des Latent-change-Modells.

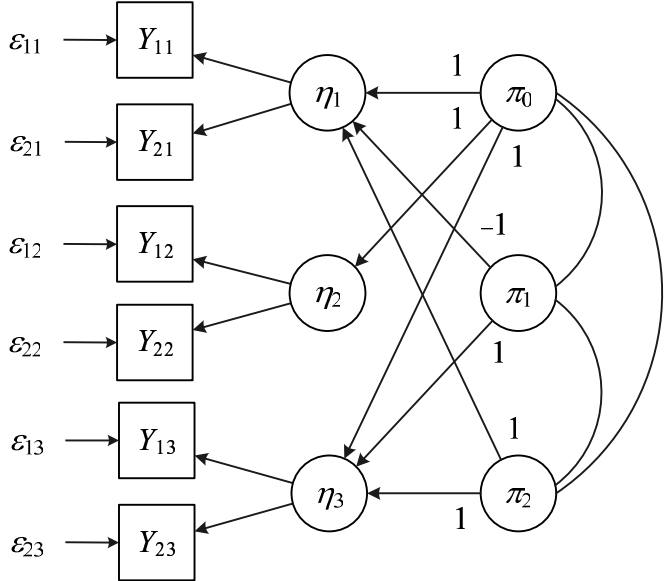


Abbildung 8. Wachstumskurvenmodell mit Zeitkodierung $\lambda(1) = -1$, $\lambda(2) = 0$ und $\lambda(3) := 1$. Die η -Variablen haben hier keine Residuen.

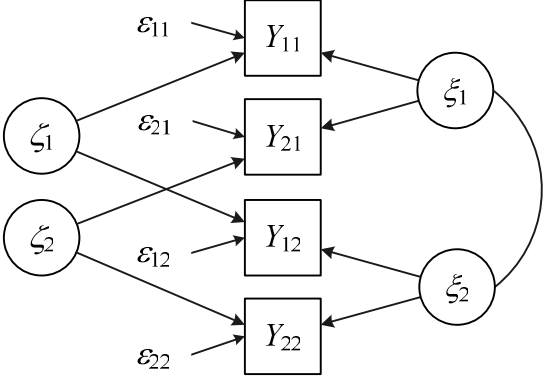


Abbildung 9. Pfaddiagramm des Multitrait-multimethod-Modells der Konfirmatorischen Faktorenanalyse (CFA)

Literatur

Akaike, H. (1974). A new look at statistical model identification. *IEEE transactions on Automatic Control*, 19, 716-723.

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika*, 52, 317-332.

Anderson, T. W. & Rubin, H. (1956). Statistical inference in factor analysis. In J. Neyman (Ed.), *Proceedings of the Third Berkeley Symposium* (pp. 111-150). Berkeley: University of California Press.

Baron, R. M. & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic and statistical considerations. *Journal of Personality and Social Psychology*, 51, 1173-1182.

Bentler, P. M. (1995). *EQS Structural equations program manual*. Encino, CA: Multivariate Software.

Bentler, P. M. & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, 88, 588-606.

Biesanz, J. C., Deeb-Sossa, N., Papadakis, A. A., Bollen, K. A., & Curran, P. J. (2004). The role of coding time in estimating and interpreting growth curve models. *Psychological Methods*, 9, 30-52.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.

Bollen, K. A. & Long, J. S. (1993). *Testing Structural Equation Models*. Newbury Park, CA: Sage.

Boomsma, A. & Hoogland, J. J. (2001). The robustness of LISREL modeling revisited. In R. Cudeck, S. du Toit, & D. Sörbom (Eds.), *Structural equation models: Present*

and future. A Festschrift in honor of Karl Jöreskog (pp. 139-168). Chicago, IL: Scientific Software International.

Boomsma, A. (2000). Reporting Analyses of Covariance Structures. *Structural Equation Modeling*, 7, 461-484.

Browne, M. W. & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Newbury Park, CA: Sage.

Byrne, B. M. (1989). *A primer of LISREL: Basic applications and programming for confirmatory factor analytic models*. New York: Springer.

Byrne, B. M. (1994). *Structural Equation Modeling with EQS and EQS/Windows: Basic concepts, applications, and programming*. Thousand Oaks, CA: Sage Publications.

Byrne, B. M. (1998). *Structural Equation Modeling with LISREL, PRELIS, and SIMPLIS: Basic concepts, applications, and programming*. Mahwah, NJ: Lawrence Erlbaum Associates.

Byrne, B. M. (2001). *Structural Equation Modeling with AMOS: Basic concepts, applications, and programming*. Mahwah, NJ: Lawrence Erlbaum Associates.

Campbell, D. T. & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research on teaching. In N.L.Gage (Ed.), *Handbook on research on teaching* (pp. 171-246). Chicago, IL: Rand McNally.

Chou, C.-P. & Bentler, P. M. (1995). Estimation and tests in structural equation modeling. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 37-55). Thousand Oaks, CA: Sage.

Cole, D. A., Martin, N. C., & Steiger, J. H. Empirical and conceptual problems with longitudinal trait-state models: Support for a trait-state-occasion model. *Psychological Methods*, (in press).

Cook, T. D. C. & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Boston, MA: Houghton Mifflin.

Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods*, 1, 16-29.

Eid, M. (1995). *Modelle der Messung von Personen in Situationen*. Weinheim: Psychologie Verlags Union.

Eid, M. (2000). Neuere testtheoretische Modelle der Veränderungsmessung. In R. R. Engel, W. Maier, & H.-J. Möller (Eds.), *Methodik von Verlaufs- und Therapiestudien in Psychiatrie und Psychotherapie* (pp. 1-9). Göttingen: Hogrefe.

Eid, M. & Hoffmann, L. (1998). Measuring variability and change with an item response model for polytomous variables. *Journal of Educational and Behavioral Statistics*, 23, 193-215.

Eid, M., Lischetzke, T., Trierweiler, L. I., & Nußbeck, F. W. (2003). Separating trait effects from trait-specific method effects in multitrait-multimethod models: A multiple-indicator CT-C(M-1) model. *Psychological Methods*, 8, 38-60.

Finkbeiner, C. (1979). Estimation for the multiple factor model when data are missing. *Psychometrika*, 44, 409-420.

Fox, J. (1984). *Linear statistical models and related methods*. New York: Wiley.

Gähde, U., Jagodzinski, W., & Steyer, R. (1992). On a structuralist reconstruction of latent state-trait theory. In H. Westmeyer (Ed.), *The structuralist program in psychology: Foundations and applications* (pp. 105-119). Göttingen: Hogrefe & Huber.

Goldberger, A. S. (1973). Structural equation models: An overview. In A. S. Goldberger & O. D. Duncan (Eds.), *Structural equation models in the social sciences* (pp. 1-18).

Grayson, D. A. & Marsh, H. W. (1994). Identification with deficient rank loading matrices in confirmatory factor analysis: Multitrait-multimethod models. *Psychometrika*, 59, 121-134.

Gulliksen, H. (1950). *Theory of mental tests*. New York: Wiley.

Hagemann, D., Naumann, E., Thayer, J. F., & Bartussek, D. (2002). Does resting EEG asymmetry reflect a trait? An application of latent state-trait theory. *Journal of Personality and Social Psychology*, 82, 619-641.

Hayduk, L. A. (1996). *LISREL issues, debates, and strategies*. Baltimore: Johns Hopkins University Press.

Hoogland, J. J. (1999). *The robustness of estimation methods for covariance structure analysis. (Unpublished doctoral dissertation)*. Groningen: University of Groningen.

Hoogland, J. J. & Boomsma, A. (1998). Robustness studies in covariance structure modeling: An overview and a meta-analysis. *Sociological Methods & Research*, 26, 329-367.

Hoyle, R. H. (1995). *Structural Equation Modeling: Concepts, issues, and applications*. Thousand Oaks: Sage.

Hoyle, R. H. & Panter, A. T. (1995). Writing about structural equation models. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 158-176). Thousand Oaks, CA: SAGE Publications.

Hu, L. & Bentler, P. M. (1995). Evaluating model fit. In R. H. Hoyle (Ed.), *Structural equation modeling. Concepts, issues, and applications* (pp. 76-99). London: Sage.

James, L. R., Mulaik, S. A., & Brett, J. M. (1982). *Causal analysis: Assumptions, models, and data*. Beverly Hills, CA: Sage.

Jamshidian, M. & Bentler, P. M. (1999). ML estimation of mean and covariance structures with missing data using complete data routines. *Journal of Educational and Behavioral Statistics*, 24, 21-41.

Johnston, J. J. (1984). *Econometric Methods*. New York: McGraw-Hill.

Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, 34, 183-202.

Jöreskog, K. G. (1971). Statistical analysis of sets of congeneric tests. *Psychometrika*, 36, 109-133.

Jöreskog, K. G. (1979). Basic ideas of factor and component analysis. In K. G. Jöreskog & D. Sörbom (Eds.), *Advances in factor analysis and Structural Equation Models* (pp. 5-20). Cambridge: Abt Books.

Jöreskog, K. G. & Sörbom, D. (1989). *LISREL 7. A guide to the program and its applications*. (2 ed.) Chicago: SPSS.

Jöreskog, K. G. & Sörbom, D. (1996a). *LISREL 8 user's reference guide*. Chicago: SSI.

Jöreskog, K. G. (1970). A general method for analysis of covariance structures. *Biometrika*, 57, 239-251.

Jöreskog, K. G. (1973). A general method for estimating a linear structural equation system. In A. S. Goldberger & O. D. Duncan (Eds.), *Structural Equation Model in the Social Sciences* (pp. 85-112).

Jöreskog, K. G. & Sörbom, D. (1981). *LISREL V: Analysis of linear structural relationships by maximum likelihood and least squares methods (Research Report 81-8)*. Uppsala: University of Uppsala, Department of Statistics.

Jöreskog, K. G. & Sörbom, D. (1993). *LISREL 8: Structural equation modeling with the SIMPLIS command language*. Hillsdale, N.J.: Erlbaum Associates.

Jöreskog, K. G. & Sörbom, D. (1996b). *LISREL 8: Structural Equation Modeling with the SIMPLIS command language*. Chicago: SSI.

Jöreskog, K. G. & Sörbom, D. (1996c). *PRELIS 2 user's reference guide: A program for multivariate data screening and data summarization*. Chicago: SSI.

Jöreskog, K. G., Sörbom, D., du Toit, S., & du Toit, M. (2001). *LISREL 8: New statistical features*. Chicago: SSI.

Kaplan, D. (2000). *Structural equation modeling: Foundation and extensions*. Thousand Oaks, CA: Sage.

Kelloway, E. K. (1998). *Using LISREL for Structural Equation Modeling. A researcher's guide*. Thousand Oaks, CA: Sage.

Kenny, D. A. & Kashy, D. A. (1992). Analysis of the multitrait-multimethod matrix by confirmatory factor analysis. *Psychological Bulletin*, 112, 165-172.

Kirschbaum, C., Steyer, R., Eid, M., Patalla, U., Schwenkmezger, P., & Hellhammer, D. H. (1990). Cortisol and behavior: 2. application of a latent state-trait model to salivary cortisol. *Psychoneuroendocrinology*, *15*, 297-307.

Kline, R. B. (1998). *Principles and practice of Structural Equation Modeling*. New York: Guilford Press.

Lawley, D. N. & Maxwell, A. E. (1971). *Factor analysis as a statistical method*. London: Butterworths.

Lee, S. & Hershberger, S. (1990). A simple rule for generating equivalent models in covariance structure modeling. *Multivariate Behavioral Research*, *25*, 313-334.

Lienert, G. A. (1989). *Testaufbau und Testanalyse*. (4. neu ausgestattete Aufl. ed.) München: Psychologie-Verlags-Union.

Lord, F. M. & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison Wesley.

MacCallum, R. C. (1986). Specification searches in covariance structure modeling. *Psychological Bulletin*, *100*, 107-120.

Marcoulides, G. A. & Schumacker, R. E. (1996). *Advanced Structural Equation Modeling: Issues and Techniques*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Marsh, H. W. & Grayson, D. (1995). Latent variable models of multitrait-multimethod data. In R. H. Hoyle (Ed.), *Structural equation modeling. Concepts, issues, and applications* (Thousands Oaks: Sage.

Marsh, H. W. (1989). Confirmatory factor analyses of multitrait-multimethod data: Many problems and a few solutions. *Applied Psychological Measurement*, *13*, 335-361.

Mischel, W. (1977a). On the future of personality measurement. *American Psychologist*, 32, 246-254.

Mischel, W. (1977b). The interaction of person and situation. In D. Magnusson & N. S. Endler (Eds.), *Personality at the crossroads* (pp. 333-352). New York: Wiley.

Mislevy, R. J. (1986). Recent developments in the factor analysis of categorical variables. *Journal of Educational Statistics*, 11, 3-31.

Mulaik, S. A., James, L. R., Van Alstine, J., Bennett, N., Lind, S., & Stilwell, C. D. (1989). Evaluation of goodness-of-fit indices for structural equation models. *Psychological Bulletin*, 105, 430-445.

Müller, R. O. (1996). *Basic principles of structural equation modeling: An introduction to LISREL and EQS*. New York: Springer.

Muthen, B. (1978). Contributions to factor analysis of dichotomous variables. *Psychometrika*, 43, 551-560.

Muthen, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika*, 49, 115-132.

Muthen, B. & Christofferson, A. (1981). Simultaneous factor analysis of dichotomous variables in several groups. *Psychometrika*, 46, 407-419.

Muthen, L. K. & Muthen, B. O. (1999). *Mplus user's guide (version 2.0)*. Los Angeles, CA: Muthen & Muthen.

Nachtigall, C., Kroehne, U., Funke, F., & Steyer, R. (2003). (Why) Should we use SEM? Pros and cons of structural equation modeling. *Methods of Psychological Research Online*, 8, 1-22.

Neale, M. C. (2000). Individual fit, heterogeneity, and missing data in multigroup structural equation modeling. In T. D. Little & K. U. Schnabel (Eds.), *Modeling longitudinal and multilevel data: Practical issues, applied approaches, and specific examples* (pp. 249-267). Mahwah, NJ: Lawrence Erlbaum.

Neale, M. C., Boker, S. M., Xie, G., & Maes, H. H. (1999). *Mx: Statistical modeling*. (5 ed.) Richmond, VA: Department of Psychiatry.

Novick, M. R. (1966). The axioms and principal results of classical test theory. *Journal of Mathematical Psychology*, 3, 1-18.

Olsson, U. H., Foss, T., Troye, S. V., & Howell, R. D. (2000). The performance of ML, GLS, and WLS estimation in structural equation modeling under conditions of misspecification and nonnormality. *Structural Equation Modeling*, 7, 557-595.

Raykov, T. (2000). On the large-sample bias, variance, and mean squared error of the conventional noncentrality parameter estimator of covariance structure models. *Structural Equation Modeling*, 7, 431-441.

Revenstorf, D. (1980). *Faktorenanalyse*. Stuttgart: Kohlhammer.

Röhr, M. (1993). *Statistische Strukturanalysen*. Stuttgart: Gustav Fischer.

Satorra, A. (2000). Scaled and adjusted restricted tests in multi-sample analysis of moment structures. In R. D. H. Heijmans, D. S. G. Pollock, & A. Satorra (Eds.), *Innovations in multivariate statistical analysis* (pp. 233-247). London: Kluwer Academic Publishers.

Satorra, A. & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye & C. C. Clogg (Eds.), *Latent variable analysis: Applications for developmental research* (pp. 399-419). Thousand Oaks, CA: Sage.

Satorra, A. & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, *66*, 507-514.

Schermelleh-Engel, K., Keith, N., Moosbrugger, H., & Hodapp, V. (2004). Decomposing person and occasion-specific effects: An extension of latent state-trait (LST) theory to hierarchical LST models. *Psychological Methods*, *9*, 198-219.

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online*, *8*, 23-74.

Schumacker, R. E. & Lomax, R. G. (1996). *A beginner's guide to Structural Equation Modeling*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Schumacker, R. E. & Marcoulides, G. A. (1998a). *Advanced Structural Equation Modeling: Interaction models*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Schumacker, R. E. & Marcoulides, G. A. (1998b). *Interaction and non-linear effects in Structural Equation*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin Company.

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, *25*, 173-180.

Steiger, J. H., Shapiro, A., & Browne, M. W. (1985). On the multivariate asymptotic distribution of sequential Chi-square statistics. *Psychometrika*, *50*, 253-264.

Steyer, R. (1992). *Theorie kausaler Regressionsmodelle*. Stuttgart: Gustav Fischer Verlag.

Steyer, R. (2003). *Wahrscheinlichkeit und Regression*. Berlin: Springer.

Steyer, R. (2005). Analyzing individual and average causal effects via structural equation models. *Methodology*, 1.

Steyer, R. & Eid, M. (2001). *Messen und Testen*. Berlin: Springer.

Steyer, R., Eid, M., & Schwenkmezger, P. (1996). Ein Latent-State-Trait-Modell für Variablen mit geordneten Antwortkategorien und seine Anwendung zur Analyse der Variabilitätssensitivität von Stimmungsskizzen. *Diagnostica*, 42, 293-312.

Steyer, R., Eid, M., & Schwenkmezger, P. (1997). Modeling true intraindividual change: True change as a latent variable. *Methods of Psychological Research Online*, 2, 21-34.

Steyer, R., Ferring, D., & Schmitt, M. J. (1992). States and traits in psychological assessment. *European Journal of Psychological Assessment*, 8, 79-98.

Steyer, R., Gabler, S., von Davier, A. A., & Nachtigall, C. (2000a). Causal regression models II: Unconfoundedness and causal unbiasedness. *Methods of Psychological Research Online*, 5, 55-86.

Steyer, R., Gabler, S., von Davier, A. A., Nachtigall, C., & Buhl, T. (2000b). Causal regression models I: Individual and average causal effects. *Methods of Psychological Research Online*, 5, 39-71.

Steyer, R., Krambeer, S., & Hannover, W. (2004). Modeling latent trait-change. In K. van Montfort, J. H. L. Oud, & A. Satorra (Eds.), *Recent developments on structural equation models: Theory and applications* (pp. 337-357). Dordrecht: Kluwer.

Steyer, R., Nachtigall, C., Wüthrich-Martone, O., & Kraus, K. (2002). Causal regression models III: Covariates, conditional, and unconditional average causal effects. *Methods of Psychological Research Online*, 7, 41-68.

Steyer, R., Partchev, I., & Shanahan, M. J. (2000). Modeling true intradividual change in structural equation models: The case of poverty and children's psychosocial adjustment. In T. D. Little & K. U. Schnabel (Eds.), *Modeling longitudinal and multilevel data: Practical issues, applied approaches, and specific examples* (pp. 109-126). Mahwah, NJ, US: Lawrence Erlbaum Associates.

Steyer, R., Schmitt, M., & Eid, M. (1999). Latent state-trait theory and research in personality and individual differences. *European Journal of Personality*, 13, 389-408.

Steyer, R. & Schmitt, M. J. (1990). Latent state-trait models in attitude research. *Quality and Quantity*, 24, 427-445.

Tisak, J. & Tisak, M. S. (2000). Permanency and ephemerality of psychological measures with application to organizational commitment. *Psychological Methods*, 5, 175-198.

Überla, K. (1971). *Faktorenanalyse*. Berlin: Springer.

West, S. G., Finch, J. F., & Curran, P. J. (1995). Structural equation models with nonnormal variables: Problems and remedies. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 56-75). Thousand Oaks, CA: Sage.

Widaman, K. F. (1985). Hierarchically Nested Covariance Structure Models for Multitrait-Multimethod Data. *Psychological Measurement*, 9, 1-26.

Widaman, K. F. (1993). Common factor analysis versus principal component analysis: Differential bias in representing model parameters? *Multivariate Behavioral Research*, 28, 263-311.

Wolf, A. & Kroehne, U. (2002). *Ein struktureller Unterschied zur Erklärung von Unterschieden zwischen computerisierten und Papier-und-Bleistift Tests (Unveröffentlichte Diplomarbeit)*. Jena: Friedrich-Schiller-Universität Jena, Institut für Psychologie.

Wothke, W. (1996). Models for multitrait-multimethod matrix analysis. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling. Issues and techniques* (pp. 7-56). Mahwah: Erlbaum.

Wothke, W. (2000). Longitudinal and multigroup modeling with missing data. In T. D. Little & K. U. Schnabel (Eds.), *Modeling longitudinal and multilevel data: Practical issues, applied approaches, and specific examples* (pp. 219-240). Mahwah, NJ: Lawrence Erlbaum.

Wright, S. (1921). Correlation and causation. *Journal of Agricultural Research*, 20, 557-585.

Wright, S. (1934). The method of path coefficients. *Annals of Mathematical Statistics*, 5, 161-215.

Zimmerman, D. W. (1975). Probability spaces, Hilbert spaces and the axioms of test theory. *Psychometrika*, 40, 395-412.

Zimmerman, D. W. & Williams, R. H. (1977). The theory of test validity and correlated errors of measurement. *Journal of Mathematical Psychology*, 16, 135-152.