

UNIVERSITÄT TRIER

UNIVERSITÄTS-RECHENZENTRUM

B. Baltes-Götz

**Einführung in die Analyse von Strukturgleichungsmodellen mit
LISREL 7 und PRELIS unter SPSS**

(2., vollständig überarbeitete Auflage)

Vorwort

Diese Benutzereinführung richtet sich an Leser mit Grundkenntnissen in multivariater Statistik (vor allem relevant: Regressions- und Faktorenanalyse), die sich in die Analyse von Strukturgleichungsmodellen mit latenten Variablen einarbeiten wollen und dabei die von Karl Jöreskog und Dag Sörbom entwickelten Programme PRELIS und LISREL 7 in der Implementierung als SPSS-Prozeduren verwenden möchten. Das Manuskript entstand als Begleitlektüre zu LISREL-Kursen des Universitäts-Rechenzentrums Trier (URT), ist aber auch für das Selbststudium geeignet. In den URT-Kursen wurde die SPSS-Version 4.1 auf der Siemens-Anlage H90-D unter dem Betriebssystem BS2000 verwendet, doch ist der Text in der vorliegenden zweiten, völlig überarbeiteten Auflage bis auf einige ergänzende Hinweise unabhängig von der Rechner-Plattform. Daher wurde im Vergleich zur ersten Auflage der Hinweis auf das Betriebssystem BS2000 aus dem Titel gestrichen.

Das Manuskript ist anwendungsorientiert und erklärt ausführlich die praktische Durchführung der Analysen mit SPSS und LISREL, jedoch wird auch die zugrundeliegende statistische Theorie besprochen, soweit dies für die korrekte Spezifikation und Interpretation einer LISREL-Analyse erforderlich ist. Im einzelnen werden folgende Themen behandelt:

- die wichtigsten statistischen Grundlagen für das Arbeiten mit LISREL,
- das LISREL-Modell und die LISREL-Terminologie,
- die meisten in LISREL verfügbaren Ausgaben zur Beschreibung, Beurteilung und Modifikation von Modellen,
- die Vorbehandlung problematischer Daten mit PRELIS,
- die Syntax der SPSS-Prozeduren PRELIS und LISREL 7.

Einige spezielle LISREL-Möglichkeiten werden, dem Einführungscharakter des Manuskriptes entsprechend, nicht behandelt:

- LISREL-Modelle mit Ordinatenabschnitten und Mittelwerten für latente Variablen
- Mehr-Gruppen-Analysen

Alle Ausführungen werden durch Anwendungsbeispiele illustriert und durch Übungsaufgaben vertieft.

Zur Durchführung der Übungsaufgaben sollte dem Leser eine SPSS-Version mit LISREL-Option zur Verfügung stehen. Angehörige der Universität Trier können im Bedarfsfall zur Einarbeitung in die SPSS-Installation auf der Siemens-Anlage H90-D die URT-Broschüre "Einführung in SPSS 4.1 unter BS2000" verwenden (erhältlich in der Skriptenstelle des URT).

Leser, die bereits mit anderen Versionen oder Installationen von LISREL gearbeitet haben, können sich auf die Lektüre von Abschnitt 6 beschränken, der die Syntax der SPSS-Prozeduren PRELIS und LISREL beschreibt.

Für das Korrekturlesen sowie zahlreiche wertvolle Anregungen, die diese Arbeit erst möglich gemacht haben, danke ich meinen Kolleginnen und Kollegen, vor allem dem Leiter des Rechenzentrums der Universität Trier, Herrn Prof. Dr.-Ing. Manfred Paul. Weiterhin danke ich den Kursteilnehmern, durch deren Aufmerksamkeit viele Fehler entdeckt wurden. Hinweise auf verbliebene Fehler und Verbesserungsvorschläge nehme ich dankbar entgegen.

Inhaltsverzeichnis

Vorwort	III
1 Einleitung	1-1
1.1 Strukturgleichungsmodelle für direkt beobachtbare Variablen	1-1
1.2 Faktorenanalytische Modelle	1-6
1.3 Strukturgleichungsmodelle mit latenten Variablen (LISREL-Modelle)	1-9
2 Grundprinzipien der Kovarianzstrukturanalyse	2-1
2.1 Einige Voraussetzungen aus der Wahrscheinlichkeitstheorie	2-1
2.2 Erklärung einer Kovarianzmatrix durch ein Strukturgleichungsmodell	2-4
2.3 Identifikation der Parameter	2-8
2.4 Schätzung der Parameter	2-10
2.4.1 Unweighted Least Squares (ULS)	2-11
2.4.2 Generalized Least Squares (GLS)	2-13
2.4.3 Maximum Likelihood (ML)	2-13
2.4.4 Instrumental Variables (IV) und Two-Stage Least Squares (TSLS)	2-16
2.4.5 Hinweise zur praktischen Anwendung der Schätzverfahren	2-16
2.5 Beurteilung der Gültigkeit eines Modells	2-17
2.5.1 Untersuchung der Parameterschätzungen	2-17
2.5.2 Der χ^2 - Modelltest	2-18
2.5.3 Andere globale Fit-Indikatoren	2-20
2.5.4 Residuen und standardisierte Residuen	2-21
2.5.5 Modifikationsindikatoren	2-25
Übungsaufgaben	2-27
3 Das LISREL-Modell	3-1
3.1 Variablentypen	3-1
3.2 Meßmodell und Strukturgleichungsmodell	3-2
3.2.1 Meßmodell	3-2
3.2.2 Das Strukturgleichungsmodell	3-3
3.2.3 LISREL-Annahmen über Erwartungswerte und Kovarianzen	3-4
3.3 Die Kovarianzmatrizen der Quellvariablen	3-6
3.4 Die Kovarianzgleichung: $\Sigma = f(\theta)$	3-7
3.5 Submodelle	3-7
3.6 Feste, freie und restringierte Parameter	3-8
3.7 Modi und Formen der Parametermatrizen	3-10
3.8 Maßeinheiten der latenten Variablen	3-12
3.9 Modelle mit direkt beobachteten exogenen Variablen ($\xi \equiv X$)	3-13
Übungsaufgaben	3-14
4 PRELIS: Der LISREL-Preprozessor für problematische Daten	4-1
4.1 Variablentypen und Assoziationsmaße	4-1
4.1.1 Variablentypen	4-1
4.1.1.1 Kontinuierliche Variablen	4-1
4.1.1.2 Ordinale Variablen	4-1
4.1.1.3 Zensierte Variablen	4-2
4.1.2 Assoziationsmaße	4-2
4.1.2.1 Normal-Scores	4-2
4.1.2.2 Optimal-Scores	4-3
4.1.2.3 Korrelationsmaße	4-4
4.1.2.4 Andere Assoziationsmaße	4-5

4.2	Asymptotisch verteilungsfreie Schätzer	4-5
4.3	Simulationsstudie: Verhalten verschiedener Schätzmethoden bei ordinalen Indikatoren	4-9
5	LISREL-Ausgaben zur Modelldiagnose und zur Modellmodifikation	5-1
5.1	Modelldiagnose	5-1
5.1.1	Fit-Beurteilung	5-4
5.1.2	Quadrierte multiple Korrelationen und Determinationskoeffizienten	5-5
5.1.3	ML-Schätzer und T-Werte	5-7
5.1.4	Direkte, indirekte und totale Effekte, Stabilitätsindex	5-9
5.1.5	Varianzen und Kovarianzen	5-11
5.1.6	Standardisierte Lösung	5-11
5.2	Modellmodifikation	5-14
	Übungsaufgaben	5-16
6	Beschreibung der SPSS-Prozeduren PRELIS und LISREL	6-1
6.1	Zur Integration von PRELIS und LISREL in das SPSS-System	6-1
6.2	Datenaustausch zwischen dem SPSS-Basissystem, PRELIS und LISREL	6-3
6.3	Allgemeine Syntax-Regeln für die SPSS-Prozeduren PRELIS und LISREL	6-5
6.4	Konventionen zur Beschreibung der PRELIS- und LISREL-Subkommandos	6-6
6.5	Beschreibung der PRELIS-Subkommandos	6-7
6.5.1	Die PRELIS-Syntax im Überblick	6-7
6.5.2	Das VARIABLES-Subkommando	6-7
6.5.3	Das MISSING-Subkommando	6-9
6.5.4	Das MAXCAT-Subkommando	6-9
6.5.5	Das TYPE-Subkommando	6-10
6.5.6	Das MATRIX-Subkommando	6-11
6.5.7	Das WRITE-Subkommando	6-12
6.5.8	Das PRINT-Subkommando	6-13
6.5.9	Das CRITERIA-Subkommando	6-14
6.6	Beschreibung der LISREL-Subkommandos	6-15
6.6.1	Die LISREL-Syntax im Überblick	6-15
6.6.2	Titelzeilen	6-16
6.6.3	Daten-Spezifikation	6-17
6.6.3.1	Das DA(ta)-Subkommando	6-18
6.6.3.2	Lesen aus einer SPSS-Matrix-Systemdatei	6-20
6.6.3.3	Einlesen von Variablen aus der aktiven SPSS-Systemdatei	6-22
6.6.3.4	Einlesen von Rohdaten aus einer Textdatei (Subkommando RA(w data))	6-22
6.6.3.5	Einlesen einer Assoziationsmatrix aus einer Textdatei (Subkommandos CM, KM, MM, OM und PM)	6-25
6.6.3.6	Einlesen von Mittelwerten und Standardabweichungen aus einer Textdatei (Subkommandos ME(ans)- und SD(evs))	6-27
6.6.3.7	Einlesen von asymptotischen (Ko-)Varianzen der Assoziations- Statistiken (Subkommandos AC und AV)	6-28
6.6.3.8	Vereinbaren von Etiketten für die manifesten Variablen (Subkommando LA(bels))	6-29
6.6.3.9	Auswählen und Umordnen der Variablen für eine Analyse (Subkommando SE(lect))	6-30
6.6.4	Modell-Spezifikation	6-31
6.6.4.1	Das MO(dell)-Subkommando	6-32
6.6.4.2	Freisetzen bzw. Fixieren von Paramtern (Subkommandos FR(ee)- bzw. FI(x))	6-33

6.6.4.3	Freisetzung von Parametern bei der automatischen Modellsuche verhindern (Subkommando NF(ree))	6-35
6.6.4.4	Deklaration freier und fixierter Parameter durch eine Mustermatrix Subkommando PA(ttern))	6-35
6.6.4.5	Deklaration von Gleichheitsrestriktionen (Subkommando EQ(ual))	6-36
6.6.4.6	Spezifikation von Werten für eine Liste von Parametern (Subkommandos VA(lue)- und ST(art))	6-36
6.6.4.7	Spezifikation von (Start-)Werten für eine Parametermatrix (Subkommando MA)	6-37
6.6.4.8	Vereinbarung von Etiketten für latente Variablen (Subkommandos LE(ta) und LK(si)	6-38
6.6.4.9	Plotten der Fit-Funktion in Abhängigkeit von einem Parameter (Subkommando PL(ot))	6-39
6.6.5	Spezifikation der Ausgabe und der Schätzmethode (Subkommando OU(tput))	6-40
Übungsaufgaben	6-45
Lösungsvorschläge für die Übungsaufgaben	A-1
Literatur	A-6

1 Einleitung

Wesentliche Aufgabe empirischer Wissenschaften ist die Formulierung und Prüfung von Theorien zur **Erklärung** empirischer Sachverhalte. Für die speziellen Probleme, die dabei in den Sozial- und Wirtschaftswissenschaften auftreten, bietet LISREL (**L**inear **S**tructural **R**elationships) als Forschungsmethode bzw. als Computerprogramm moderne Lösungsmöglichkeiten.

Eine Besonderheit sozial- bzw. wirtschaftswissenschaftlicher Forschung liegt darin, daß zur Prüfung vieler Kausalhypothesen kontrollierte Experimente mit Manipulation der interessierenden Faktoren z.B. aus moralischen oder pragmatischen Gründen nicht in Frage kommen, so daß stattdessen eine Überprüfung anhand der statistischen Beziehungen (Korrelationen) zwischen beobachteten Variablen erfolgen muß. Eine typische sozialwissenschaftliche Theorie besteht aus mehreren Gleichungen, in denen jeweils die Abhängigkeit einer Variablen von einer oder mehreren anderen Variablen erklärt wird. Wenn dabei der Anspruch erhoben wird, daß die Gleichungen nicht nur statistische Assoziationen beschreiben, sondern **kausale Mechanismen**, bezeichnet man das Gleichungssystem als "**Strukturgleichungsmodell**". Jöreskog & Sörbom (1989, S. 1) formulieren den Anspruch von Strukturgleichungsmodellen so:

"A structural equation model is used to specify the phenomenon under study in terms of putative cause-and-effect variables and their indicators. Because each equation in the model represents a causal link rather than a mere empirical association, the structural parameters do not, in general, coincide with coefficients of regressions among observed variables. Instead, the structural parameters represent relatively unmixed, invariant and autonomous features of the mechanism that generates the observed variables."

Um die Koeffizienten in Strukturgleichungen mit Hilfe von Stichprobenstatistiken (z.B. durch Regressionskoeffizienten) schätzen zu können, bedarf es einer Methodologie, die wesentlich über die klassischen Verfahren der Regressions- und Varianzanalyse hinausgeht.

Ein weiteres Grundproblem der sozial- und wirtschaftswissenschaftlichen Methodologie besteht darin, daß wichtige theoretische Begriffe nur unzulänglich operationalisiert sind, so daß Theorien mit diesen Begriffen anhand von fehlerhaften Indikatoren überprüft werden müssen. Gerade bei diesem Problem hat LISREL durch die Integration strukturanalytischer und faktorenanalytischer Methoden erhebliche Fortschritte gebracht.

1.1 Strukturgleichungsmodelle für direkt beobachtbare Variablen

Betrachten wir zunächst Modelle, in denen mit Ausnahme der Störvariablen ausschließlich direkt beobachtbare Variablen auftreten. Viele dieser Modelle können mit den klassischen Methoden der Regressionsanalyse analysiert werden.

Beispiel 1.1: (Berufliche Ambitionen von Jugendlichen)

In einer Untersuchung zu den Bedingungen der beruflichen Ambitionen von Jugendlichen wurden bei 767 12-jährigen Jungen folgende Variablen erhoben:¹

X_1	Intelligenz des Jungen
X_2	Bildungsgrad des Vaters
X_3	Beruf des Vaters
Y_1	Schulleistung des Jungen
Y_2	Berufliche Ambitionen des Jungen

¹ Das Beispiel basiert auf einem Datensatz von Kerchoff (1974), der in der Literatur schon mehrfach analysiert wurde. Es soll nicht unerwähnt bleiben, daß die völlige Fixierung auf väterliche Karriere-Merkmale als nachdrückliche Aufforderung zur Lösung des Emanzipationsproblems verstanden werden muß.

Über das Zusammenwirken der Variablen wird folgendes Modell (bzw. folgende Theorie) aufgestellt:

$$Y_1 = \gamma_{11}X_1 + \gamma_{12}X_2 + \zeta_1 \quad (1.1.1)$$

$$Y_2 = \gamma_{23}X_3 + \beta_{21}Y_1 + \zeta_2 \quad (1.1.2)$$

In der ersten Gleichung wird u.a. behauptet, daß eine Erhöhung der Intelligenz um eine Maßeinheit bei Konstanzhaltung des väterlicher Bildungsgrades im Mittel zu einer Veränderung der beruflichen Ambitionen um γ_{11} Maßeinheiten führt. Analog sind die übrigen γ - und β -Koeffizienten zu interpretieren.

Die Modellgleichungen machen Aussagen über **zentrierte Variablen**, d.h. über Variablen mit Erwartungswert Null, und enthalten daher keine additiven Konstanten. LISREL ist selbstverständlich auch bei Modellen mit additiven Konstanten anwendbar. Bei der vorliegenden Fragestellung sind die (unbedingten) Erwartungswerte der beteiligten Variablen allerdings uninteressant und sie werden daher durch Betrachtung der zentrierten Variante $Z := Z' - E(Z')$ zu jeder interessierenden Variablen Z' der Einfachheit halber eliminiert.

In den Modellgleichungen werden **lineare und additive Effekte** der jeweiligen unabhängigen Variablen auf die abhängige Variable postuliert. Diese Einschränkung kann jedoch durch geschickte Vorbehandlung der unabhängigen Variablen (z.B. logarithmische Transformation, Bildung von Produktvariablen) teilweise überwunden werden.

Die X -Variablen tauchen in keiner Modellgleichung als abhängige Variable auf. Ihr Bedingungsgrund liegt vollständig außerhalb des betrachteten Variablensystems. Sie werden daher als "**exogen**" (griechisch: von externem Ursprung) bezeichnet. Für die Y -Variablen werden demgegenüber kausale Einflüsse durch andere Variablen des betrachteten Systems angenommen. Sie heißen daher "**endogen**". Genauer gesagt dürfen die Y -Variablen von exogenen Variablen sowie von anderen endogenen Variablen abhängen. Das Modell enthält für jede Y -Variable eine Gleichung.

In der Regel kann eine endogene Variable nicht vollständig durch die exogenen und die übrigen endogenen Variablen des Modells erklärt werden. Daher wird in die i -te Gleichung die **Fehler-, Residual-, oder Störvariable** ζ_i aufgenommen. Sie repräsentiert alle nicht im Modell enthaltenen kausalen Einflüsse auf die endogene Variable Y_i .

Die Stör- oder Residualvariablen sind offenbar auch von externem Ursprung, allerdings soll der Ausdruck "exogen" für die X -Variablen reserviert bleiben. In diesem Manuskript werden gelegentlich exogene Variablen und Störvariablen gemeinsam als "**Quellvariablen**" bezeichnet; dies ist aber kein üblicher statistischer Terminus.

Die Stärke eines kausalen Effektes wird durch den zugehörigen "**Strukturkoeffizienten**" γ_{ij} bzw. β_{ij} ausgedrückt, wobei folgende Bezeichnungsvereinbarungen gelten:

- Der Effekt einer exogenen Variablen X_j auf eine endogene Variable Y_i wird mit γ_{ij} bezeichnet.
- Der Effekt einer endogenen Variablen Y_j auf eine andere endogene Variable Y_i wird mit β_{ij} bezeichnet.

Der erste Index gibt also jeweils die abhängige Variable und der zweite Index die unabhängige Variable an. Im Rahmen der Modellentwicklung möchte man die Strukturkoeffizienten möglichst genau schätzen und Hypothesen über ihre wahren Ausprägungen prüfen (z.B. $H_0: \gamma_{11} = 0$ vs. $H_1: \gamma_{11} \neq 0$).

Die **Varianzen und Kovarianzen der Quellvariablen** treten zwar in den Gleichungen nicht explizit auf, sind aber ebenfalls essentielle Bestandteile des Modells. Ein Grundprinzip der in diesem Kurs behandelten Forschungsmethoden ist nämlich die Aufklärung der Kovarianzen und Varianzen beobachteter Variablen. Wie in Abschnitt 2 gezeigt wird, sind bei diesen Erklärungs Bemühungen neben den Strukturkoeffizienten in den Gleichungen die Kovarianzen und Varianzen der Quellvariablen als wesentliche Modellparameter

beteiligt. Die Varianz von X_i soll mit ϕ_{ii} und die Kovarianz von X_i und X_j mit ϕ_{ij} bezeichnet werden, $i, j=1, \dots, 3$, für die Varianz von ζ_i schreibt man ψ_{ii} und für die Kovarianz von ζ_i und ζ_j schreibt man ψ_{ij} , $i, j=1, 2$.

Für unser spezielles Modell soll allerdings Unkorreliertheit der Residualvariablen ζ_1 und ζ_2 angenommen werden, so daß die folgenden, frei schätzbaren Parameter verbleiben:

$$(\gamma_{11}, \gamma_{12}, \gamma_{23}, \beta_{21}, \phi_{11}, \phi_{21}, \phi_{22}, \phi_{31}, \phi_{32}, \phi_{33}, \psi_{11}, \psi_{22})$$

Besonders bei größeren Modellen ist die graphische Darstellung durch ein **Pfaddiagramm** sehr zu empfehlen. Die folgende Abbildung 1.1 zeigt das Pfaddiagramm zu unserem Beispiel 1.1:

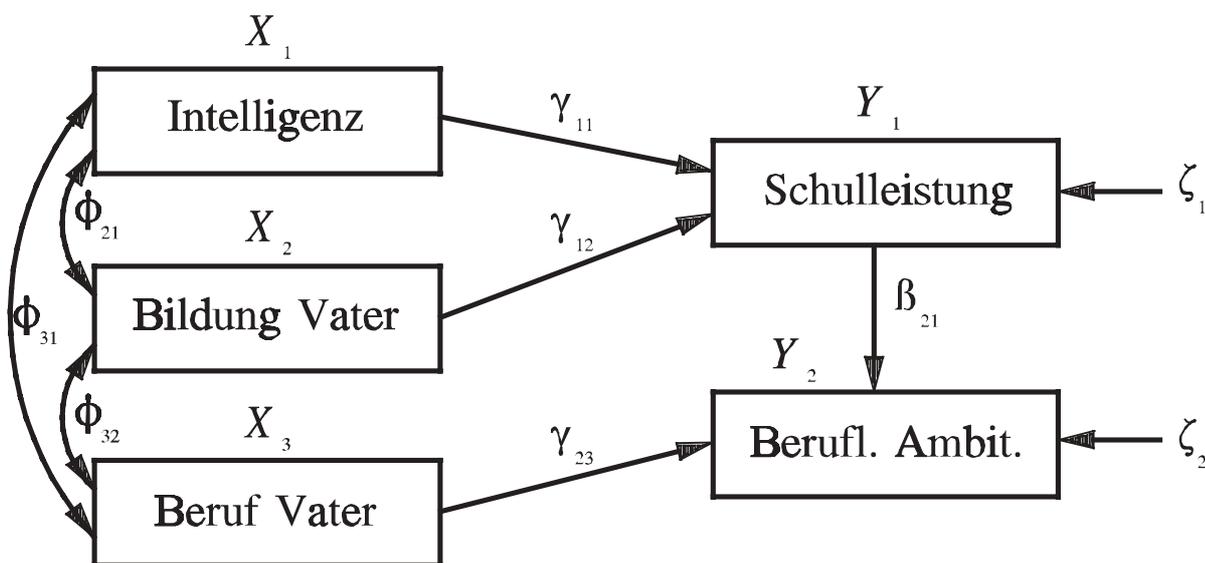


Abb. 1.1 Pfaddiagramm zu den beruflichen Ambitionen von Jugendlichen

Alle in den Modellgleichungen behaupteten direkten Effekte sind durch gerichtete Pfeile dargestellt. Erlaubt das Modell eine Kovariation zwischen zwei Quellvariablen, so wird diese durch einen gebogenen Doppelpfeil dargestellt. Durch Weglassen eines gerichteten Pfeils oder eines gebogenen Doppelpfeils wird also *explizit* behauptet, daß der zugehörige Koeffizient gleich Null sei. Unser Modell sieht z.B. keinen direkten Effekt von X_1 (Intelligenz) auf Y_2 (berufliche Ambitionen) vor; es wird nur ein indirekter, von Y_1 (Schulleistung) vermittelter, Effekt angenommen.

Da LISREL meist mit dem Ziel einer **Kausalanalyse** eingesetzt wird, müssen wir kurz auf das methodologisch und wissenschaftstheoretisch anspruchsvolle Thema der kausalen Interpretierbarkeit statistischer Ergebnisse eingehen. Für eine gründliche Behandlung der Thematik wird auf Steyer (1988, 1992) verwiesen.

Damit die geschätzten Strukturkoeffizienten der exogenen (X -)Variablen **kausal** interpretiert werden können, muß für die X -Variablen **Unkorreliertheit mit allen Störvariablen** angenommen werden. Diese Unkorreliertheit sorgt nämlich unter gewissen Zusatzannahmen für die **Invarianz** der geschätzten

Strukturkoeffizienten aller exogenen Variablen bei Erweiterung einer Modellgleichung um beliebige zusätzliche Einflußgrößen, die vorher Bestandteil des Residuums waren. Diese Invarianz ist das zentrale Kriterium moderner Kausalitätsberiffe¹. In der klassischen Experimental-Methodologie wird die Unkorreliertheits-Voraussetzung und damit die Invarianzbedingug durch **Randomisierung**, d.h. durch zufällige Auswahl der bei einem Fall zu realisierenden experimentellen Bedingung, sichergestellt. Wie eingangs erwähnt, ist allerdings in den Wirtschafts- und Sozialwissenschaften die Experimental-Technologie häufig aus ethischen und/oder pragmatischen Gründen nicht einsetzbar. Bei der ersatzweise gewählten Strukturgleichungsanalyse von beobachteten Variablen muß dann die Unkorreliertheit der X-Variablen mit den Störvariablen direkt angenommen werden. Geschieht dies zu Unrecht, so können grob verzerrte Schätzer für kausale Effekte resultieren (man spricht vom "**omitted variable**"-Fehler). Dies wird nun an einem sehr einfachen Strukturgleichungsmodell demonstriert, dessen einzige Gleichung die Abhängigkeit einer endogenen Variablen Y von einer exogenen Variable X beschreibt:

Ein nicht sehr aufgeklärter Forscher hat die Hypothese aufgestellt, daß die Geburtenrate (Anzahl der jährlichen Geburten pro 10000 Einwohner, Abk.: Y) von der Anzahl der Klapperstörche (Abk.: X) abhängt und möchte anhand der Daten aus schwedischen Landkreisen die Stärke γ des kausalen Effektes von X auf Y schätzen. Sein Modell lautet:

$$Y = \gamma X + \zeta$$

Unter der kritischen Annahme der Unkorreliertheit von X und ζ schätzt der Forscher aus einer Stichproben-Kovarianzmatrix mit den beobachteten Variablen X und Y den Effekt $\hat{\gamma} = 0.6$. Er folgert, daß die Zunahme der Storchenpopulation um ein Exemplar eine Steigerung der Geburtenrate um ca. 0.6 bewirkt.

Die Residualvariable ζ enthält in unserem Beispiel u.a. als bedeutende Einflußgröße den Industrialisierungsgrad, welcher die Geburtenrate negativ beeinflusst. Seine Nicht-Berücksichtigung allein gefährdet die Interpretierbarkeit von $\hat{\gamma}$ noch nicht. Allerdings ist der Industrialisierungsgrad ebenfalls negativ mit der Anzahl der Störche korreliert. Die Unkorreliertheits-Bedingung ist also verletzt und damit darf $\hat{\gamma}$ nicht als kausaler Effekt von X auf Y interpretiert werden. In der Tat ist $\hat{\gamma}$ eine drastisch verzerrte Schätzung für den wahren Effekt, der vermutlich exakt Null beträgt. Offenbar führt der negative Effekt des Industrialisierungsgrades auf Y und X zu einer positiven Korrelation zwischen diesen Variablen. Daraus resultiert bei einer Analyse unter der falschen Annahme der Unkorreliertheit von X und ζ die unsinnige Effektschätzung $\hat{\gamma}$. Ihre Nichtigkeit erweist sich z.B. bei Erweiterung obiger Strukturgleichung um die unabhängige Variable Industrialisierungsgrad. Dann wird vermutlich für γ eine Schätzung nahe Null resultieren, d.h. die Invarianzbedingung gilt nicht.

Bei endogenen (Y -)Variablen, die in einer Modellgleichung in explanatorischer Position auftreten (z.B. Y_1 in (1.1.2)), kann ohne Schaden für die kausale Interpretierbarkeit der geschätzten Strukturkoeffizienten eine von Null verschiedene Korrelation mit der zugehörigen Störvariablen zugelassen werden, sofern diese im Modell geeignet berücksichtigt wird (s.u.). Sie *muß* für eine erfolgreiche Modellierung vieler empirischer Systeme sogar zugelassen werden, z.B. bei Vorliegen einer wechselseitigen Abhängigkeit zwischen zwei endogenen Variablen.

Anhand von Beispiel 1.1 soll demonstriert werden, wann eine Korrelation zwischen einer endogenen unabhängigen Variablen und der zugehörigen Störvariablen nicht zu verzerrten Schätzungen für den Strukturkoeffizienten führt. Wir wollen annehmen, daß Y_1 mit ζ_2 , also der Störvariablen aus der zweiten Modellgleichung (1.1.2), korreliert sei. Da Y_1 modellgemäß die Summe aus einer Linearkombination der exogenen Variablen und ζ_1 ist und außerdem die exogenen Variablen gemäß der eben diskutierten

¹ Eine exakte Untersuchung notwendiger und hinreichender Bedingungen für starke kausale Regressionsmodelle findet sich in Steyer (1992, S. 128ff).

Generalvoraussetzung mit ζ_2 unkorreliert sind, ist unsere Annahme gleichbedeutend mit der Unterstellung einer Kovariation der beiden Störvariablen ζ_1 und ζ_2 ¹. Sie tritt z.B. dann auf, wenn im Modell exogene Variablen (X_4, X_5, \dots) mit Einfluß auf Y_1 und Y_2 weggelassen wurden, was bei komplexen sozialen Sachverhalten häufig der Fall sein dürfte. Diese Kovarianz von ζ_1 und ζ_2 kann aber im LISREL-Modell für unser Problem als frei schätzbarer Parameter ψ_{12} berücksichtigt werden (s.o.). Damit wird die artifizielle Inflation des Schätzers für den direkten Effekt β_{21} von Y_1 auf Y_2 verhindert; es tritt kein "omitted variable" - Fehler auf.

Eine analoge Berücksichtigung der Korrelation zwischen einer exogenen unabhängigen Variablen und einer Störvariablen ζ_i ist nicht möglich.

Für eine **Regressionsgleichung** wird üblicherweise gefordert, daß *alle* unabhängigen Variablen mit dem Residuum unkorreliert sind. Dies macht Systeme von Regressionsgleichungen für die Modellierung vieler empirischer Phänomene ungeeignet. Bei einem Strukturgleichungsmodell wird demgegenüber nur für exogene unabhängige Variablen die Unkorreliertheit mit den Residuen vorausgesetzt. Daher erfordert die Analyse von Strukturgleichungsmodellen eine über traditionelle Regressionstechniken hinausgehende Methodologie, die z.B. in LISREL realisiert ist.

Aus dem Gesagten folgt, daß ein System aus Regressionsgleichungen auch ein Strukturgleichungsmodell ist, falls für die Regressionskoeffizienten eine kausale Interpretierbarkeit angenommen wird.

Wenn wir in unserem Beispiel 1.1, z.B. aus den eben besprochenen kausalitäts-theoretischen Gründen, eine Kovariation von ζ_1 und ζ_2 zulassen und ψ_{12} als frei schätzbaren Parameter in unser Modell aufnehmen, dann ist die Strukturgleichung (1.1.2) wegen der resultierenden Korrelation zwischen Y_1 und ζ_2 keine Regressionsgleichung.

Wenn wir aber in der Überzeugung, daß der wahre Wert von ψ_{12} Null ist, zugunsten eines möglichst einfachen Modells die Unkorreliertheit von ζ_1 und ζ_2 postulieren, dann wird (1.1.2) zur Regressionsgleichung. Das Pfaddiagramm in Abb. 1.1 zeigt genau diese Modellvariante: ζ_1 und ζ_2 sind nicht durch einen gebogenen Doppelpfeil verbunden. In dieser Variante erfüllt unser Strukturgleichungsmodell die beiden Bedingungen für sogenannte "**rekursive**" Systeme:

- Die endogenen Variablen können so nummeriert werden, daß endogene Variablen keine kausalen Effekte auf andere endogene Variablen mit niedrigeren Nummern ausüben.
- Die Residualvariablen aus verschiedenen Gleichungen sind unkorreliert.

Für die Modellierung empirischer Systeme durch rekursive Strukturgleichungsmodelle hat sich der Ausdruck "**Pfadanalyse**" eingebürgert.

Strukturgleichungsmodelle, die eine der obigen Bedingungen verletzen, heißen "**nonrekursiv**"². Besonders beeindruckende Vertreter dieser Klasse sind Systeme, in denen kausale Sequenzen auftreten, die direkt (z.B.: Y_1 beeinflusst Y_2 und Y_2 beeinflusst Y_1) oder indirekt (z.B. Y_1 beeinflusst Y_2 , Y_2 beeinflusst Y_3 und Y_3 beeinflusst Y_1) zum Ausgangsort zurückführen. In Beispiel 1.3 wird ein nonrekursives System mit zwei wechselseitig voneinander abhängigen latenten Variablen η_1 und η_2 vorgestellt (siehe Abb. 1.3).

¹ Diese Aussagen folgen aus Theoremen über Varianzen, Kovarianzen und Korrelationen, die in Abschnitt 2.1 wiederholt werden.

² Der Begriff "rekursiv" verursacht gelegentlich Mißverständnisse, z.B. bei EDV-gebildeten Personen, weil ein "Zurückkommen" doch ausgerechnet bei den als "nonrekursiv" bezeichneten Modellen auftritt. Man denke jedoch daran, daß in der Logik und in der Mathematik eine "rekursive Definition" vorschreibt, wie spätere Ausdrücke aus unmittelbar vorausgehenden bestimmt werden können. Ist jedoch in einem Strukturgleichungsmodell Y_2 nur unter Verwendung von Y_1 zu ermitteln, andererseits aber Y_1 nur bei Bekanntheit von Y_2 bestimmbar, dann ist offenbar ein systematischer Aufbau im Sinne der rekursiven Definition nicht möglich. Ein solches System wird daher zurecht als "nonrekursiv" bezeichnet.

Wir werden im folgenden die in Abb. 1.1 gezeigte rekursive Variante von Beispiel 1.1 zugrundelegen, wohl wissend, daß die Unabhängigkeitsannahme für ζ_1 und ζ_2 keine nebensächliche Vereinbarung ist, sondern eine handfeste empirische Behauptung. Wir negieren damit z.B. die Existenz von exogenen Variablen (X_4, X_5, \dots) mit Einfluß auf Y_1 und Y_2 , die im Modell weggelassen wurden, und riskieren verzerrte Schätzungen für den Strukturkoeffizienten β_{21} , falls wir uns irren. Allerdings läßt uns LISREL in diesem Fall nicht kommentarlos mit verzerrten Schätzern allein, sondern es erlaubt durch mehrere Indikatoren für die Güte unseres Modells eine empirische Prüfung der Unabhängigkeitsannahme. Damit gibt uns Beispiel 1.1 Gelegenheit, LISREL anhand einer relativ einfachen Problemstellung kennenzulernen, die auch noch mit den Methoden der klassischen Regressionsanalyse bearbeitet werden kann. Andererseits wird sich jedoch zeigen, daß sich aufgrund der Verfügbarkeit von **Modellanpassungstests** der Einsatz von LISREL auch bei rekursiven Strukturgleichungsmodellen lohnen kann.

1.2 Faktorenanalytische Modelle

Von zentraler Bedeutung für Beispiel 1.1 sowie für alle vergleichbaren Modelle mit direkt beobachtbaren Variablen ist die Annahme, daß alle in explanatorischer Funktion auftretenden Variablen **fehlerfrei gemessen** werden können. Anderenfalls können nämlich die interessierenden Effekt-Parameter γ_{ij} und β_{ij} nicht unverzerrt geschätzt werden (siehe Aufgabe 2.1; vgl. Bollen 1989, Kapitel 5).

Die meisten sozial- bzw. wirtschaftswissenschaftlichen Modelle enthalten aber **latente Variablen (theoretische Begriffe, hypothetische Konstrukte)**, z.B. sozioökonomischer Status, Leistungsmotivation, Einstellung zu Ausländern, für die es keine perfekte Meßmethode gibt, sondern nur eine Anzahl mehr oder weniger fehlerhafter empirischer Indikatoren.

Postuliert ein Wissenschaftler eine Strukturgleichung und hat er dabei für eine exogene Variable lediglich fehlerhafte Indikatoren zur Verfügung, so bietet die regressionsanalytische Methodologie nur zwei unbefriedigende Auswege an:

- Regressionsrechnung unter Verwendung einer Funktion der fehlerhaften Indikatoren (z.B. Summe) Dabei erhält man jedoch potentiell für *alle* Strukturkoeffizienten der Gleichung verzerrte Schätzungen (vgl. Bollen 1989, S. 164). Diese lassen sich nur in einfachen Situationen korrigieren, sofern die Reliabilitäten der Indikatoren bekannt sind (z.B. Minderungskorrektur für Regressions- und Korrelationskoeffizienten).
- Weglassen der betroffenen exogenen Variablen Hierbei können drastisch verzerrte Schätzungen für die Strukturkoeffizienten zu den verbliebenen exogenen Variablen resultieren, falls diese nicht unabhängig von der weggelassenen Variablen sind (vgl. obiges Klapperstorch-Beispiel).

Man muß also die regressionsanalytische Methodologie so erweitern, daß unter Verwendung fehlerhafter Indikatoren konsistente Schätzungen für Effekte auf der Ebene latenter Variablen möglich werden.

Wertvolle Ansätze für solche Erweiterungen liefert die **Faktorenanalyse**. Ihr Grundmodell besteht aus Regressionsgleichungen, in denen jeweils eine beobachtete Variable auf die linear kombinierte Wirkung latenter Variablen und einen Restanteil zurückgeführt wird. Das Modell erklärt die Korrelationen (oder Kovarianzen) zwischen den beobachteten Variablen durch deren Abhängigkeit von den latenten Variablen und durch die Korrelationen (oder Kovarianzen) zwischen den latenten Variablen.

Die Faktorenanalyse wird in der Test- bzw. Meßtheorie erfolgreich zur Lösung der methodologischen Problemen beim Übergang von direkt beobachtbaren ("**manifesten**") zu latenten Variablen eingesetzt (siehe z.B. Steyer 1988). Im Kontext der (klassischen) Testtheorie läßt sich so etwa ein Paralleltestmodell

explizieren und empirisch prüfen. Wichtige testtheoretische Größen wie z.B. die Reliabilität sind durch die Parameter eines faktorenanalytischen Modells darstellbar.

Beispiel 1.2: (Ein Latent-State-Modell der Angst¹)

Im Zusammenhang mit der Messung von Veränderungen bei einer latenten psychologischen Variablen wie z.B. der Angst tritt das Problem auf, die Stabilität der latenten Variablen und die Reliabilität der manifesten Indikatoren analytisch zu trennen, damit z.B. nicht aus Reliabilitätsmängeln fälschlich auf hohe Variabilität der latenten Variablen geschlossen wird. Anhand des nun zu beschreibenden Beispiels für ein Strukturgleichungsmodell mit manifesten *und* latenten Variablen wird später demonstriert, wie eine getrennte Beurteilung von Stabilität und Reliabilität erfolgen kann.

Ein Test zur Messung von Zustandsangst wurde bei 179 Studenten der Universität Trier zweimal im Abstand von zwei Monaten erhoben. Die Ergebnisse für je eine Testhälfte werden als ein (fehlerbehafteter) Indikator für die latente Variable Zustandsangst aufgefaßt. Die latente Zustandsangst wird als zeitlich variabel angenommen, so daß für die beiden Zeitpunkte unterschiedliche, gleichwohl korrelierte, latente Variablen postuliert werden.

Es werden also folgende Variablen betrachtet:

X_1	Erste Testhälfte, erhoben zum ersten Zeitpunkt
X_2	Zweite Testhälfte, erhoben zum ersten Zeitpunkt
X_3	Erste Testhälfte, erhoben zum zweiten Zeitpunkt
X_4	Zweite Testhälfte, erhoben zum zweiten Zeitpunkt
ξ_1	Latente Zustandsangst zum ersten Zeitpunkt
ξ_2	Latente Zustandsangst zum zweiten Zeitpunkt

Die Bezeichnung von manifesten Variablen durch lateinische und von latenten Variablen durch griechische Buchstaben wird in der LISREL-Literatur weitgehend einheitlich verwendet. Aus diesem Grund wurde in Beispiel 1.1 die (nicht beobachtbare!) Residualvariable zu Y_i durch ζ_i , $i=1,2$, bezeichnet.

Die Annahmen des zu prüfenden **Meßmodells** werden in Gleichungen spezifiziert, die im Unterschied zu den Strukturgleichungen des ersten Beispiels manifeste *und* latente Variablen enthalten:

$$\begin{array}{rcll} X_1 & = & \lambda_{11}\xi_1 & + \delta_1 \\ X_2 & = & \lambda_{21}\xi_1 & + \delta_2 \\ X_3 & = & & \lambda_{32}\xi_2 + \delta_3 \\ X_4 & = & & \lambda_{42}\xi_2 + \delta_4 \end{array}$$

Mit λ_{ij} wird die Ladung der manifesten Variablen X_i auf der latenten Variablen ξ_j bezeichnet, $i=1..4$, $j=1,2$.

Für die Residualvariablen δ_i , $i=1,..,4$, zu den manifesten Variablen X_i muß wiederum aus den oben dargestellten kausalitäts-theoretischen Gründen vorausgesetzt werden, daß sie mit allen latenten (ξ)-Variablen unkorreliert sind, damit die Ladungen unverzerrt geschätzt werden können. Aus modellimmanenten Gründen wird zusätzlich postuliert, daß δ_i und δ_j , $i \neq j$, unkorreliert sind. Außerdem wird wieder der Einfachheit halber für alle Variablen der Erwartungswert Null angenommen.

Es werden ferner folgende **Gleichheits-Restriktionen** behauptet:

$$\begin{array}{l} \lambda_{11} = \lambda_{21} = \lambda_{32} = \lambda_{42} \\ \text{Var}(\delta_1) = \text{Var}(\delta_2) = \text{Var}(\delta_3) = \text{Var}(\delta_4) \end{array}$$

¹ Dieses Beispiel einschließlich zugehöriger empirischer Daten hat mir dankenswerterweise Rolf Steyer überlassen (vgl. Steyer 1988, S. 361ff).

Dies ist insgesamt ein recht strenges Meßmodell, das sehr viele prinzipiell mögliche Kovarianzmatrizen für die X -Variablen verbietet. Es erlaubt eine Prüfung der relativ restriktiven testtheoretischen Annahme der τ_k -Äquivalenz für die manifesten Variablen (vgl. Steyer, 1988, S. 345ff).

Die Kovarianzmatrix der ξ -Variablen darf beliebig sein, folglich dürfen die beiden latenten Variablen beliebige Varianzen haben und beliebig miteinander kovariieren.

Damit enthält unser Modell insgesamt die folgenden, frei schätzbaren Parameter (die beiden Gruppen gleich gesetzter Parameter sind jeweils durch *einen* Repräsentanten vertreten):

$$(\lambda_{11}, \phi_{11}, \phi_{21}, \phi_{22}, \text{Var}(\delta_1))$$

Wie LISREL-Kenner schon bemerkt haben werden, ist dieser Parametervektor noch "überbesetzt". Überlegungen in den Abschnitten 2.2 und 2.3 werden dazu führen, den Parameter λ_{11} , und damit wegen der Gleichheitsrestriktion alle Ladungsparameter, auf den Wert 1 zu fixieren, so daß dann nur noch 4 frei schätzbare Parameter verbleiben.

Die folgende Abbildung 1.2 zeigt das Pfaddiagramm nach jetzigem Stand unserer Überlegungen:

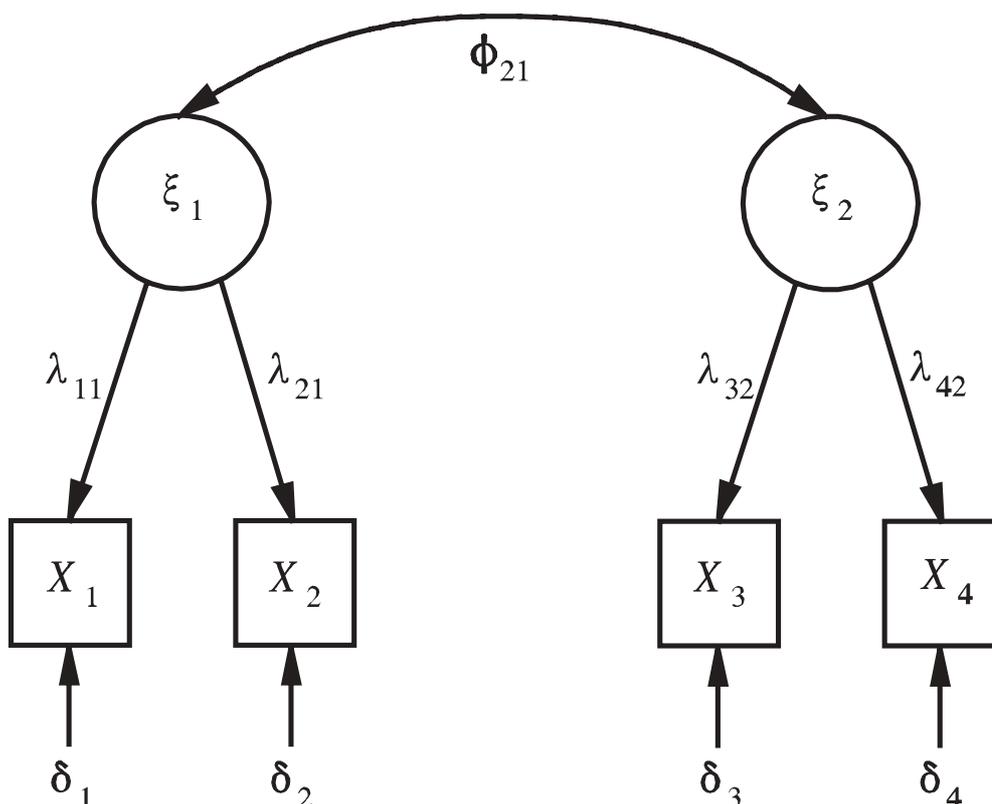


Abb. 1.2 Pfaddiagramm für das Latent-State-Modell der Angst

Es ist üblich, latente Variablen in Kreisen (oder Ovalen) und manifeste Variablen in Quadraten (oder Rechtecken) darzustellen. Residualvariablen werden allerdings nicht eingerahmt.

Faktorenanalytische Forschung kann **exploratorisch** oder **konfirmatorisch** betrieben werden. Bei einer rein exploratorischen Faktorenanalyse geht man von der Kovarianz- oder Korrelationsmatrix für q beobachtete Variablen aus und überläßt einem Algorithmus (z.B. Hauptachsenextraktion und anschließen-

de Varimaxrotation) die Ermittlung eines faktorenanalytischen Modells. Bei der moderneren konfirmatorischen Faktorenanalyse formuliert man wie in unserem Beispiel 1.2 vor der Datenerhebung ein faktorenanalytisches Modell und testet anschließend, wie gut dieses Modell die Stichproben-Kovarianz- oder Korrelationsmatrix der beobachteten Variablen erklären kann. Ein wesentlicher Vorteil des konfirmatorischen Vorgehens ist die Verfügbarkeit von Signifikanztests für die Gültigkeit des Modells und auch für einzelne Modellparameter. Eine exploratorische Faktorenanalyse liefert nur Parameterschätzungen.

1.3 Strukturgleichungsmodelle mit latenten Variablen (LISREL-Modelle)

Viele sozialwissenschaftliche Theorien enthalten mehrere Strukturgleichungen wie in unserem Beispiel 1.1, wobei jedoch latente Variablen an Stelle der dortigen beobachtbaren X - und Y -Variablen stehen. Man ist natürlich an den "reinen" Effekten zwischen den latenten Variablen interessiert, die in den Beziehungen zwischen fehlerbehafteten Indikatoren nur verzerrt und abgeschwächt erscheinen. Allerdings sind nur die fehlerhaften Indikatoren verfügbar. Man muß folglich Modelle formulieren, die

- einerseits die angenommenen kausalen Hypothesen auf der Ebene latenter Variablen enthalten und
- andererseits für jede latente Variable eine Verankerung in der Beobachtungsbasis in Form eines Meßmodells wie in unserem Beispiel 1.2 besitzen.

Genau dies ermöglicht LISREL. Das Programm kann u.a. bei folgenden Fragestellungen eingesetzt werden:

- Analyse von Strukturgleichungsmodellen für manifeste Variablen (vgl. Beispiel 1.1)
Auch hier ist LISREL oft den klassischen regressionsanalytischen Methoden überlegen (z.B. bei nonrekursiven Modellen).
- Konfirmatorische Faktorenanalyse
Z.B. Schätzung und Prüfung testtheoretischer Modelle (vgl. Beispiel 1.2)
- Schätzung minderungskorrigierter Korrelationen
- Strukturgleichungsmodelle mit latenten Variablen (vgl. Beispiel 1.3)
- Vergleiche der Kovarianzmatrizen oder der Faktorladungsmatrizen aus mehreren Populationen

Beispiel 1.3: (Konstruiertes Modell)

Das folgende konstruierte Modell aus Jöreskog & Sörbom (1989, Abschnitt 1.3) demonstriert, welche Analysemöglichkeiten LISREL bietet. Es enthält drei latente exogene Variablen (die ξ -Variablen), gemessen durch sieben Indikatoren (die X -Variablen), sowie zwei latente endogene Variablen (die η -Variablen), gemessen durch vier Indikatoren (die Y -Variablen). Das Modell enthält Residuen für die Indikatoren der η -Variablen und für die Indikatoren der ξ -Variablen, stellt also Meßfehler bei endogenen und bei exogenen Variablen in Rechnung. Dabei wird die Störvariable zu Y_i im Unterschied zu Beispiel 1.1 mit ε_i (statt ζ_i) bezeichnet.

Für die beiden η -Variablen wird eine reziproke Abhängigkeit behauptet, weshalb für ihre Residuen (ζ_1 und ζ_2) eine von Null verschiedene Kovarianz vorgesehen ist.

Da an dieser Stelle nur ein Eindruck von einem komplexen LISREL-Strukturgleichungsmodell mit latenten Variablen vermittelt werden soll, genügt die Wiedergabe des Pfaddiagramms. Die zugehörigen Gleichungen werden später zur Erläuterung eines vollständigen LISREL-Modells herangezogen.

Daß im Pfaddiagramm die Ladungen der X - bzw. Y -Variablen zwecks Unterscheidbarkeit durch ein entsprechendes Superskript gekennzeichnet sind, bedarf eigentlich keiner Erwähnung (z.B. $\lambda_{21}^{(x)}$). Eventuell irritieren allerdings die auf Eins fixierten Faktorladungen. Wir werden später sehen, daß damit keine empirischen Behauptungen aufgestellt, sondern die Maßeinheiten der latenten Variablen verankert werden.

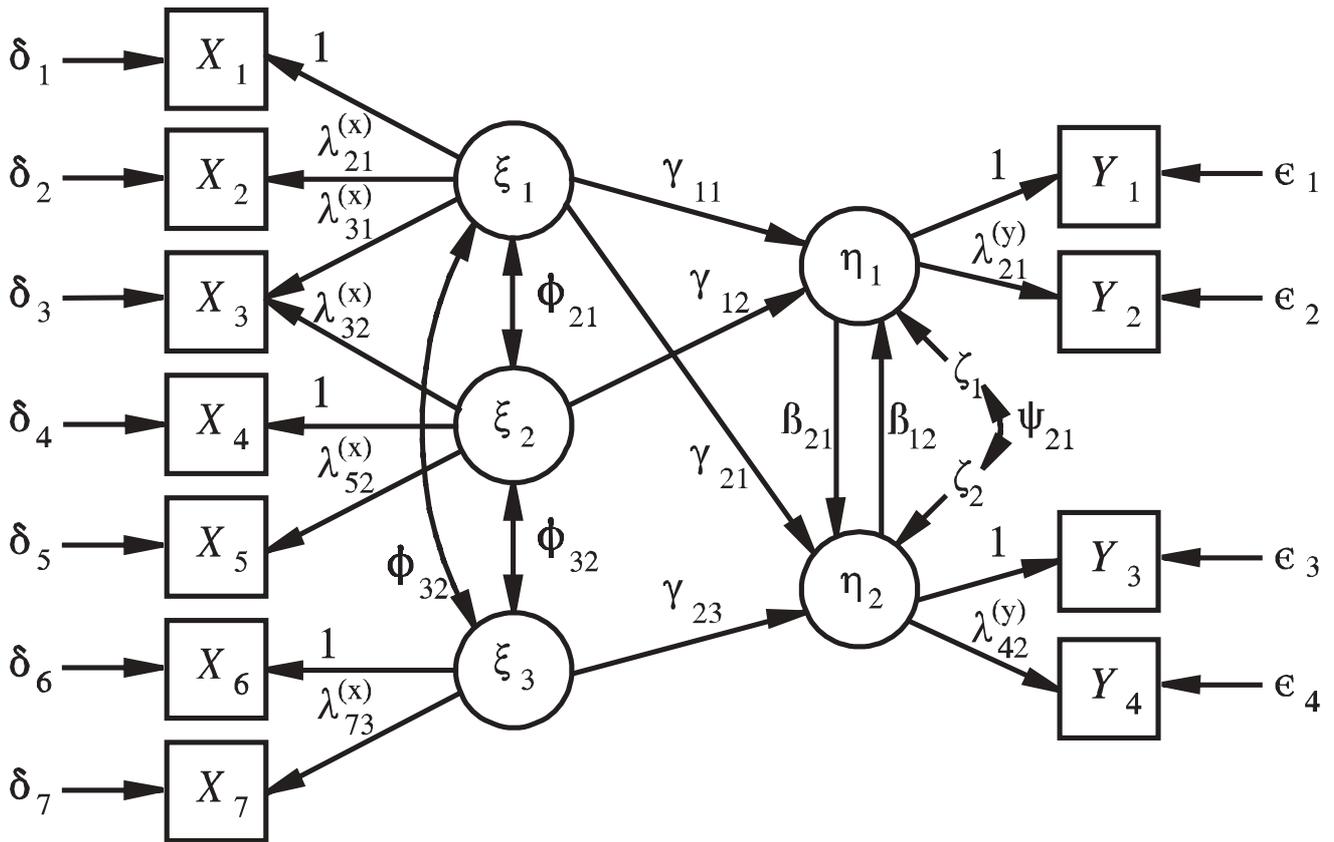


Abb. 1.3 Pfaddiagramm für das konstruierte Beispiel

2 Grundprinzipien der Kovarianzstrukturanalyse

Alle in der Einleitung angesprochenen Methoden gehen aus von einer Kovarianzmatrix (oder Korrelationsmatrix) manifester Variablen und überprüfen Hypothesen bzw. Theorien über die manifesten Variablen anhand ihrer Fähigkeit, diese Kovarianzmatrix (oder Korrelationsmatrix) zu erklären. Daher kann die Anwendung dieser Methoden treffend als "**Kovarianzstrukturanalyse**" bezeichnet werden. Die Grundprinzipien dieser Methoden-Familie sollen im 2. Abschnitt entwickelt werden.

Wegen seiner Integration mehrerer traditioneller Methoden unterstützt LISREL eine sehr große Klasse kovarianzstrukturanalytischer Modelle.

2.1 Einige Voraussetzungen aus der Wahrscheinlichkeitstheorie

Eine elementare Annahme der Kovarianzstrukturanalyse lautet, daß die Verteilung der beobachteten Variablen hinreichend gut durch die **Momente erster und zweiter Ordnung (d.h. durch Erwartungswerte, Varianzen und Kovarianzen)** beschrieben werden kann. Diese Annahme ist im Fall multivariater Normalverteilung der manifesten Variablen perfekt erfüllt.

Zur Wiederholung sollen die Definitionen dieser Momente und ihre wichtigsten Eigenschaften notiert werden, wobei auf mathematische Exaktheit (z.B. Existenzvoraussetzungen der Definitionen) verzichtet wird.

Erwartungswert

Der Erwartungswert einer Zufallsvariablen wird in diesem Manuskript selten explizit betrachtet, weil durchweg für alle Variablen Zentriertheit (d.h. ein Erwartungswert von Null) angenommen wird, was im Rahmen der klassischen Kovarianzstrukturanalyse ohne Beschränkung der Allgemeinheit möglich ist. Jedoch basieren viele der später einzuführenden Definitionen auf dem Erwartungswert, so daß unser kleiner Exkurs in die mathematische Statistik mit diesem elementaren Begriff beginnen soll.

Für eine diskrete Zufallsvariable X mit der Verteilung $\{x_i, P(X=x_i)\}_{i=1,2,\dots}$ ist der Erwartungswert definiert durch:

$$E(X) := \sum_i x_i P(X=x_i)$$

Für eine stetige (kontinuierliche) Zufallsvariable X mit der Dichte f ist der Erwartungswert definiert durch:

$$E(X) := \int_{-\infty}^{+\infty} x f(x) dx$$

Offenbar ist die Integralformel die natürliche Verallgemeinerung der Summenformel für den Fall überabzählbar unendlich vieler Variablen-Ausprägungen. Aber mathematisch weniger interessierte Leser brauchen sich keine Sorgen zu machen: Die Integralrechnung ist für das Verständnis dieses Kurses irrelevant.

Varianz

Die Varianz einer Zufallsvariablen X ist definiert durch:

$$\text{Var}(X) := E[(X - E(X))^2]$$

Aus der Definition folgt durch Ausmultiplizieren und wegen der Linearität des Erwartungswertes (d.h.: $E(aX) = aE(X)$, $a \in \mathbb{R}$; $E(X+Y) = E(X) + E(Y)$):

$$\text{Var}(X) = E(X^2) - E(X)^2,$$

so daß bei $E(X) = 0$ gilt: $\text{Var}(X) = E(X^2)$.

Kovarianz

Die Kovarianz der beiden Zufallsvariablen X und Y ist definiert durch:

$$\text{Cov}(X,Y) := E[(X - E(X))(Y - E(Y))]$$

Aus der Definition folgt (mit denselben Argumenten wie bei der Varianz):

$$\text{Cov}(X,Y) = E(XY) - E(X)E(Y),$$

so daß bei $E(X) = E(Y) = 0$ gilt: $\text{Cov}(X,Y) = E(XY)$.

Durch Einsetzen von X für Y sieht man außerdem sofort, daß die Varianz der Zufallsvariablen X gerade die Kovarianz von X mit sich selber ist.

Die Kovarianzmatrix eines Zufallsvektors

Ist $X := (X_1, X_2, \dots, X_q)^T$, $q \in \mathbb{N}$ ein Spaltenvektor aus Zufallsvariablen, so wird die Kovarianzmatrix $\text{Cov}(X)$ von X wie folgt definiert:

$$\text{Cov}(X) := \begin{bmatrix} \text{Var}(X_1) & & & & \\ \text{Cov}(X_2, X_1) & \text{Var}(X_2) & & & \\ \text{Cov}(X_3, X_1) & \text{Cov}(X_3, X_2) & \text{Var}(X_3) & & \\ \cdot & \cdot & \cdot & \cdot & \\ \text{Cov}(X_q, X_1) & \text{Cov}(X_q, X_2) & \text{Cov}(X_q, X_3) & \dots & \text{Var}(X_q) \end{bmatrix}$$

Die obere Dreiecksmatrix ist dabei aus Symmetriegründen weggelassen.

Korrelation

Die Korrelation der beiden Zufallsvariablen X und Y ist definiert durch:

$$\text{Cor}(X,Y) := \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$$

Bei standardisierten Variablen (d.h. $\text{Var}(X) = \text{Var}(Y) = 1$) gilt also: $\text{Cor}(X,Y) = \text{Cov}(X,Y)$. Die Korrelationsmatrix eines Zufallsvektors ist analog zur Kovarianzmatrix definiert.

Kovarianzen von Linearkombinationen aus Zufallsvariablen

Die Beispiele aus der Einleitung haben gezeigt, daß in einer Strukturgleichung eine abhängige Variable als Linearkombination von unabhängigen Variablen dargestellt wird. Bei der Analyse von Strukturgleichungsmodellen (mit oder ohne latente Variablen) spielt die Berechnung der Varianzen und Kovarianzen von Linearkombinationen aus den Varianzen und Kovarianzen der Ausgangsvariablen eine entscheidende Rolle.

Für die Kovarianz von $X := a_1X_1 + a_2X_2$ und $Y := b_1Y_1 + b_2Y_2$ gilt:

$$\text{Cov}(X,Y) = a_1b_1\text{Cov}(X_1,Y_1) + a_1b_2\text{Cov}(X_1,Y_2) + a_2b_1\text{Cov}(X_2,Y_1) + a_2b_2\text{Cov}(X_2,Y_2) \quad (2.1)$$

Damit gilt speziell für die Varianz von X :

$$\text{Var}(X) = a_1^2\text{Var}(X_1) + 2a_1a_2\text{Cov}(X_1,X_2) + a_2^2\text{Var}(X_2)$$

Für unkorrelierte Ausgangsvariablen X_1 und X_2 folgt: $\text{Var}(X) = a_1^2\text{Var}(X_1) + a_2^2\text{Var}(X_2)$.

Dichte der multivariaten Normalverteilung

Ist die gemeinsame Verteilung der Zufallsvariablen $X := (X_1, X_2, \dots, X_q)^T$, $q \in \mathbb{N}$, multivariat normal, so ist sie perfekt durch den Vektor $\mu := (E(X_1), E(X_2), \dots, E(X_q))^T$ der Erwartungswerte und die Kovarianzmatrix $\Sigma := \text{Cov}(X)$ festgelegt, wie folgende Funktionsvorschrift für die Dichte der q -dimensionalen Normalverteilung zeigt ($|\Sigma|$ = Determinante von Σ):

$$f_{\mu,\Sigma}(x) := \frac{1}{\sqrt{(2\pi)^q |\Sigma|}} \exp\left[-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right], \quad x \in \mathbb{R}^q$$

Übungen:

- i) Beweisen Sie Formel (2.1) unter Verwendung der Linearität des Erwartungswertes (s.o., Def. der Varianz).
- ii) Es gelte $\text{Var}(X_1) = 1$, $\text{Var}(X_2) = 2$, $\text{Var}(X_3) = 3$, $\text{Cov}(X_1, X_2) = 0.5$, $\text{Cov}(X_1, X_3) = \text{Cov}(X_2, X_3) = 0$. Berechnen Sie folgende Varianzen bzw. Kovarianzen:

- Var($2X_1$)
- Var($2X_1 + X_3$)
- Var($3X_1 + 2X_2$)
- Cov($4X_1 + X_3, X_2 + X_3$)

Eine ausführliche Behandlung der Regeln für das Rechnen mit Kovarianzen findet sich z.B. in Kenny (1979, Kap. 2).

Mit den Vereinbarungen $X := (X_1, X_2, \dots, X_q)^T$, $a := (a_1, a_2, \dots, a_q) \in \mathbb{R}^q$ können wir die Linearkombination $a_1X_1 + a_2X_2 + \dots + a_qX_q$ bequem schreiben als: Ax , wobei a ein Zeilenvektor aus Zahlen und X ein Spaltenvektor aus Zufallsvariablen ist. Ordnet man die Koeffizientenvektoren von k Linearkombinationen untereinander zu einer $(k \times q)$ -Matrix A an, so läßt sich der Spaltenvektor der k Linearkombinationen schreiben als: AX . Durch Verallgemeinerung der Formel (2.1) ergibt sich die folgende, oft benötigte Formel für die Kovarianzmatrix $\text{Cov}(AX)$ der Linearkombinationen AX :

$$\text{Cov}(AX) = A\text{Cov}(X)A^T$$

Im allgemeinen LISREL-Modell können acht verschiedene Parametermatrizen in der Rolle von A auftreten und beim Arbeiten mit der LISREL-Methodologie müssen daher oft Kovarianzmatrizen des Typs $\text{Cov}(AX)$ berechnet werden.

Wir haben im aktuellen Abschnitt 2.1 bislang über Populationsparameter gesprochen. In einer konkreten Forschungssituation sind natürlich nur Stichprobenschätzer für die Parameter der manifesten Variablen verfügbar. Als Schätzer für den Erwartungswert der Variablen X verwendet man das Stichprobenmittel:

$$\bar{x} := \frac{1}{N} \sum_{i=1}^N x_i$$

Die Kovarianz der Variablen X und Y kann erwartungstreu geschätzt werden durch:

$$s_{XY} := \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})$$

Alle oben vorgestellten Rechenregeln für Kovarianzen gelten erfreulicherweise uneingeschränkt auch für die Stichproben-Schätzer.

Abschließend sollen noch zwei LISREL-Bezeichnungskonventionen eingeführt werden. Für die Kovarianzmatrix des Zufallsvektors aus allen manifesten Variablen schreiben wir " Σ ". Wie wir bereits wissen, besteht das Ziel einer Kovarianzstrukturanalyse darin, eine möglichst gute Erklärung für diese Matrix Σ zu liefern. Bei der Suche nach einem guten Modell spielt natürlich die allein verfügbare Matrix mit den Stichproben-Schätzern für die Varianzen und Kovarianzen der manifesten Variablen eine entscheidende Rolle. Von dieser zufallsabhängigen Matrix, die mit " S " bezeichnet wird, hängen alle Parameterschätzungen und alle Hypothesentests ab. Daher sollte diese Matrix vor Beginn der Analyse sehr sorgfältig auf mögliche Artefakte, z.B. durch Ausreißer, geprüft werden (vgl. Bollen, 1989, S. 24ff).

2.2 Erklärung einer Kovarianzmatrix durch ein Strukturgleichungsmodell

Mit den Ergebnissen aus Abschnitt 2.1 kann die Formulierung aus der Einleitung, daß bei einer Faktorenanalyse die Kovarianzen (oder Korrelationen) zwischen q beobachteten Variablen durch deren Abhängigkeit von n ($< q$) latenten Variablen und durch die Korrelationen (oder Kovarianzen) zwischen den latenten Variablen erklärt würden, vollständig nachvollzogen werden. Für die Varianzen und Kovarianzen

der manifesten Variablen in Beispiel 1.2 folgt mit obigen Rechenregeln aus den Gleichungen und sonstigen Annahmen des Modells (Bezeichnungen gemäß Abschnitt 1: $\phi_{ii} = \text{Var}(\xi_i)$, $\phi_{ij} = \text{Cov}(\xi_i, \xi_j)$):

$$\begin{aligned}\text{Var}(X_1) &= \lambda_{11}^2 \phi_{11} + 2\lambda_{11} \text{Cov}(\xi_1, \delta_1) + \text{Var}(\delta_1) \\ &= \lambda_{11}^2 \phi_{11} + \text{Var}(\delta_1)\end{aligned}$$

$$\text{Var}(X_2) = \lambda_{21}^2 \phi_{11} + \text{Var}(\delta_2)$$

$$\text{Var}(X_3) = \lambda_{32}^2 \phi_{22} + \text{Var}(\delta_3)$$

$$\text{Var}(X_4) = \lambda_{42}^2 \phi_{22} + \text{Var}(\delta_4)$$

$$\begin{aligned}\text{Cov}(X_1, X_2) &= \text{Cov}(\lambda_{11} \xi_1 + \delta_1, \lambda_{21} \xi_1 + \delta_2) \\ &= \lambda_{11} \lambda_{21} \phi_{11} + \lambda_{11} \text{Cov}(\xi_1, \delta_2) + \lambda_{21} \text{Cov}(\delta_1, \xi_1) + \text{Cov}(\delta_1, \delta_2) \\ &= \lambda_{11} \lambda_{21} \phi_{11}\end{aligned}$$

$$\begin{aligned}\text{Cov}(X_1, X_3) &= \text{Cov}(\lambda_{11} \xi_1 + \delta_1, \lambda_{32} \xi_2 + \delta_3) \\ &= \lambda_{11} \lambda_{32} \phi_{21}\end{aligned}$$

Übung: Berechnen Sie die übrigen Kovarianzen.

Nun müssen noch die Gleichheits-Restriktionen:

$$\begin{aligned}\lambda_{11} &= \lambda_{21} = \lambda_{32} = \lambda_{42} \\ \text{Var}(\delta_1) &= \text{Var}(\delta_2) = \text{Var}(\delta_3) = \text{Var}(\delta_4)\end{aligned}$$

berücksichtigt werden.

Außerdem ist zu beachten, daß die Maßeinheiten (aus statistischer Sicht: die Varianzen) der beiden latenten Variablen beliebig festgelegt werden können und auch festgelegt werden müssen, damit die Ladungen interpretiert werden können. Betrachten wir die Bestimmungsgleichung für X_1 :

$$X_1 = \lambda_{11} \xi_1 + \delta_1$$

Durch diese Gleichung ist die Varianz von ξ_1 nicht festgelegt, denn wir können X_1 äquivalent durch die latente Variable $\xi_1^* := a\xi_1$ mit der Varianz $a^2 \text{Var}(\xi_1)$ erklären:

$$X_1 = \lambda_{11}^* \xi_1^* + \delta_1, \quad \text{mit } \lambda_{11}^* := \frac{1}{a} \lambda_{11}$$

Diese Beliebigkeit bei der Wahl von a , d.h. bei der Festlegung der Maßeinheit bzw. der Varianz von ξ_1^* , läßt sich durch Fixieren der Ladung λ_{11} (z.B. auf den Wert Eins) beseitigen. Man kann die Varianzen aller latenter Variablen eines Modells festlegen, indem man jeweils die Ladung des besten Indikators auf Eins fixiert. Damit wird **keine** empirische Behauptung aufgestellt, sondern lediglich eine Unbestimmtheit aus dem Modell beseitigt. Bei einer anderen Festlegung resultiert ein Modell, das zwar numerisch andere Parameter besitzt, aber exakt dieselbe Kovarianzmatrix impliziert. Daher wählt man den bequemen Wert Eins. Wir werden uns mit Problemen dieser Art unter dem Stichwort "Identifikation" noch öfter beschäftigen (vgl. Abschnitt 2.3).

In unserem Fall sind durch die Fixierung von λ_{11} und λ_{32} auf 1 wegen der ersten Gleichheits-Restriktion alle Ladungen auf Eins festgelegt.

Mit der Bezeichnungsvereinbarung:

$$\vartheta := \text{Var}(\delta_1) = \text{Var}(\delta_2) = \text{Var}(\delta_3) = \text{Var}(\delta_4)$$

impliziert unser faktorenanalytisches Modell folgende Kovarianzmatrix der manifesten Variablen:

$$\Sigma = \begin{bmatrix} \phi_{11} + \vartheta & & & \\ \phi_{11} & \phi_{11} + \vartheta & & \\ \phi_{21} & \phi_{21} & \phi_{22} + \vartheta & \\ \phi_{21} & \phi_{21} & \phi_{22} & \phi_{22} + \vartheta \end{bmatrix} \quad (2.2)$$

Man kann die Kovarianzmatrix Σ der manifesten Variablen als **empirische Gesetzmäßigkeit** auffassen, die durch die Entwicklung eines LISREL-Modells erklärt werden soll. Dieses Vorgehen entspricht dem Grundschemata des wissenschaftlichen Erklärens, wobei nach Überzeugung der Wissenschaftsphilosophie eine Erklärung mit Hilfe theoretischer Begriffe (latenter Variablen) besonders erfolgversprechend ist.

Die Darstellung der Kovarianzmatrix der manifesten Variablen als Funktion der Modellparameter ist ein wesentliches Merkmal der Kovarianzstrukturanalyse mit LISREL, unabhängig davon, ob das Modell latente Variablen enthält (wie in Beispiel 1.2) oder nicht (wie in Beispiel 1.1). Neben den Kausal- bzw. Strukturkoeffizienten werden für diese Darstellung auch alle Varianzen und Kovarianzen der Quellvariablen (exogenen Variablen und Residualvariablen) benötigt. Der Vektor aller Modellparameter in diesem Sinn wird meist mit " θ " bezeichnet. In Beispiel 1.2 haben wir nach Auswertung der Strukturgleichungen, Restriktionen und Identifikationsüberlegungen folgenden Vektor erhalten:

$$\theta = (\phi_{11}, \phi_{21}, \phi_{22}, \vartheta)$$

Er enthält alle freien Parameter sowie einen Repräsentanten für jede Gruppe gleich gesetzte Parameter.

Die Darstellung der Kovarianzmatrix Σ als Funktion dieses Parametervektors haben wir bereits in Gleichung (2.2) formuliert. Sie ist von der allgemeinen Gestalt:

$$\Sigma = f(\theta) \quad (2.3)$$

Eine solche Darstellung werden wir später gelegentlich als die "Fundamental-Hypothese" des jeweiligen Modells bezeichnet.

In folgender Tabelle ist die von Steyer (1988, S. 312) mitgeteilte empirische Kovarianzmatrix aus einer Stichprobe von $N = 179$ Studenten wiedergegeben, die wir für unsere bald anstehenden konkreten Analyseschritte benötigen:

	X_1	X_2	X_3	X_4
X_1	24.670			
X_2	21.895	25.135		
X_3	10.353	10.624	27.239	
X_4	11.665	12.636	25.258	28.683

Nun wäre es eigentlich an der Zeit, die Stichproben-Kovarianzmatrix von Kerchoff (1974) zu präsentieren, an der sich unsere Theorie bewähren soll. Aber mit Entsetzen stellen wir fest, daß Kerchoff lediglich die Korrelationsmatrix publiziert hat. Zum Glück müssen wir unsere schöne LISREL-Methodologie jetzt nicht einpacken, jedoch muß die Analyse einer Korrelationsmatrix an Stelle der aus statistischen Gründen im allgemeinen erforderlichen Kovarianzmatrix speziell gerechtfertigt werden. Die entsprechenden Argumente sind relativ kompliziert, so daß hier nur auf das Original-Handbuch zu LISREL 7 (Jöreskog & Sörbom 1989, Abschnitt 1.21) sowie auf die Arbeit von Lee (1985) verwiesen werden kann. In Abschnitt 4.3 werden wir an einem einfacheren Beispiel hinreichende Bedingungen für die Analyse einer Korrelationsmatrix kennenlernen.

Folgende Tabelle enthält die von Kerchoff (1974) mitgeteilten Korrelationen:

	INTELLNZ	BLDNGVAT	BERUFVAT	SCHULLST	AMBITION
INTELLNZ	1				
BLDNGVAT	.277	1			
BERUFVAT	.250	.611	1		
SCHULLST	.572	.294	.248	1	
AMBITION	.335	.303	.331	.478	1

2.3 Identifikation der Parameter

Wir haben im letzten Abschnitt gesehen, wie die Kovarianzmatrix Σ der manifesten Variablen als Funktion des Parametervektors θ ausgedrückt werden kann. Nun sind allerdings die Parameter unbekannt und wir möchten Informationen über sie aus den Daten gewinnen. Andererseits können wir die Varianzen und Kovarianzen der manifesten Variablen immerhin anhand der Stichprobe schätzen. Wenn es gelingt, die Gleichung

$$\Sigma = f(\theta)$$

nach den unbekanntem Parametern aufzulösen, kann man Σ durch die Matrix S der Stichprobenstatistiken ersetzen und Schätzer für die gesuchten Parameter berechnen.

Ein Parameter θ_v heißt "**identifiziert**", wenn für ihn das Auflösen gelingt, d.h. wenn er auf mindestens eine Weise als Funktion g_v der Elemente von Σ dargestellt werden kann:

$$\theta_v = g_v(\Sigma)$$

Ein Modell heißt "**identifiziert**", wenn alle Parameter identifiziert sind.

Es gibt leider keine praktikable notwendige **und** hinreichende Bedingung für die Identifikation eines **beliebigen** Strukturgleichungsmodelles. Allerdings sind für einige spezielle Modelle nützliche Identifikationsregeln bekannt (vgl. Bollen 1989), z.B. sind rekursive Modelle (wie unser Beispiel 1.1) stets identifiziert.

Im allgemeinen muß zum Nachweis der Identifikation eines konkreten Modells für jeden Parameter seine Darstellbarkeit durch die Varianzen und Kovarianzen der manifesten Variablen gezeigt werden. Die entscheidenden Hilfsmittel bei diesem Nachweis sind die in Abschnitt 2.1 eingeführten Rechenregeln für Kovarianzen.

Nach diesen etwas abstrakten Ausführungen soll nun die Identifikation der Parameter in Beispiel 1.2 nachgewiesen werden. Wir müssen das Gleichungssystem:

$$\begin{bmatrix} \text{Var}(X_1) \\ \text{Cov}(X_2, X_1) & \text{Var}(X_2) \\ \text{Cov}(X_3, X_1) & \text{Cov}(X_3, X_2) & \text{Var}(X_3) \\ \text{Cov}(X_4, X_1) & \text{Cov}(X_4, X_2) & \text{Cov}(X_4, X_3) & \text{Var}(X_4) \end{bmatrix} = \begin{bmatrix} \phi_{11} + \vartheta & & & \\ \phi_{11} & \phi_{11} + \vartheta & & \\ \phi_{21} & \phi_{21} & \phi_{22} + \vartheta & \\ \phi_{21} & \phi_{21} & \phi_{22} & \phi_{22} + \vartheta \end{bmatrix}$$

nach den Parametern in $\theta = (\phi_{11}, \phi_{21}, \phi_{22}, \vartheta)$ auflösen. Wir erhalten mühelos das Ergebnis:

$$\begin{aligned} \phi_{11} &= \text{Cov}(X_2, X_1) \\ \phi_{21} &= \text{Cov}(X_3, X_1) \\ \phi_{22} &= \text{Cov}(X_4, X_3) \\ \vartheta &= \text{Var}(X_1) - \text{Cov}(X_2, X_1) \end{aligned}$$

Damit steht also einer Schätzung der Parameter prinzipiell nichts mehr im Wege.

Bei genauerem Hinsehen fällt auf, daß unser System sogar noch sechs weitere Bestimmungsgleichungen für Parameter liefert:

$$\begin{aligned} \phi_{21} &= \text{Cov}(X_3, X_2) \\ \phi_{21} &= \text{Cov}(X_4, X_1) \\ \phi_{21} &= \text{Cov}(X_4, X_2) \\ \vartheta &= \text{Var}(X_2) - \text{Cov}(X_2, X_1) \\ \vartheta &= \text{Var}(X_3) - \text{Cov}(X_4, X_3) \\ \vartheta &= \text{Var}(X_4) - \text{Cov}(X_4, X_3) \end{aligned}$$

Für ϕ_{21} und ϑ haben wir jeweils mehrere Bestimmungsgleichungen zur Verfügung. Parameter mit dieser Eigenschaft werden als "**überidentifiziert**" oder "**überbestimmt**" bezeichnet. Ein Modell wird so genannt, wenn es identifiziert ist und mindestens einen überidentifizierten Parameter enthält.

Bei einem überidentifizierten Modell ist die Anzahl t der Parameter kleiner als die Anzahl nicht-redundanter Elemente in der zu erklärenden Kovarianzmatrix und damit genügen diese Modelle dem wissenschaftstheoretischen Gebot der Sparsamkeit. Die Überbestimmtheit hat zwei Konsequenzen, die in späteren Abschnitten noch genauer beleuchtet werden:

- Verschiedene Bestimmungsgleichungen für einen Parameter liefern zufallsbedingt im allgemeinen auch dann unterschiedliche Ergebnisse, wenn das Strukturgleichungsmodell stimmt. Zu eindeutigen Schätzern gelangt man durch folgendes Minimierungsprinzip: Man wählt den Vektor $\hat{\theta}$ der geschätzten Parameter so, daß der "Abstand" zwischen der Stichprobenkovarianzmatrix S und der durch das Modell postulierten Kovarianzmatrix $f(\hat{\theta})$ möglichst klein wird. Der Abstands begriff wird in Abschnitt 2.4 über Parameterschätzung präzisiert.
- Das Strukturgleichungsmodell impliziert für die Kovarianzmatrix der manifesten Variablen Restriktionen, die empirisch falsch sein können. Unter der Annahme multivariater Normalverteilung der manifesten Variablen erlaubt dies eine inferenzstatistische Beurteilung des Modells. Damit genügen überbestimmte Modelle der wissenschaftsphilosophischen Forderung nach Falsifizierbarkeit.

Ein Strukturgleichungsmodell heißt "**genau identifiziert**" oder "**saturiert**", wenn alle Parameter identifiziert sind und kein Parameter überidentifiziert ist, d.h. wenn **genau ein** Lösungsvektor θ für die Gleichung (2.3) existiert. Für den Schätzer $\hat{\theta}$ gilt dann:

$$S = f(\hat{\theta}),$$

d.h. die Stichprobenkovarianzmatrix S kann exakt reproduziert werden. Dies ist häufig der Fall, wenn die Anzahl t der Parameter in θ identisch ist mit der Anzahl nicht-redundanter Elemente in S . Letztere Anzahl beträgt bei k manifesten Variablen $\frac{1}{2}k(k+1)$. Genau identifizierte Modelle sind nicht prüfbar. Sofern sie stimmen, können sie aber trotzdem sinnvoll sein, weil sie immerhin die Schätzung und (bei multivariater Normalverteilung der manifesten Variablen) die Testung der einzelnen Parameter erlauben.

Eine notwendige (aber nicht hinreichende!) Bedingung für die Identifikation eines Modelles lautet

$$t \leq \frac{1}{2}k(k+1) \quad (2.4)$$

Bei einem nicht-identifizierten ("unterbestimmten") Modell existieren unendlich viele verschiedene Vektoren $\theta^* \in \mathbb{R}^t$, welche dieselbe Kovarianzmatrix $f(\theta^*)$ liefern und damit natürlich auch denselben Abstand zu einer Stichprobenkovarianzmatrix S . Bei der Parameterschätzung nach der oben erwähnten Abstand-Minimierungsmethode muß daher die iterative Suche nach dem globalen Minimum scheitern. Folglich können die nicht-identifizierten Parameter nicht geschätzt werden.

Wir haben in Abschnitt 2.2 bereits ein Beispiel für ein nicht-identifiziertes Modell kennengelernt: Nimmt man in das Latent-State-Modell der Angst einen freien Parameter λ für die (als gleich angenommenen) Ladungen auf anstatt der Fixierung auf Eins, so sind λ , ϕ_{11} , ϕ_{21} , ϕ_{22} nicht identifiziert, obwohl fünf Parametern zehn Varianzen bzw. Kovarianzen gegenüberstehen, denn die Parametervektoren

$$(\lambda, \phi_{11}, \phi_{21}, \phi_{22}, \vartheta) \text{ und } (\alpha\lambda, \frac{1}{\alpha^2}\phi_{11}, \frac{1}{\alpha^2}\phi_{21}, \frac{1}{\alpha^2}\phi_{22}, \vartheta), \quad \alpha > 0$$

liefern dieselbe Kovarianzmatrix. Der Parameter ϑ ist hingegen nach wie vor überbestimmt. Dieses Beispiel zeigt, daß ein Modell sowohl identifizierte als auch nicht-identifizierte Parameter enthalten kann und daß die Ungleichung (2.4) keine Gewähr für Identifikation bietet.

LISREL entdeckt in vielen Fällen Identifikationsprobleme und gibt entsprechende Warnungen aus. Es kann jedoch auch passieren, daß LISREL den Defekt nicht bemerkt und sinnlose Schätzer (z.B. negative Varianzen) berechnet. Es ist daher ratsam, die Identifikation eines Modells vor dem Einsatz von LISREL nachzuprüfen. In günstigen Fällen kann man dazu eine der Identifikationsregeln aus Bollen (1989) verwenden. Ist dies nicht möglich, so kontrolliere man zunächst die notwendige Bedingung (2.4). Ist sie erfüllt, versuche man, mit Hilfe der Rechenregeln für Kovarianzen das Gleichungssystem (2.3) nach den Parametern aufzulösen.

2.4 Schätzung der Parameter durch Minimierung einer Fit-Funktion

Bei einer Kovarianzstrukturanalyse möchte man einerseits die Daten erklären und andererseits anhand der Daten die unbekannteren Parameter schätzen. Natürlich ist das Schätzen der einzelnen Parameter nur sinnvoll, wenn die Erklärung der Daten durch das Modell korrekt ist. Folglich sollten als nächstes mögliche Modelltests besprochen werden. Allerdings basieren die Modelltests auf dem Ergebnis der Parameterschätzung unter Annahme der Modellgültigkeit. Das muß so sein, weil das Modell im jetzigen Stadium noch sehr unfertig ist: Es enthält viele unspezifizierte Parameter. Daher werden wir uns zunächst

mit der Parameterschätzung beschäftigen und erst im nächsten Abschnitt Methoden zur Prüfung der Gültigkeit des Modells kennenlernen.

Das Schätzen der Parameter setzt, wie oben besprochen, die Identifikation des Modells voraus. Dann gilt für je zwei verschiedene Parametervektoren θ und θ' , daß sie verschiedene Kovarianzmatrizen implizieren, d.h. $f(\theta) \neq f(\theta')$. Daher kann ein Schätzer $\hat{\theta}$ nach folgendem Prinzip ermittelt werden:

Bestimme $\hat{\theta}$ so, daß der Abstand zwischen der Stichprobenkovarianzmatrix S und der von $\hat{\theta}$ implizierten Kovarianzmatrix $f(\hat{\theta})$ minimal wird.

Weil bei den folgenden Darstellungen mehrere, verschieden indizierte Thetas auftreten, soll deren Bedeutung vorab klargestellt werden:

- θ Der gesuchte Vektor der wahren Parameter. Dies ist ein festes Element aus dem \mathbb{R}^l .
- $\hat{\theta}$ Stichproben-Schätzer für θ
- θ^* Beliebiges Element aus der Menge aller potentiellen Parametervektoren. (Dies ist eine geeignet eingeschränkte Teilmenge des \mathbb{R}^l . θ^* ist also ein Platzhalter, der beliebige Werte aus einer sehr großen Menge von Vektoren annehmen kann, während θ ein Vektor aus festen (aber unbekannt) Zahlen ist.

Für ein beliebiges θ^* soll " $f(\theta^*)$ " im folgenden durch " Σ^* " abgekürzt werden.

Aus dem Grundprinzip wurden durch Definition einer speziellen **Fitfunktion** $F(S, \Sigma^*)$ zur Messung des Abstands zwischen S und Σ^* mehrere konkrete Schätzverfahren entwickelt. F ist dabei eine nicht-negative Funktion der Komponenten von θ^* , die nur bei perfektem Fit ($\Sigma^* = S$) den Wert 0 annimmt. Sie ist im allgemeinen nicht-linear und so kompliziert, daß ihre Minimierung ein iteratives Verfahren erfordert. Dabei müht sich der Computer schrittweise durch einen oft sehr hochdimensionalen Parameter-Raum und sucht nach dem tiefsten Tal in der Funktionsoberfläche.

Die auf der Siemens-Anlage derzeit installierte Version LISREL VI bietet folgende Schätzverfahren an:

2.4.1 Unweighted Least Squares (ULS)

Hier wird folgende Fitfunktion minimiert:

$$F_{\text{ULS}}(S, \Sigma^*) := \frac{1}{2} \text{tr}[(S - \Sigma^*)^T (S - \Sigma^*)]$$

Mit "tr" wird dabei eine Funktion bezeichnet, die für eine quadratische Matrix die Summe ihrer Hauptdiagonalelemente liefert. $F_{\text{ULS}}(S, \Sigma^*)$ ist nichts anderes als die Summe aller quadrierten Differenzen zwischen korrespondierenden Elementen aus S und Σ^* . $F_{\text{ULS}}(S, \Sigma^*)$ wird offensichtlich umso kleiner, je ähnlicher sich S und Σ^* werden.

Beispiel: Bei

$$S = \begin{bmatrix} 1.1 & 0.6 \\ 0.6 & 2.2 \end{bmatrix}, \quad \Sigma^* = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 2 \end{bmatrix}$$

$$\Rightarrow S - \Sigma^* = \begin{bmatrix} 0.1 & 0.1 \\ 0.1 & 0.2 \end{bmatrix}$$

ergibt sich $F_{\text{ULS}}(S, \Sigma^*) = \frac{1}{2}(0.1^2 + 0.1^2 + 0.1^2 + 0.2^2) = 0.035$.

Für den bei Minimierung von $F_{\text{ULS}}(S, \Sigma^*)$ resultierenden Schätzer spricht, daß er **ohne Voraussetzungen über die Verteilung** der manifesten Variablen **konsistent** ist, d.h. ($N = \text{Stichprobenumfang}$):

$$\lim_{N \rightarrow \infty} P(|\hat{\theta}_N - \theta| > \varepsilon) = 0 \quad \forall \varepsilon > 0$$

Der Parameter θ wird also asymptotisch richtig geschätzt. Die Konsistenz-Eigenschaft besitzen auch alle anderen von LISREL berechneten Schätzer.

Sind die manifesten Variablen **multivariat normalverteilt**, dann erlaubt die ULS-Methode auch die Berechnung von Standardfehlern zu den Parameterschätzungen (vgl. Abschnitt 2.4.3) und einen globalen Modelltest (vgl. Abschnitt 2.5.2). Allerdings sind unter dieser Voraussetzung die anschließend vorzustellenden Schätzmethode ML und GLS überlegen.

Nachteile des ULS-Schätzers:

- Unter der klassischen Annahme multivariater Normalität der gemeinsamen Verteilung aller manifesten Variablen ist die ULS-Schätzmethode der GLS- und der ML-Schätzmethode unterlegen (vgl. Abschnitt 2.4.3).
- Die Fitfunktion F_{ULS} ist nicht skaleninvariant.

Zur Erläuterung dieser Aussage sind zwei Definitionen erforderlich (vgl. Jöreskog & Sörbom 1989, Abschnitt 1.20):

Ein **Strukturgleichungsmodell** mit der Fundamental-Hypothese $\Sigma = f(\Theta)$ heißt **skaleninvariant**, wenn für jede Diagonalmatrix D mit positiven Einträgen und für jeden Parametervektor Θ ein anderer Parametervektor $\tilde{\Theta}$ existiert mit:

$$f(\tilde{\Theta}) = Df(\Theta)D$$

Die Matrix D enthält Skalenfaktoren für die manifesten Variablen; sie bewirkt also Wechsel der Maßeinheiten (z.B. Gewichtsmessung in Gramm statt Kilogramm). Trägt man in die Hauptdiagonale von D die Kehrwerte der Standardabweichungen der manifesten Variablen ein, so ist DSD eine Korrelationsmatrix.

Eine **Fitfunktion** $F(S, \Sigma^*)$ heißt **skaleninvariant**, wenn für jede Diagonalmatrix D mit positiven Einträgen gilt:

$$F(DSD, D\Sigma^*D) = F(S, \Sigma^*)$$

Eine skaleninvariante Fitfunktion verhält sich "vernünftig", wenn bei der Analyse eines skaleninvarianten Modells Maßeinheiten für manifeste Variablen verändert werden, d.h. wenn statt S die Matrix DSD verwendet wird. In diesem Fall implizieren die neuen Parameter-Schätzungen in $\hat{\theta}_D$ die Kovarianzmatrix $Df(\hat{\theta})D$, die in "natürlicher" Relation zu der ursprünglich implizierten Kovarianzmatrix $f(\hat{\theta})$ steht. Damit ergeben sich auch die neuen Parameter-Schätzungen in $\hat{\theta}_D$ selber durch Berücksichtigung der Maßstabsänderungen in D unmittelbar aus den alten Schätzungen in $\hat{\theta}$ (Jöreskog & Sörbom 1989).

Die Fitfunktion F_{ULS} ist nicht skaleninvariant und daher können durch Skalenfaktoren bei den ULS-Schätzungen Veränderungen auftreten, die nicht verträglich sind mit den Maßstabsänderungen. Weil der Maßstab sozialwissenschaftlicher Indikatoren meist willkürlich ist, wird daher oft empfohlen, ULS-Schätzer aus der Korrelationsmatrix statt aus der Kovarianzmatrix zu bestimmen (vgl. Long, 1983).

2.4.2 Generalized Least Squares (GLS)

Beim GLS-Verfahren wird die Matrix $(S - \Sigma^*)$ der Differenzen mit S^{-1} (Inverse von S) gewichtet, bevor wie in 2.4.1 die Summe der quadrierten Elemente berechnet wird. Dies führt zu der Fitfunktion ($I =$ Einheitsmatrix):

$$\begin{aligned} F_{\text{GLS}}(S, \Sigma^*) &:= \frac{1}{2} \text{tr}[(S^{-1}(S - \Sigma^*))^T S^{-1}(S - \Sigma^*)] \\ &= \frac{1}{2} \text{tr}[(I - S^{-1}\Sigma^*)^T (I - S^{-1}\Sigma^*)] \end{aligned}$$

Auch $F_{\text{GLS}}(S, \Sigma^*)$ wird offensichtlich umso kleiner, je näher Z^* an S heranrückt. Eine technische Voraussetzung der GLS-Methode ist die Invertierbarkeit von S .

Der GLS-Schätzer ist bei multivariater Normalverteilung der manifesten Variablen asymptotisch äquivalent mit dem nun vorzustellenden ML-Schätzer und benimmt sich in Simulationsstudien auch bei nicht-normalen Verhältnissen sehr ähnlich. Folglich gelten für ihn dieselben Verteilungs-Voraussetzungen, Vor- und Nachteile. Da für die Berechnung von $F_{\text{GLS}}(S, \Sigma^*)$ nur die Matrix S regulär sein muß, bei der Berechnung der ML-Fitfunktion $F_{\text{ML}}(S, \Sigma^*)$ aber die Regularität von S und von Σ^* benötigt wird (s.u.), kann man gelegentlich mit Erfolg auf die GLS-Methode ausweichen, wenn das ML-Verfahren wegen ungünstiger Startwerte scheitert.

2.4.3 Maximum Likelihood (ML)

Die ML-Methode wird anders begründet als die Kleinst-Quadrat-Verfahren. Man setzt zunächst voraus, daß die k manifesten Variablen **multivariat normalverteilt** seien. Unter dieser Voraussetzung hat Wishart die Wahrscheinlichkeitsdichte $W(S; \Sigma, N-1)$ für eine Stichprobenkovarianzmatrix S bei unabhängiger Ziehung von N Fällen aus einer Population mit Kovarianzmatrix Σ hergeleitet. Das Maß für die Ähnlichkeit von S und Σ^* bzw. für den Fit eines potentiellen Parametervektors θ^* basiert auf dem Quotienten aus der Wahrscheinlichkeitsdichte von S bei Annahme von $\Sigma = \Sigma^*$ und der Wahrscheinlichkeitsdichte von S bei Annahme von $\Sigma = S$:

$$\log\left(\frac{W(S; \Sigma^*, N-1)}{W(S; S, N-1)}\right) = -\frac{1}{2}(N-1)[\text{tr}(S\Sigma^{*-1}) + \log|\Sigma^*| - \log|S| - k]$$

Dabei wird vorausgesetzt, daß die Determinanten $|S|$ und $|\Sigma^*|$ positiv sind. Der Likelihood-Quotient zeigt an, inwiefern die Likelihood von S unter Annahme von Σ^* schlechter ist als die maximal mögliche Likelihood und kann folglich zwischen 0 und 1 variieren. Der logarithmierte Quotient nimmt daher Werte von $-\infty$ bis 0 an und erreicht seinen maximalen Wert 0 genau dann, wenn $\Sigma^* = S$. Statt diesen Ausdruck zu maximieren kann man wegen des negativen Vorzeichens auch den Ausdruck in eckigen Klammern minimieren. Dafür haben sich die Autoren von LISREL entschieden und folgende Fitfunktion gewählt:

$$F_{\text{ML}}(S; \Sigma^*) = \text{tr}(S\Sigma^{*-1}) + \log|\Sigma^*| - \log|S| - k$$

$F_{\text{ML}}(S, \Sigma^*)$ wird genau dann minimal, wenn der Zähler des Likelihood-Quotienten die maximale Likelihood der beobachteten Kovarianzmatrix S bei Gültigkeit des behaupteten Strukturgleichungsmodells, d.h. unter den Restriktionen, die das Modell für die Kovarianzmatrix der manifesten Variablen impliziert, annimmt. Dadurch ist der Name "Maximum-Likelihood-Schätzer" begründet. Im Nenner des Likelihood-Quotienten steht die maximale Likelihood von S in einem Normalverteilungsmodell mit beliebiger, d.h. unrestringier-

ter, Kovarianzmatrix. Er bleibt während der Minimierungsprozedur konstant und ist daher für die Schätzung irrelevant. Zur Testung des Modells ist jedoch der Likelihood-Quotient wesentlich (siehe nächsten Abschnitt).

Der ML-Schätzer hat **bei multivariater Normalverteilung der manifesten Variablen und Gültigkeit des Modells** über die Konsistenz hinaus zahlreiche günstige Eigenschaften:

i) Er ist **asymptotisch erwartungstreu**, d.h.:

$$\lim_{N \rightarrow \infty} E(\hat{\theta}_N) = \theta$$

ii) Er ist **asymptotisch effizient**, d.h. mit wachsendem N wird seine Varianz, also sein Fehler, kleiner oder gleich der Varianz jedes beliebigen anderen Schätzers.

iii) Seine Verteilung konvergiert mit wachsendem N gegen eine **Normalverteilung**, d.h. die Stichprobenschätzer $\hat{\theta}_i$ sind asymptotisch normalverteilt um den Erwartungswert θ_i . Man kann ferner eine Schätzung für den asymptotischen **Standardfehler** $SE(\hat{\theta}_i)$ des Schätzers $\hat{\theta}_i$ berechnen und somit Vertrauensintervalle konstruieren oder Hypothesen über θ_i prüfen. LISREL berechnet auf Wunsch für jeden geschätzten Koeffizienten θ_i den geschätzten Standardfehler und zur Beurteilung der Nullhypothese:

$$H_0^i: \theta_i = 0$$

einen sogenannten **T-Wert**, der definiert ist durch den Quotienten aus dem Schätzer und seinem geschätzten Standardfehler:

$$T(\hat{\theta}_i) := \frac{\hat{\theta}_i}{\hat{SE}(\hat{\theta}_i)}$$

Bei zweiseitiger Alternativhypothese lehnt man obige H_0^i ab, wenn $|T(\hat{\theta}_i)|$ größer ist als das 97.5% - Fraktile der Standardnormalverteilung (≈ 2). Statt H_0^i kann man auch allgemeiner die Hypothese:

$$\tilde{H}_0^i: \theta_i = c$$

prüfen, wobei c eine geeignete Konstanten ist.

iv) Die zur ML-Schätzung verwendete Fit-Funktion F_{ML} erlaubt einen **globalen Modelltest**, d.h. eine Prüfung der Nullhypothese, daß die Populations-Kovarianzmatrix von der Form ist, die das Modell impliziert, gegen die Alternativhypothese, daß sie den vom Modell auferlegten Restriktionen nicht genügt (vgl. Abschnitt 2.5.2).

v) Die ML-Fitfunktion ist **skaleninvariant**.

Sofern das untersuchte Modell seinerseits skaleninvariant ist, bleiben die ML-Schätzergebnisse im wesentlichen invariant bei Veränderung der Maßeinheiten von manifesten Variablen (vgl. Abschnitt 2.4.1). Während der globale Modelltest, die T-Werte (s.o), die Modifikationsindikatoren (vgl. Abschnitt 2.5.5) und die standardisierten Residuen (vgl. Abschnitt 2.5.4) unverändert bleiben, werden die Parameterschätzer an die veränderten Maßeinheiten der manifesten Variablen angepaßt (siehe Jöreskog & Sörbom 1989).

Der entscheidende Nachteil des ML-Schätzers ist die starke Voraussetzung der multivariaten Normalverteilung der manifesten Variablen. Allerdings wurde inzwischen gezeigt, daß diese Voraussetzung liberalisiert werden kann (vgl. Bollen 1989, S. 114), und in Simulationsstudien hat sich der ML-Schätzer als robust gegen mäßige Verletzungen der Normalität erwiesen. Die ML-Schätzmethode sollte jedoch in folgenden Fällen *nicht* verwendet werden:

- Die Verteilung der manifesten Variablen weicht erheblich von der Normalverteilung ab.
- Einige der manifesten Variablen sind ordinal-skaliert, d.h. sie können z.B. nur wenige verschiedene, monoton geordnete Werte annehmen (vgl. Abschnitt 4.1). Dann sollte anstatt der Kovarianzmatrix eine Matrix mit polychorischen und polyserialen Korrelationen analysiert werden.

In diesen Fällen sind die in Abschnitt 4.2 im Zusammenhang mit der Aufbereitung problematischer Daten durch PRELIS vorzustellenden, asymptotisch verteilungsfreien Schätzmethoden **WLS** und **DWLS** indiziert.

Für unser Beispiel 1.2 (Latent-State-Modell der Angst) liefert LISREL folgende Ergebnisse:

i) ML-Schätzungen

LATENT-STATE-MODELL DER ANGST				
LISREL ESTIMATES (MAXIMUM LIKELIHOOD)				
LAMBDA X				
	KSI 1	KSI 2		
X1	1.000	0.000		
X2	1.000	0.000		
X3	0.000	1.000		
X4	0.000	1.000		
PHI				
	KSI 1	KSI 2		
KSI 1	21.971			
KSI 2	11.319	25.182		
THETA DELTA				
	X1	X2	X3	X4
X1	2.855			
X2	0.000	2.855		
X3	0.000	0.000	2.855	
X4	0.000	0.000	0.000	2.855

Die im Ergebnisprotokoll verwendeten Beschriftungen entsprechen weitgehend den in der Einleitung verwendeten Bezeichnungen der Parameter. Mit "THETA DELTA" ist die Kovarianzmatrix der Residualvariablen δ_i gemeint.

ii) Standardfehler

LATENT-STATE-MODELL DER ANGST				
STANDARD ERRORS				
PHI				
	KSI 1	KSI 2		
KSI 1	2.483			
KSI 2	2.054	2.823		
THETA DELTA				
	X1	X2	X3	X4
X1	0.214			
X2	0.000	0.214		
X3	0.000	0.000	0.214	
X4	0.000	0.000	0.000	0.214

iii) T-Werte

LATENT-STATE-MODELL DER ANGST				
T-VALUES				
	PHI			
	KSI 1	KSI 2		
KSI 1	8.850			
KSI 2	5.512	8.921		
	THETA DELTA			
	X1	X2	X3	X4
X1	13.342			
X2	0.000	13.342		
X3	0.000	0.000	13.342	
X4	0.000	0.000	0.000	13.342

Bei Voraussetzung der Modellgültigkeit werden in unserem Beispiel 1.2 also alle Nullhypothesen H_0^i abgelehnt.

Wegen ihrer vielen positiven Eigenschaften ist die ML-Methode unter LISREL-Anwendern sehr beliebt. Sie konnte zumindest bis zur Einführung der asymptotisch verteilungsfreien Schätzverfahren (vgl. Abschnitt 4.2) als "LISREL-Standard-Methode" bezeichnet werden. Dies zeigt sich z.B. im Untertitel des LISREL VI - Manuals (Jöreskog & Sörbom 1986): "Analysis of Linear Structural Relationships by the Method of Maximum Likelihood".

2.4.4 Instrumental Variables (IV) und Two-Stage Least Squares (TSLS)

Die bislang besprochenen Schätzverfahren arbeiten iterativ. Ihr Erfolg und ihre Schnelligkeit hängen ab von der Stelle im Parameterraum, an der sie gestartet werden. Für die meisten Modelle kann LISREL geeignete Startwerte automatisch berechnen. Dazu verwendet es eines der folgenden Schätzverfahren:

- Instrumental Variables (IV)
- Two-Stage Least Squares (TSLS)

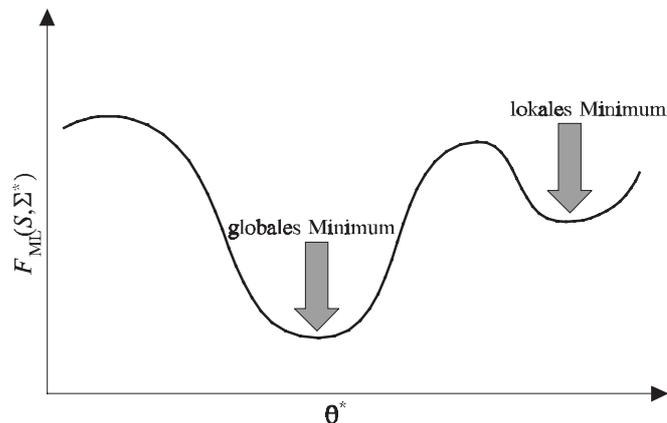
Sie verwenden zur Schätzung eines bestimmten Parameters θ_i , jeweils nur einen Teil der verfügbaren Information, sind aber immerhin konsistent und liefern bei einfachen Modellen dieselben Ergebnisse wie die iterativen Schätzverfahren.

2.4.5 Hinweise zur praktischen Anwendung der Schätzverfahren

Für alle iterativen Schätzverfahren von LISREL (ULS, GLS, ML, WLS, DWLS) gelten folgende Anmerkungen zu möglichen Problemen bei der Anwendung:

i) Wenn das Iterationsverfahren nicht konvergiert oder ein Minimum außerhalb des zulässigen Parameterraumes auftritt (z.B. negativer Varianzschätzer) kommen vor allem zwei Ursachen in Frage: Das Modell ist nicht identifiziert oder es paßt sehr schlecht zu den Daten. Ein schlechter Fit kann auch durch zu geringe Stichprobengröße entstehen (vgl. unten).

ii) Es kann passieren, daß der Algorithmus in einem lokalen Minimum landet, das kein globales Minimum ist. Dies wird in folgender Abbildung für den einfachen Spezialfall eines einparametrischen Modells demonstriert.



Gegen diese Gefahr, die sich in Simulationsstudien allerdings als relativ gering erwiesen hat, kann man sich durch mehrfache Schätzversuche ausgehend von verschiedenen Startwerten absichern (siehe unten).

iii) Der Rechenaufwand für ein iteratives Schätzverfahren ist bei größeren Modellen enorm, was allerdings aufgrund der Leistungsstärke unserer SIEMENS-Anlage kein Problem darstellt.

Ein wichtiges Thema bei Verfahren, die auf asymptotischer Statistik beruhen, ist die **minimal erforderliche Stichprobengröße**. In der Literatur (z.B. Loehlin 1987, S. 60-61) werden folgende Empfehlungen gegeben:

- Mindestens $N \geq 100$, besser $N \geq 200$
- Mindestens $N \geq 5 * t$, wobei t die Anzahl der zu schätzenden Parameter ist.

Die Schätzverfahren WLS und DWLS erfordern noch größere Stichproben (vgl. Abschnitt 6.5.9).

2.5 Beurteilung der Gültigkeit eines Modells

In diesem Abschnitt werden Kriterien zur Beurteilung der Gültigkeit eines Strukturgleichungsmodells besprochen. Es geht also um die Frage, ob das Modell eine akzeptable Erklärung der empirischen Kovarianzmatrix liefert. Der positive Ausgang dieser Prüfung ist eine Voraussetzung für die Interpretation von Parameterschätzungen und anderen Ergebnissen.

2.5.1 Untersuchung der Parameterschätzungen

Zunächst müssen alle Schätzer auf Plausibilität überprüft werden. Unzulässige Werte (z.B. negative Varianzen, Korrelationen mit einem Betrag größer als Eins) können folgende Ursachen haben:

- Das Modell paßt sehr schlecht zu den Daten.
 - Das Modell ist nicht identifiziert.
 - Die Stichprobengröße ist so klein, daß die asymptotischen Eigenschaften der Schätzer (z.B. Konsistenz) noch nicht wirksam sind.
 - Die Eingabedaten sind mangelhaft.
- Bei Verwendung einer Kovarianz- oder Korrelationsmatrix, deren Koeffizienten wegen **paarweisem Ausschluß fehlender Werte** auf unterschiedlichen Stichproben beruhen, tritt oft das Problem auf, daß diese Matrix nicht positiv semidefinit ist, d.h. eine notwendige Eigenschaft von Kovarianz- bzw.- Korrelationsmatrizen nicht besitzt. In diesem Fall muß eine alternative Behandlung fehlender Werte gewählt werden.
- Die Anweisungen an LISREL zur Analyse des Modells sind fehlerhaft.

2.5.2 Der χ^2 - Modelltest

In Abschnitt 2.4 haben wir gesehen, daß die Fitfunktion $F_{ML}(S, \Sigma^*)$ bis auf den Vorfaktor $\frac{1}{2}(N-1)$ mit dem logarithmierten Likelihoodquotienten:

$$-2 \log \left(\frac{W(S; \Sigma^*, N-1)}{W(S; S, N-1)} \right)$$

identisch ist. Nach einer allgemeinen Theorie der mathematischen Statistik ist das Minimum dieses logarithmierten Likelihoodquotienten bei Gültigkeit des Modells asymptotisch (d.h. für $N \rightarrow \infty$) χ^2 -verteilt mit:

$$df = \frac{1}{2}k(k+1) - t,$$

Freiheitsgraden, wobei wiederum k die Anzahl der manifesten Variablen und t die Anzahl der freien Parameter im Modell bedeutet. Damit erlaubt also die folgende Prüfgröße ($\hat{\Sigma}$ sei die vom geschätzten Parametervektor implizierte Kovarianzmatrix):

$$\chi^2 := (N-1) \cdot F_{ML}(S, \hat{\Sigma}) \quad (2.5)$$

eine Prüfung der Nullhypothese, daß die Populations-Kovarianzmatrix von der Form ist, die das Modell impliziert, gegen die Alternativhypothese, daß sie den vom Modell auferlegten Restriktionen nicht genügt. Lasch aber einprägsam formuliert, können wir testen, ob die Stichproben-Kovarianzmatrix von dem behaupteten Modell erzeugt worden ist.

LISREL berechnet die χ^2 -Statistik bei Verwendung der ML-Schätzmethode gemäß (2.5), bei Wahl einer anderen Schätzmethode (ULS, GLS, DWLS oder WLS) aus dem minimalen Wert der jeweiligen Fitfunktion, z.B. bei GLS gemäß:

$$\chi^2 := (N-1) \cdot F_{GLS}(S, \hat{\Sigma}).$$

Das zusätzlich ausgegebene P-level gibt die approximative Wahrscheinlichkeit für das Auftreten eines gleich großen oder größeren χ^2 -Wertes bei Gültigkeit des Modells an.

Es soll an dieser Stelle nachdrücklich daran erinnert werden, daß ein χ^2 -Modelltest basierend auf der ML-, GLS-, ULS- oder DWLS-Fitfunktion nur unter folgenden Voraussetzungen gültig ist:

- Die manifesten Variablen sind multivariat normalverteilt.
- Die Analyse basiert auf der Stichproben-**Kovarianzmatrix**; bei Verwendung der Korrelationsmatrix ist der Test im allgemeinen nicht korrekt.
- Die Stichprobe ist hinreichend groß (vgl. Abschnitt 2.4).

Bei Verwendung der in Abschnitt 4.2 vorzustellenden WLS-Schätzmethode kann auf die beiden ersten Voraussetzungen verzichtet werden, allerdings ist ein erheblich größerer Stichprobenumfang erforderlich.

Für unser Modell in Beispiel 1.2 und die Daten in Abschnitt 2.2 liefert LISREL bei Wahl der ML-Methode¹ folgendes Ergebnis (die Fit-Indizes in den letzten drei Zeilen werden später erläutert):

CHI-SQUARE WITH	6 DEGREES OF FREEDOM =	7.53 (P = .274)
	GOODNESS OF FIT INDEX =	0.980
	ADJUSTED GOODNESS OF FIT INDEX =	0.966
	ROOT MEAN SQUARE RESIDUAL =	0.668

Weil das P-level größer als 0.05 ist, muß die Nullhypothese nicht verworfen werden. Unser Modell hat diesen Test also bestanden. Allerdings darf man wie üblich aus einem nicht-signifikanten Ergebnis keinesfalls auf die Gültigkeit der Nullhypothese (also des Modells) schließen.

Ein wesentlicher Unterschied zwischen diesem Modelltest und der üblichen Testpraxis (z.B. beim t-Test zum Vergleich zweier Populations-Erwartungswerte) liegt darin, daß hier die Beibehaltung der H_0 im Interesse des Forschers liegt, während er bei traditionellen Testproblemen über eine Ablehnung der Nullhypothese (z.B. $H_0: \mu_1 = \mu_2$) erfreut ist.

Die LISREL-Urheber (vgl. z.B. Jöreskog & Sörbom 1986, S. I.39) raten davon ab, die χ^2 -Statistik im Sinne eines strengen Hypothesentests zu interpretieren, denn:

- Die χ^2 -Statistik kann bei Verletzungen der Normalverteilungsannahme erheblich zu groß werden, so daß auch gut passende Modelle verworfen werden.
- Ein Strukturgleichungsmodell ist in der sozialwissenschaftlichen Forschung meist als *Approximation* der Realität gedacht. Insofern besteht das Ziel der statistischen Analyse nicht darin, eine apriori als falsch anzunehmende Nullhypothese zu testen, sondern vielmehr darin, die Angemessenheit des Modells zu beurteilen.

Jöreskog & Sörbom empfehlen daher die Interpretation des χ^2 -Wertes als deskriptiven Fit-Index, insbesondere auch zum Vergleich verschiedener Modelle.

Einige Anmerkungen sind auch zur **Rolle der Stichprobengröße** bei den von LISREL angebotenen inferenzstatistischen Ergebnissen angebracht. Wie oben gezeigt wurde, enthält die χ^2 -Statistik den Faktor $(N-1)$, d.h. bei gleicher Matrix S und gleichem Modell führt ein doppelt so großes N zu einer ungefähr doppelt so großen χ^2 -Statistik. Bei Gültigkeit des Modells ist dies eine sinnvolle Normierung, weil bei wachsendem N eine immer bessere Übereinstimmung zwischen S und der vom Modell implizierten, wahren Populationsmatrix zu erwarten ist. Ist das Modell jedoch streng genommen falsch, was nach obigen Überlegungen in der Sozialwissenschaft als sicher angenommen werden kann, so konvergiert S bei wachsendem N nicht gegen die vom Modell implizierte Kovarianzmatrix und der χ^2 -Wert steigt mit N drastisch an. Dies bedeutet, daß die Power des χ^2 -Tests zur Aufdeckung des als sicher anzunehmenden Modelldefekts mit N ansteigt. Mit wachsendem N geht also auch für ein "gut passendes" Modell die Wahrscheinlichkeit gegen Eins, daß es (streng genommen zurecht) als falsch verworfen wird.

Der t-Test zur Beurteilung der Nullhypothese $H_0^i: \theta_i = 0$ über einen einzelnen Parameter ist nach dem üblichen Schema angelegt: Der Forscher hat den Parameter θ_i ins Modell aufgenommen, weil er ihn für wichtig hält, und freut sich über das Scheitern der H_0^i . Eine Steigerung der Stichprobengröße erhöht in

¹ Diese (bei LISREL voreingestellte) Schätzmethode liegt auch allen später berichteten Ergebnissen zu den Beispielen 1.1, 1.2 und 1.3 zugrunde.

seinem Sinne die Power des t-Tests, eine falsche H_0^i zu verwerfen. Der t-Test hängt nämlich wesentlich vom Standardfehler des Schätzers ab, der bei wachsendem N gegen 0 strebt. Dies zeigt folgende Formel, worin Δ eine von N unabhängige Größe ist:

$$\text{SE}(\hat{\theta}_i) = \frac{\Delta}{\sqrt{N-1}}$$

2.5.3 Andere globale Fit-Indikatoren

LISREL liefert neben der χ^2 -Statistik noch drei andere Maße für die globale Modellanpassung, die rein deskriptiv zu verstehen sind. Sie haben derzeit in der Anwendungspraxis eine deutlich geringere Bedeutung als die χ^2 -Statistik.

i) Goodness-of-fit - Index (GFI)

Der GFI mißt, vereinfacht ausgedrückt, welchen Anteil an der gewichteten Summe der quadrierten, nicht-redundanten Elemente der Stichproben-Kovarianzmatrix S vom Modell erklärt wird, wobei die Gewichtungsmatrix W^{-1} von der Schätzmethode abhängt.

$$\text{GFI} := 1 - \frac{\text{tr}[(S - \hat{\Sigma})^T W^{-1} (S - \hat{\Sigma})]}{\text{tr}[S^T W^{-1} S]}$$

Sein Wert erreicht bei perfektem Fit den Wert Eins und geht bei schlechtem Fit gegen Null; allerdings sind auch negative Werte möglich. Im Unterschied zur χ^2 -Statistik ist der GFI unabhängig von der Stichprobengröße (vgl. Abschnitt 2.5.2).

Bei der ML-Schätzmethode wird die Gewichtungsmatrix $(\hat{\Sigma}^T \hat{\Sigma})^{-1}$ verwendet, so daß hier gilt:

$$\text{GFI} := 1 - \frac{\text{tr}[(\hat{\Sigma}^{-1} S - I)^T (\hat{\Sigma}^{-1} S - I)]}{\text{tr}[(\hat{\Sigma}^{-1} S)^T (\hat{\Sigma}^{-1} S)]}$$

Bei perfektem Fit gilt:

$$\hat{\Sigma} = S \quad \Rightarrow \quad \hat{\Sigma}^{-1} S = I \quad \Rightarrow \quad \text{GFI} = 1$$

Bei der ULS-Schätzmethode wird mit der Einheitsmatrix gewichtet, so daß eine besonders anschauliche Definition resultiert:

$$\text{GFI} := 1 - \frac{\text{tr}[(S - \hat{\Sigma})^T (S - \hat{\Sigma})]}{\text{tr}[(S^T S)]}$$

Hier wird einfach die Summe der quadrierten elementweisen Diskrepanzen zwischen S und $\hat{\Sigma}$ durch die Summe der quadrierten Elemente von S dividiert. Dieser Quotient wird von 1 subtrahiert.

ii) Adjusted goodness-of-fit - Index (AGFI)

Bei dieser Verfeinerung des GFI wird die Anzahl der verwendeten Parameter in Rechnung gestellt:

$$\text{AGFI} := 1 - \frac{k(k+1)}{2d} (1 - \text{GFI})$$

Dabei bedeutet k die Anzahl der manifesten Variablen und d die Anzahl der Freiheitsgrade im Modells, also $d = \frac{1}{2}k(k+1) - t$, wobei t die Anzahl der Parameter ist. Der AGFI bestraft also im Sinne des wissenschaftsphilosophischen Sparsamkeitsprinzips eine hohe Parameterzahl.

iii) Root Mean Square Residual (RMR)

Das RMR-Maß ist definiert als Wurzel aus der mittleren quadrierten Diskrepanz zwischen S und $\hat{\Sigma}$:

$$\text{RMR} := \left[\frac{1}{\frac{1}{2}k(k+1)} \sum_{i=1}^k \sum_{j=1}^i (s_{ij} - \hat{\sigma}_{ij})^2 \right]^{\frac{1}{2}}$$

In der Formel steht " s_{ij} " bzw. " $\hat{\sigma}_{ij}$ " für das Element (i,j) der Matrix S bzw. $\hat{\Sigma}$. RMR ist nicht normiert und kann daher nur in Relation zu den Varianzen und Kovarianzen der manifesten Variablen interpretiert werden. Dies gelingt am besten, wenn alle manifesten Variablen dieselbe Maßeinheit besitzen oder wenn sie standardisiert sind (also S eine Korrelationsmatrix ist).

In unserem Beispiel ist der RMR-Wert von 0.668 gemessen an den Varianzen der manifesten Variablen (≈ 25) als günstig zu beurteilen.

Übung: Zeigen Sie, daß RMR wie folgt aus der ULS-Fitfunktion $F_{\text{ULS}}(S, \Sigma^*)$ berechnet werden kann:

$$\text{RMR} = \left[\frac{1}{k(k+1)} 4F_{\text{ULS}}(S, \hat{\Sigma}) \right]^{\frac{1}{2}}$$

2.5.4 Residuen und standardisierte Residuen

Wenn ein globaler Fit-Indikator unbefriedigend ausfällt, müssen die Diskrepanzen zwischen S und $\hat{\Sigma}$ im Detail untersucht werden. Dazu kann LISREL ausgeben:

- Die vom Modell implizierte Kovarianzmatrix $\hat{\Sigma} = f(\hat{\theta})$
Bezeichnung im Ergebnisprotokoll: "FITTED COVARIANCE MATRIX"
- Die Residuen $S - \hat{\Sigma}$
Bezeichnung im Ergebnisprotokoll: "FITTED RESIDUALS"
- Standardisierte Residuen

Der Quotient aus dem einfachen Residuum $(s_{ij} - \hat{\sigma}_{ij})$ zum Element s_{ij} von S und seinem asymptotischen Standardfehler wird als "standardisiertes Residuum" bezeichnet. Bei Verwendung der ML- bzw. GLS-Schätzmethode und gültiger Normalverteilungsannahme oder bei korrekter Anwendung der WLS-Schätzmethode sind die standardisierten Residuen in großen Stichproben standardnormalverteilt, sofern das Modell gültig ist.

Bezeichnung im Ergebnisprotokoll: "STANDARDIZED RESIDUALS"

Für unser Beispiel 1.2 ermittelt LISREL folgende Werte:

```

LATENT-STATE-MODELL DER ANGST

      FITTED COVARIANCE MATRIX
            X1          X2          X3          X4
X1  -----
X2  24.826          24.826
X3  11.319          11.319          28.037
X4  11.319          11.319          25.182          28.037

      FITTED RESIDUALS
            X1          X2          X3          X4
X1  -----
X2  -0.156          0.309
X3  -0.966          -0.695          -0.798
X4  0.346          1.317          0.076          0.646

SUMMARY STATISTICS FOR FITTED RESIDUALS
SMALLEST FITTED RESIDUAL = -0.966
MEDIAN FITTED RESIDUAL = 0.000
LARGEST FITTED RESIDUAL = 1.317

STEMLEAF PLOT
- 1 | 0
- 0 | 87
- 0 | 21
  0 | 133
  0 | 6
  1 | 3

      STANDARDIZED RESIDUALS
            X1          X2          X3          X4
X1  -----
X2  -0.178          0.351
X3  -1.505          -1.083          -0.852
X4  0.538          2.050          0.711          0.690

SUMMARY STATISTICS FOR STANDARDIZED RESIDUALS
SMALLEST STANDARDIZED RESIDUAL = -1.505
MEDIAN STANDARDIZED RESIDUAL = 0.087
LARGEST STANDARDIZED RESIDUAL = 2.050

STEMLEAF PLOT
- 1 | 51
- 0 | 972
  0 | 4577
  1 |
  2 | 0

```

Eine grobe Signifikanz-Beurteilung der standardisierten Residuen kann durch Vergleich ihres Betrages mit dem kritischen Wert 2.5 ($\approx 99.5\%$ - Fraktile der Standardnormalverteilung) erfolgen. Danach unterscheiden sich in unserem Beispiel 1.2 alle standardisierten Residuen nur durch zufällige Stichprobenvariationen von Null.

Nach obigen Überlegungen ist bei Gültigkeit des Modells als empirische Verteilung der $\frac{1}{2}k(k+1)$ standardisierten Residuen ebenfalls approximativ die Standardnormalverteilung zu erwarten.¹ Zur Prüfung

¹ Dies gilt streng genommen nur bei Unabhängigkeit der normalisierten Residuen, die im allgemeinen **nicht** gegeben ist.

dieser Bedingung kann LISREL 7 zwei semigraphische Ausgaben erzeugen: einen Stem-and-Leaf - Plot und einen Quantil-Plot.

Der **Stem-and-Leaf (Stamm-und-Blatt) - Plot** wird für die einfachen und für die standardisierten Residuen erstellt. Er zeigt die empirische Verteilung in einem Histogramm-ähnlichen Format. Alle beobachteten Werte, welche die ersten d Dezimalstellen gemeinsam haben, bilden eine Klasse; die gemeinsamen d Ziffern sind der zugehörige Stamm. Das Blatt eines beobachteten Wertes ist seine $(d+1)$ -te Ziffer.

Beispiel (mit $d = 1$):

Wert	Stamm	Blatt
1.105	1	1
1.536	1	5
2.050	2	0

Bei der Stamm-und-Blatt-Darstellung werden die Stämme aller Klassen untereinander geschrieben. Nach einem senkrechten Strich folgen dann pro Klasse die Blätter der einzelnen Beobachtungen in aufsteigender Sortierung. Die Plots für unser Beispiel 1.2 sind in der letzten Ergebnis-Box wiedergegeben.

Der **Quantil-Plot (Q-Plot)** zeigt die Übereinstimmung zwischen der empirischen Verteilungsfunktion der standardisierten Residuen und der theoretischen Verteilungsfunktion der Standardnormalverteilung. Zum Verständnis dieses Plots ist noch ein wenig Wahrscheinlichkeitstheorie nötig: Hat das standardisierte Residuum sr_{ij} den Rangplatz rg_{ij} in der nach Größe geordneten Reihe aller standardisierten Residuen, so hat die empirische Verteilungsfunktion F_e an der Stelle sr_{ij} den Wert

$$F_e(sr_{ij}) := rg_{ij} / \frac{1}{2}k(k+1).$$

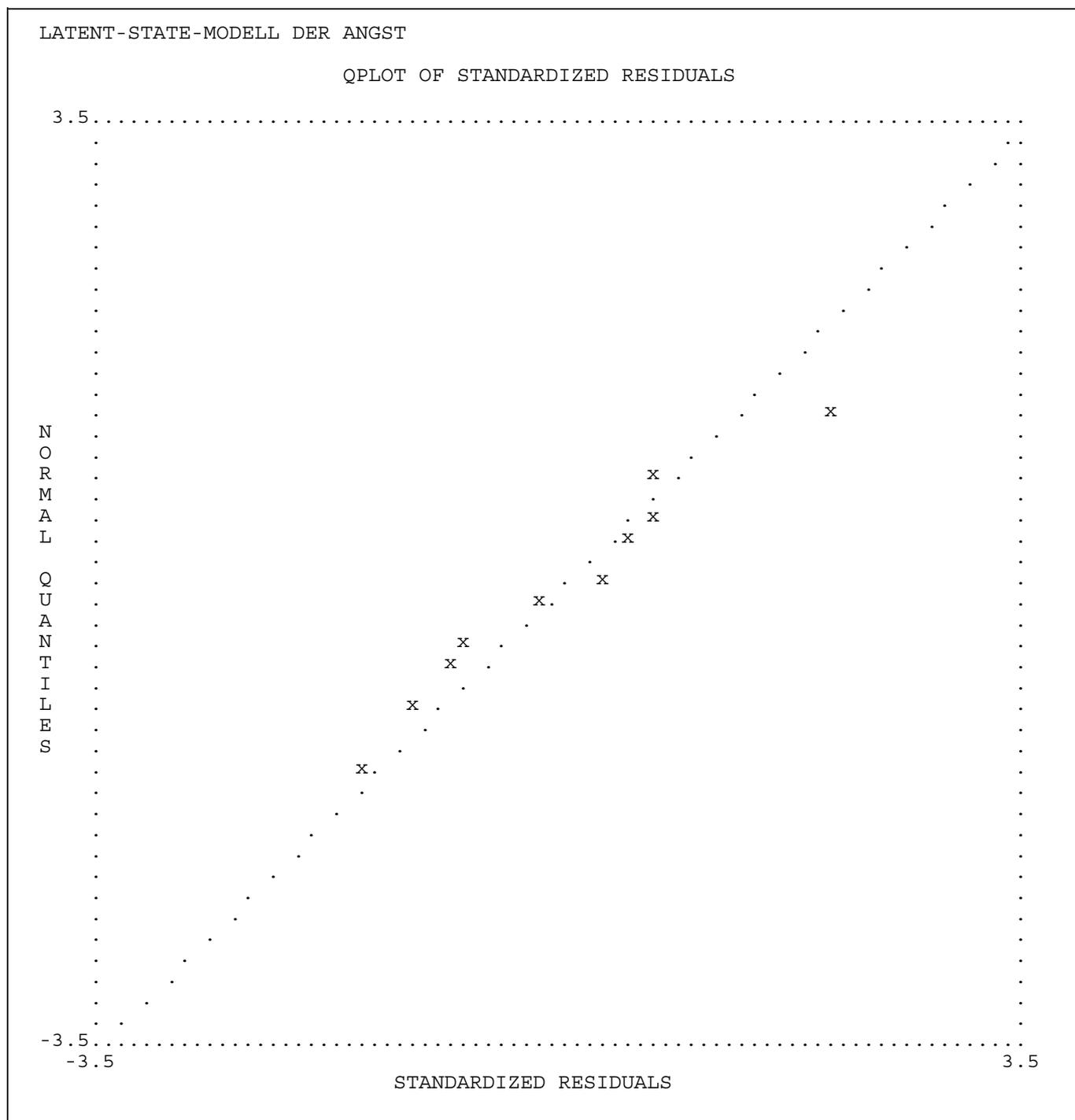
Die empirische Verteilung hat also bis zur Stelle sr_{ij} gerade $rg_{ij} / \frac{1}{2}k(k+1)$ % ihrer Gesamtmasse 1 vergeben. F_e stimmt dann gut mit der theoretischen Verteilungsfunktion F_t überein, wenn die Standardnormalverteilung in der Nähe von sr_{ij} ebenfalls die Masse $rg_{ij} / \frac{1}{2}k(k+1)$ erreicht, d.h. wenn gilt:

$$sr_{ij} \approx F_t^{-1}(rg_{ij} / \frac{1}{2}k(k+1)) = F_t^{-1}(F_e(sr_{ij}))$$

Im Q-Plot werden die sr_{ij} , $i=1, \dots, k$, $j=1, \dots, i$ gegen $F_t^{-1}(F_e(sr_{ij}))$ abgetragen. Bei Gültigkeit des Modells sollte sich folglich eine Ursprungsgrade mit Steigung 1 ergeben. Der Q-Plot enthält auch eine entsprechende Referenzlinie. Ein schlecht passendes Modell erbringt größere Residuen, so daß im Q-Plot (bei Gültigkeit der Normalverteilungsannahme) eine Gerade mit einer Steigung < 1 resultiert. Eine starke Nichtlinearität des Q-Plots deutet auf Spezifikationsfehler oder auf eine Verletzung der Normalitätsannahme hin.

Residuen mit dem Wert Null werden vom Q-Plot ausgeschlossen. LISREL nimmt hier an, daß der perfekte Fit trivialerweise gelang, z.B. weil für das entsprechende Element von S ein eigener freier Parameter im Modell enthalten ist.

Für Beispiel 1.2 erstellt LISREL folgenden Q-Plot:



Überraschenderweise erklärt das aktuelle LISREL-Handbuch (Jöreskog & Sörbom 1989, S. 34), Q-Plots mit Steigung 1 seien indikativ für einen mäßigen Fit, während bei gutem Fit eine Steigung > 1 zu erwarten sei. Hier handelt es sich offenbar um Erfahrungswerte mit LISREL VI, die auf einem Programmfehler dieser älteren Version basieren. LISREL 7 - Anwender sollten die Interpretationsempfehlung des Handbuchs ignorieren.

2.5.5 Modifikationsindikatoren

Während die Residuen Hinweise dazu geben, welche Elemente der Kovarianzmatrix S vom Modell schlecht erklärt werden, dienen die Modifikationsindikatoren zur Ortung von Parameter-Fixierungen oder Gleichheits-Restriktionen, die empirisch ungültige Einschränkungen für die Kovarianzmatrix der manifesten Variablen implizieren und dadurch den Fit beeinträchtigen. Der Modifikationsindikator zu einem fixierten oder restringierten Parameter gibt an, welche Reduktion der χ^2 -Statistik zu erwarten ist, wenn er freigesetzt und anschließend das (erweiterte) Modell neu geschätzt wird. Damit kann ein Modifikationsindikator bei Verwendung der ML-, GLS- oder WLS-Schätzmethode anhand der χ^2 -Verteilung mit einem Freiheitsgrad beurteilt werden (vgl. Abschnitt 5). Je nach bevorzugter Regel können daher Werte größer 4 (\approx 95% - Fraktile) oder 7 (\approx 99% - Fraktile) als Kritik an der zugrundeliegenden Fixierung oder Restriktion aufgefaßt werden, wobei in der Praxis meist die 7 als kritische Grenze verwendet wird (vgl. z.B. Kühnel, 1993). LISREL gibt weiterhin für jeden fixierten bzw. restringierten Parameter seine erwartete Veränderung bei Freisetzung an (Bezeichnung im Ergebnisprotokoll: "ESTIMATED CHANGE"). Zusätzliche Hinweise zur Verwendung der Modifikationsindikatoren folgen in Abschnitt 5 "Modell-diagnose und Modellmodifikation".

Die Modifikationsindikatoren für Beispiel 1.2:

```

LATENT-STATE-MODELL DER ANGST
MODIFICATION INDICES AND ESTIMATED CHANGE

      MODIFICATION INDICES FOR LAMBDA X
      KSI 1          KSI 2
X1    0.083         0.450
X2    0.083         0.450
X3    3.663         0.692
X4    3.663         0.692

      ESTIMATED CHANGE FOR LAMBDA X
      KSI 1          KSI 2
X1   -0.011        -0.025
X2    0.011         0.025
X3   -0.075        -0.031
X4    0.075         0.031

NO NON-ZERO MODIFICATION INDICES FOR PHI

      MODIFICATION INDICES FOR THETA DELTA
      X1          X2          X3          X4
X1    0.412
X2    0.506         0.363
X3    0.854         3.361         0.343
X4    1.499         4.552         0.506         0.459

      ESTIMATED CHANGE FOR THETA DELTA
      X1          X2          X3          X4
X1    0.361
X2   -0.305         0.338
X3    0.370        -0.734        -0.333
X4   -0.490         0.854         0.305        -0.386

      MAXIMUM MODIFICATION INDEX IS      4.55 FOR ELEMENT ( 4, 2 ) OF THETA DELTA
    
```

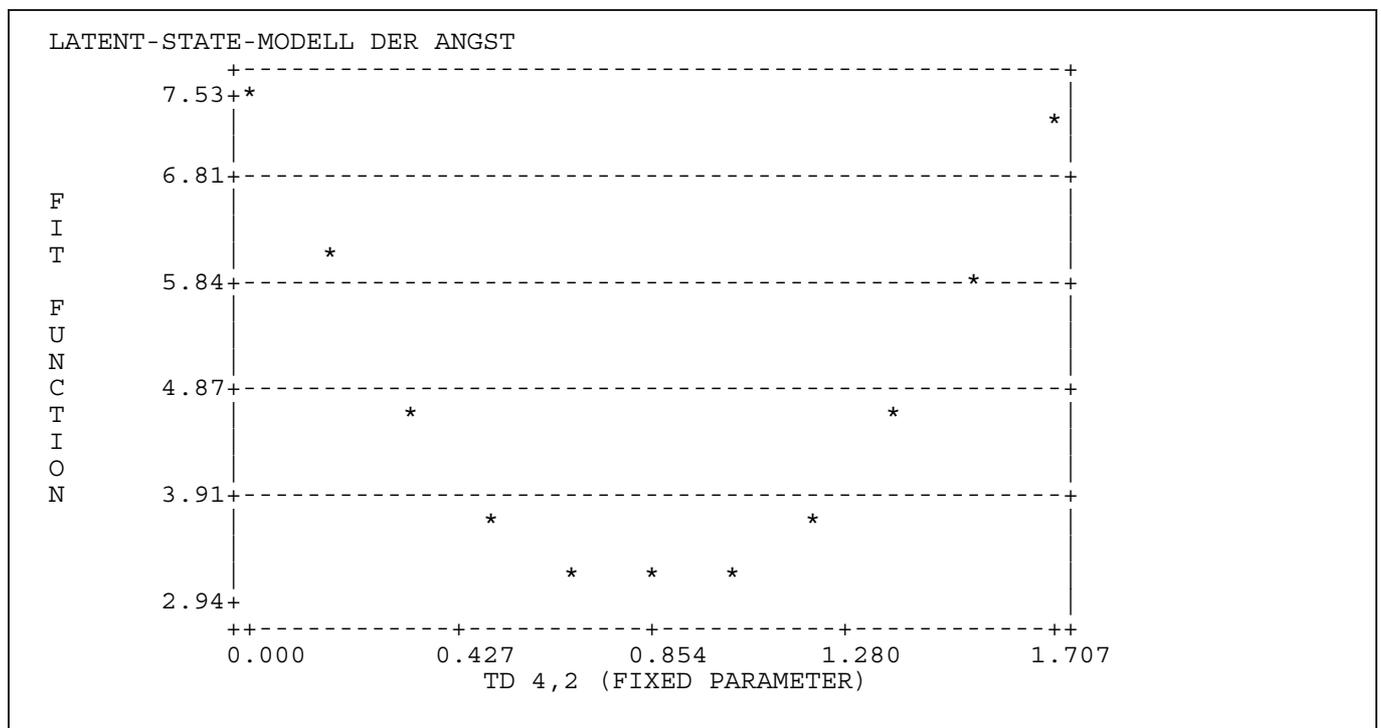
Der Modifikationsindex für $Cov(\delta_4, \delta_2)$ (im Modell auf 0 fixiert) ist mit 4.55 als einziger größer als 4. Die beiden aus der zweiten Testhälfte stammenden Indikatoren haben vermutlich über den Einfluß der

latenten Variablen hinaus eine gemeinsame Wurzel. Durch Freisetzen von $\text{Cov}(\delta_4, \delta_2)$ reduziert sich die χ^2 -Statistik um 4.59 von 7.53 auf 2.94:

CHI-SQUARE WITH	5 DEGREES OF FREEDOM =	2.94 (P = .709)
	GOODNESS OF FIT INDEX =	0.992
	ADJUSTED GOODNESS OF FIT INDEX =	0.984
	ROOT MEAN SQUARE RESIDUAL =	0.538

LISREL erstellt auf Wunsch einen Plot der Fit-Funktion in Abhängigkeit von einem interessierenden Parameter (fixiert oder frei). Dabei wird für 11 verschiedene, automatisch in einem relevanten Intervall gewählte Parameter-Ausprägungen die Fit-Funktion bzgl. aller freier Modellparameter minimiert. Ein solcher Plot verursacht also annähernd den Aufwand von 11 LISREL-Analysen.

Für $\text{Cov}(\delta_4, \delta_2)$ aus Modell 1.2 erhalten wir:



Übungsaufgaben

Aufgabe 2.1 (Minderung von Strukturkoeffizienten durch Meßfehler bei explanatorischen Variablen)

Im Zusammenhang mit Beispiel 1.1 wurde darauf hingewiesen, daß in Strukturgleichungsmodellen für direkt beobachtbare Variablen zumindest für die in explanatorischer Funktion auftretenden Variablen vorausgesetzt werden muß, daß sie fehlerfrei gemessen werden können, weil anderenfalls die Schätzer für die Effektparameter verfälscht sind. Diese Aussage kann nun mit den Hilfsmitteln aus Abschnitt 2.1 an einem einfachen Beispiel demonstriert werden. Es gelte für die fehlerfrei gemessene Variable X_T das Modell:

$$Y = \gamma_T \cdot X_T + \zeta_T$$

Nun stehe an Stelle von X_T nur die fehlerbehaftete Variable $X = X_T + \delta$ zur Verfügung ($\text{Var}(\delta) > 0$, $\text{Cov}(X_T, \delta) = 0$, $\text{Cov}(\zeta_T, \delta) = 0$). Offenbar ist X eine durch δ "verwässerte" Variante von X_T , der man nur noch einen abgeschwächten Einfluß auf Y zutrauen kann. Beweisen Sie bitte, daß der Strukturkoeffizient γ aus dem de facto untersuchten Modell:

$$Y = \gamma \cdot X + \zeta,$$

tatsächlich kleiner ist als der (wahre) Effekt γ_T , daß nämlich gilt:

$$\gamma = \gamma_T \cdot \text{Rel}(X) < \gamma_T.$$

Dabei bezeichnet $\text{Rel}(X)$ die Reliabilität von X , die definiert ist durch:

$$\text{Rel}(X) := \frac{\text{Var}(X_T)}{\text{Var}(X)}$$

Hinweis: Da die betrachteten Modelle jeweils nur aus einer einzigen Gleichung bestehen, sind die auftretenden Strukturkoeffizienten nichts anderes als Regressionskoeffizienten, und für den Koeffizienten b_{YX} aus der einfachen Regression von Y auf X gilt bekanntlich:

$$b_{YX} = \frac{\text{Cov}(X, Y)}{\text{Var}(X)}$$

Aufgabe 2.2 (Berufliche Ambitionen von Jugendlichen; Auswertung per Regressionsanalyse)

Das Modell in Beispiel 1.1 kann als rekursives Strukturgleichungsmodell mit manifesten Variablen ("Pfadmodell") auch mit regressionsanalytischen Methoden behandelt werden. Rechnen Sie bitte mit der SPSS-Prozedur REGRESSION unter Verwendung der Korrelationsmatrix in Abschnitt 2.2 für beide Strukturgleichungen eine multiple Regressionsanalyse.

Hinweise: - Sie müssen die Korrelationsmatrix zunächst mit dem Kommando MATRIX DATA einlesen.
- Im REGRESSION-Kommando müssen Sie das Einlesen einer Korrelationsmatrix anfordern durch das Subkommando: MATRIX=IN(*)

Aufgabe 2.3 (Latent-State-Modell der Angst; exploratorische Faktorenanalyse)

Rechnen Sie mit der SPSS-Prozedur FACTOR eine exploratorische Faktorenanalyse mit der Kovarianzmatrix in Abschnitt 2.2. Wählen Sie zur Faktorenextraktion die Hauptachsenmethode mit Kommunalitäteniteration und probieren Sie sowohl die schiefwinklige OBLIMIN-Rotation als auch die (im Beispiel sicher ungeeignete) orthogonale VARIMAX-Rotation aus.

Hinweise: - Sie müssen die Kovarianzmatrix zunächst mit dem Kommando MATRIX DATA einlesen.
- Da FACTOR nur Korrelationsmatrizen verarbeiten kann, müssen Sie mit dem Kommando MCONVERT aus der Kovarianzmatrix eine Korrelationsmatrix erstellen lassen.

- a) Wählen von den beiden Lösungen (orthogonal vs. schiefwinklig) die beste aus und begründen Sie Ihre Entscheidung.
- b) Welche Aspekte des Modells werden durch die exploratorische Faktorenanalyse bestätigt?
- c) Liefern die Ergebnisse Hinweise auf Schwächen des Modells?

Titelseite

3 Das LISREL-Modell

Die bisherige Darstellung der Kovarianzstrukturanalyse war im wesentlichen unabhängig von LISREL wengleich gelegentlich seine Terminologie übernommen und seine Ausgaben zur Illustration eingesetzt wurden. Nun wird das LISREL-Modell systematisch und allgemein dargestellt.

Vorab soll noch einmal an die elementare Annahme der Kovarianzstrukturanalyse und damit auch der LISREL-Analyse erinnert werden, daß die Verteilung der beobachteten Variablen hinreichend gut durch die **Momente erster und zweiter Ordnung (d.h. durch Erwartungswerte, Varianzen und Kovarianzen)** beschrieben werden kann. Diese Annahme ist z.B. im Fall multivariater Normalverteilung der manifesten Variablen perfekt erfüllt.

3.1 Variablentypen

In einem LISREL-Modell können die folgenden sieben Vektoren von Zufallsvariablen auftreten:

Tab. 3.1 Variablentypen in LISREL

Beschreibung	Bezeichnung für den Zufallsvektor		Bezeichnung für die Komponentenzahl des Zufallsvektors	
	math.	LISREL	math.	LISREL
Vektor der latenten endogenen Variablen	η	ETA	m	NE
Vektor der latenten exogenen Variablen	ξ	KSI	n	NK
Vektor der Indikatoren der latenten endogenen Variablen	Y	Y	p	NY
Vektor der Indikatoren der latenten exogenen Variablen	X	X	q	NX
Vektor der Residualvar. der η -Variablen	ζ	ZETA		
Vektor der Residualvar. der Y-Variablen	ε	EPSILON		
Vektor der Residualvar. der X-Variablen	δ	DELTA		

Alle Variablentypen sind in dem konstruierten Beispiel 1.3 vertreten, das in den nächsten Abschnitten zur Illustration dienen wird:

$$\begin{array}{ll}
 \eta = (\eta_1, \eta_2), & m = 2 \\
 \xi = (\xi_1, \xi_2, \xi_3), & n = 3 \\
 Y = (Y_1, Y_2, Y_3, Y_4), & p = 4 \\
 X = (X_1, X_2, X_3, X_4, X_5, X_6, X_7), & q = 7 \\
 \zeta = (\zeta_1, \zeta_2) & \\
 \varepsilon = (\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4) & \\
 \delta = (\delta_1, \delta_2, \delta_3, \delta_4, \delta_5, \delta_6, \delta_7) &
 \end{array}$$

3.2 Meßmodell und Strukturgleichungsmodell

Ein allgemeines LISREL-Modell besteht aus zwei Teilmodellen:

- Die Gleichungen zur Verankerung der latenten Variablen in der Beobachtungsbasis der manifesten Variablen bilden das Meßmodell.
- Die Gleichungen über kausale Wirkungen zwischen Variablen bilden das Strukturgleichungsmodell.

3.2.1 Meßmodell

Die Y -Variablen werden im Sinne der konfirmatorischen Faktorenanalyse erklärt durch die η -Variablen (\approx gemeinsame Faktoren) und die Residualvariablen in ε (\approx spezifische Faktoren). Analog werden die X -Variablen erklärt durch die ξ -Variablen und die Residualvariablen in δ . Dadurch werden die theoretischen Begriffe mit beobachtbaren Variablen verknüpft und damit einer empirischen Erforschung zugänglich gemacht. Das Meßmodell lautet:

$$\begin{aligned} Y &= \Lambda_Y \eta + \varepsilon \\ X &= \Lambda_X \xi + \delta \end{aligned} \tag{3.1}$$

Dabei enthält das Element $\lambda_{ij}^{(Y)}$ der $(p \times m)$ -Ladungsmatrix Λ_Y die Ladung der Variablen Y_i auf dem Faktor η_j und das Element $\lambda_{ij}^{(X)}$ der $(q \times n)$ -Ladungsmatrix Λ_X die Ladung der Variablen X_i auf dem Faktor ξ_j . Die Residualvariablen in den Vektoren ε und δ werden auch als "**Fehler in den Variablen**" bezeichnet.

In Beispiel 1.3 enthält das Meßmodell für die Y -Variablen die Gleichungen:

$$\begin{aligned} Y_1 &= \eta_1 + \varepsilon_1 \\ Y_2 &= \lambda_{21}^{(Y)} \eta_1 + \varepsilon_2 \\ Y_3 &= \eta_2 + \varepsilon_3 \\ Y_4 &= \lambda_{42}^{(Y)} \eta_2 + \varepsilon_4 \end{aligned}$$

Dies ergibt in Matrixschreibweise:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \lambda_{21}^{(Y)} & 0 \\ 0 & 1 \\ 0 & \lambda_{42}^{(Y)} \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{bmatrix}$$

Das Meßmodell für die X-Variablen besteht aus den Gleichungen:

$$\begin{aligned}
 X_1 &= \xi_1 + \delta_1 \\
 X_2 &= \lambda_{21}^{(X)} \xi_1 + \delta_2 \\
 X_3 &= \lambda_{31}^{(X)} \xi_1 + \lambda_{32}^{(X)} \xi_2 + \delta_3 \\
 X_4 &= \xi_2 + \delta_4 \\
 X_5 &= \lambda_{52}^{(X)} \xi_2 + \delta_5 \\
 X_6 &= \xi_3 + \delta_6 \\
 X_7 &= \lambda_{73}^{(X)} \xi_3 + \delta_7
 \end{aligned}$$

In Matrixschreibweise:

$$\begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \\ X_6 \\ X_7 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ \lambda_{21}^{(X)} & 0 & 0 \\ \lambda_{31}^{(X)} & \lambda_{32}^{(X)} & 0 \\ 0 & 1 & 0 \\ 0 & \lambda_{52}^{(X)} & 0 \\ 0 & 0 & 1 \\ 0 & 0 & \lambda_{73}^{(X)} \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{bmatrix} + \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \\ \delta_6 \\ \delta_7 \end{bmatrix}$$

In jeder Spalte von Λ_Y bzw. Λ_X wurde eine Ladung auf 1 fixiert, um die Maßeinheit der betreffenden latenten Variablen festzulegen (vgl. die Erläuterungen zu Beispiel 1.2 in Abschnitt 2.2).

3.2.2 Das Strukturgleichungsmodell

Das Strukturgleichungsmodell gibt die kausale Abhängigkeit der η -Variablen von den ξ -Variablen und anderen η -Variablen an. Es lautet in Matrixschreibweise:

$$\begin{aligned}
 \eta &= B\eta + \Gamma\xi + \zeta \\
 \Leftrightarrow (I - B)\eta &= \Gamma\xi + \zeta
 \end{aligned} \tag{3.2}$$

Die $(m \times m)$ -Matrix B enthält die (direkten) Effekte von η -Variablen auf andere η -Variablen. Bei einem Element β_{ij} von B gibt der erste Index die beeinflusste Variable und der zweite Index die beeinflussende Variable an. In der Hauptdiagonalen von B müssen Nullen stehen, da keine Variable sich *direkt* selbst beeinflussen darf. Das Element γ_{ij} der $(m \times n)$ -Matrix Γ gibt den direkten Effekt der exogenen, latenten Variablen ξ_j auf die endogene, latente Variable η_i an. Die Residualvariablen im Vektor ζ werden auch als "**Fehler in den Gleichungen**" bezeichnet.

Die Äquivalenz-Behauptung in Formel (3.2) wird gerechtfertigt durch die LISREL-Voraussetzung, daß die **Matrix $(I - B)$ regulär** ist. Um den Hintergrund dieser nicht restriktiven Annahme einzusehen, betrachten wir die zweite Zeile von Formel (3.2). Diese äquivalente Schreibweise des Strukturgleichungsmodells enthält für m Linearkombinationen aus η -Variablen jeweils eine Darstellung in Abhängigkeit von den ξ -Variablen. Jede Zeile von $(I - B)$ wählt eine Linearkombination der η -Variablen als Zielvariable

aus. Bei Singularität von $(I - B)$ wäre mindestens eine Zielvariable eine Linearkombination der übrigen, woraus eine Darstellung in Abhängigkeit der ξ -Variablen bereits festliegen würde und die fragliche Strukturgleichung überflüssig wäre oder gar einen Widerspruch produzieren würde.

Das Strukturgleichungsmodell zu Beispiel 1.3 lautet:

$$\begin{aligned}\eta_1 &= \beta_{12}\eta_2 + \gamma_{11}\xi_1 + \gamma_{12}\xi_2 + \zeta_1 \\ \eta_2 &= \beta_{21}\eta_1 + \gamma_{21}\xi_1 + \gamma_{23}\xi_3 + \zeta_2\end{aligned}$$

In Matrix-Schreibweise:

$$\begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} = \begin{bmatrix} 0 & \beta_{12} \\ \beta_{21} & 0 \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} & 0 \\ \gamma_{21} & 0 & \gamma_{23} \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{bmatrix} + \begin{bmatrix} \zeta_1 \\ \zeta_2 \end{bmatrix}$$

3.2.3 LISREL-Annahmen über Erwartungswerte und Kovarianzen

Bei den im vorliegenden Manuskript dargestellten LISREL-Modellen wird angenommen, daß alle Variablen den Erwartungswert 0 haben, d.h.

$$E(Y) = E(\varepsilon) = E(X) = E(\delta) = E(\eta) = E(\xi) = E(\zeta) = 0.$$

Über LISREL-Modelle mit freien Parametern für die Erwartungswerte latenter Variablen informieren z.B. Jöreskog & Sörbom (1989, Kapitel 10).

In *jedem* LISREL-Modell werden Über die Kovarianzen der Residuen mit anderen Variablen folgende Annahmen gemacht:

- 1) Im Strukturmodell sind die Fehler in den Gleichungen (ζ -Variablen) unkorreliert mit den exogenen Variablen (ξ -Variablen), d.h. $E(\xi\xi^T) = 0$.
 $\xi\xi^T$ ist eine $(n \times m)$ -Matrix, deren Elemente (i,j) die Produktvariablen $\xi_i\xi_j$ sind. $E(\xi\xi^T)$ ist die Matrix der zugehörigen elementweisen Erwartungswerte. Da bei zwei zentrierten Variablen Z_1 und Z_2 (also $E(Z_1) = E(Z_2) = 0$) nach einer Rechenregel aus Abschnitt 2.1 die Kovarianz $\text{Cov}(Z_1, Z_2)$ identisch ist mit dem Erwartungswert $E(Z_1Z_2)$ der Produktvariablen, wird mit $E(\xi\xi^T) = 0$ also vorausgesetzt, daß alle Kovarianzen zwischen ξ - und ζ -Variablen gleich Null sind. In Abschnitt 1.1 wurde der kausalitäts-theoretische Hintergrund für diese Forderung ausführlich diskutiert.
- 2) Im Meßmodell für die Y-Variablen sind die ε -Variablen unkorreliert mit den η -Variablen, d.h. $E(\eta\varepsilon^T) = 0$ und analog dazu sind im Meßmodell für die X-Variablen alle δ -Variablen unkorreliert mit allen ξ -Variablen, d.h. $E(\xi\delta^T) = 0$.
 Hier wird wie in 1) mit kausalitäts-theoretischer Begründung Unkorreliertheit der Residuen mit allen bedingenden Variablen gefordert.

- 3) Die Fehler in den Gleichungen (ζ -Variablen) sind unkorreliert mit den Fehlern in den Variablen (ε - und δ -Variablen), d.h. $E(\zeta\varepsilon^T) = E(\zeta\delta^T) = 0$.

In 2) wurde mit kausalitäts-theoretischer Begründung bereits vorausgesetzt, daß die ε -Variablen nicht mit den η -Variablen korrelieren. In 3) wird nun sinnvollerweise zusätzlich gefordert, daß sie auch nicht mit den ζ -Variablen korrelieren, also mit dem Anteil der η -Variablen, der durch die ζ -Variablen nicht erklärt werden kann.

Eine Korrelation zwischen ζ - und δ -Variablen ist mit der Grundkonzeption eines LISREL-Modells nicht vereinbar. Denn einerseits gehört eine δ -Variable zu den latenten, exogenen Variablen, wobei sie sich von den ξ -Variablen durch ihre Meßvariablen-Spezifität unterscheidet, und andererseits ist eine ζ -Variable auf latenter, endogener Ebene angesiedelt. Folglich ist die (ζ - δ)-Kovarianz vermutlich auf einen bislang nicht berücksichtigten Einfluß einer latenten, exogenen Variablen auf die betroffene (latente, endogene) η -Variable zurückzuführen. Die Beseitigung einer solchen Modellschwäche durch Aufnahme einer zusätzlichen ξ -Variablen mit Einfluß auf die betroffene η -Variable ist allerdings im allgemeinen aufgrund von Identifikationsproblemen nicht möglich.

- 4) Die ε -Variablen sind unkorreliert mit den δ -Variablen, d.h. $E(\varepsilon\delta^T) = 0$.

Gelegentlich ist diese Annahme unplausibel. Man kann sie dann umgehen durch Formulierung eines äquivalenten LISREL-Modells, das ausschließlich η - und Y -Variablen enthält. Verblüffenderweise ist eine solche Umformulierung bei *jedem* LISREL-Modell möglich (Jöreskog & Sörbom, 1989, Kap. 6).

Aus Voraussetzungen 2) und 3) folgt, daß die δ -Variablen mit den η -Variablen unkorreliert sind, d.h. $E(\delta\eta^T)$, wie folgende kurze Überlegung für δ_j und η_i zeigt:

$$\begin{aligned} \text{Cov}(\eta_i, \delta_j) &= \text{Cov}\left(\sum_{v=1}^n \gamma_{i,v} \xi_v + \zeta_i, \delta_j\right) \\ &= \text{Cov}\left(\sum_{v=1}^n \gamma_{i,v} \xi_v, \delta_j\right) + \text{Cov}(\zeta_i, \delta_j) \\ &= 0 + 0 \\ &= 0 \end{aligned}$$

Weiterhin folgt aus der Voraussetzung 3), daß die ε -Variablen auch mit den ξ -Variablen unkorreliert sind, d.h. $E(\varepsilon\xi^T) = 0$, wie folgende Überlegung für ε_i zeigt:

$$\begin{aligned} 0 &= \text{Cov}(\eta_i, \varepsilon_i) \\ &= \text{Cov}\left(\sum_{v=1}^n \gamma_{i,v} \xi_v + \zeta_i, \varepsilon_i\right) \\ &= \sum_{v=1}^n \gamma_{i,v} \text{Cov}(\xi_v, \varepsilon_i) + \text{Cov}(\zeta_i, \varepsilon_i) \\ &= \sum_{v=1}^n \gamma_{i,v} \text{Cov}(\xi_v, \varepsilon_i) \end{aligned}$$

Weil die letzte Gleichung für beliebige Ausprägungen der (noch unspezifizierten) γ -Parameter gilt, folgt aus einfachen Sätzen der Linearen Algebra, daß alle Kovarianzen $\text{Cov}(\xi_v, \varepsilon_i)$, $v = 1, \dots, n$, gleich Null sein müssen.

3.3 Die Kovarianzmatrizen der Quellvariablen

In einem wohldefinierten LISREL-Modell müssen alle Varianzen und Kovarianzen der Quellvariablen (dies sind die exogenen Variablen und die Residualvariablen) spezifiziert sein. Sie müssen also entweder durch Modellparameter dargestellt oder explizit fixiert sein. Den Grund kennen wir seit Abschnitt 2.2: Wir brauchen die Varianzen und Kovarianzen der Quellvariablen unbedingt, um die Kovarianzmatrix der manifesten Variablen durch das Modell erklären zu können. Das allgemeine LISREL-Modell enthält vier Vektoren mit Quellvariablen: ξ , ζ , ε , δ . Nach obigen Annahmen sind Variablen aus verschiedenen dieser Vektoren unkorreliert. Folglich werden zur Vervollständigung des Modells nur noch vier Parametermatrizen für die Kovarianzen innerhalb der vier Vektoren benötigt:

Tab. 3.2 Kovarianzmatrizen der Quellvariablen

Beschreibung	Bezeichnung			Ordnung
	Math.	LISREL-Ausgabe	LISREL-Subkommandos	
Kovarianzmatrix der ξ -Variablen	Φ	PHI	PH	$NK \times NK$
Kovarianzmatrix der ζ -Variablen	Ψ	PSI	PS	$NE \times NE$
Kovarianzmatrix der ε -Variablen	Θ_ε	THETA EPS	TE	$NY \times NY$
Kovarianzmatrix der δ -Variablen	Θ_δ	THETA DELTA	TD	$NX \times NX$

In Beispiel 1.3 sehen die vier Kovarianzmatrizen der Vektoren mit Quellvariablen folgendermaßen aus:

$$\Phi = \begin{bmatrix} \phi_{11} & & & \\ \phi_{21} & \phi_{22} & & \\ \phi_{31} & \phi_{32} & \phi_{33} & \\ & & & & \end{bmatrix}$$

$$\Psi = \begin{bmatrix} \psi_{11} & & & \\ \psi_{21} & \psi_{22} & & \\ & & & & \end{bmatrix}$$

$$\Theta_\varepsilon = \text{diag}(\theta_{11}^{(\varepsilon)}, \theta_{22}^{(\varepsilon)}, \theta_{33}^{(\varepsilon)}, \theta_{44}^{(\varepsilon)}) := \begin{bmatrix} \theta_{11}^{(\varepsilon)} & & & & \\ 0 & \theta_{22}^{(\varepsilon)} & & & \\ 0 & 0 & \theta_{33}^{(\varepsilon)} & & \\ 0 & 0 & 0 & \theta_{44}^{(\varepsilon)} & \end{bmatrix}$$

$$\Theta_\delta = \text{diag}(\theta_{11}^{(\delta)}, \theta_{22}^{(\delta)}, \dots, \theta_{77}^{(\delta)})$$

3.4 Die Kovarianzgleichung: $\Sigma = f(\theta)$

Das allgemeine LISREL-Modell wird also definiert durch die vier Matrizen B , Γ , Λ_Y und Λ_X mit Einflußgewichten und die vier Kovarianzmatrizen Φ , Ψ , Θ_ε , Θ_δ von Quellvariablen. Diese acht Parametermatrizen legen gemäß den obigen Modellgleichungen (3.1) und (3.2) und den zugehörigen Annahmen die $((p + q) \times (p + q))$ - Kovarianzmatrix Σ der manifesten Variablen fest. Die genaue Form der Abhängigkeit wird durch folgende Kovarianzgleichung beschrieben:

$$\Sigma = \begin{bmatrix} \Lambda_Y(I - B)^{-1}(\Gamma\Phi\Gamma^T + \Psi)(I - B^T)^{-1}\Lambda_Y^T + \Theta_\varepsilon & \Lambda_Y(I - B)^{-1}\Gamma\Phi\Lambda_X^T \\ \Lambda_X\Phi\Gamma^T(I - B^T)^{-1}\Lambda_Y^T & \Lambda_X\Phi\Lambda_X^T + \Theta_\delta \end{bmatrix} \quad (3.3)$$

Sie ist die Ausformulierung der in Abschnitt 2 als "Fundamental-Hypothese der Kovarianzstrukturanalyse" bezeichneten Gleichung $\Sigma = f(\theta)$, wobei in θ die nichtredundanten und frei schätzbaren Elemente aus den acht Parametermatrizen gesammelt sind¹.

Zur Herleitung von Gleichung (3.3) benötigt man neben den Regeln aus Abschnitt 2.1 noch einige Theoreme zum Rechnen mit Matrizen (siehe z.B. Bollen 1989).

3.5 Submodelle

Im Rahmen des allgemeinen LISREL-Modells lassen sich zahlreiche Submodelle analysieren, die nur eine Teilmenge der Zufallsvektoren Y , X , η , ξ und dementsprechend weniger Parametermatrizen enthalten. Im LISREL-Aufruf wird ein Submodell dadurch ausgewählt, daß im MO-Subkommando (siehe Abschnitt 6.6.4.1) nur für die benötigten Zufallsvektoren die Komponentenzahl spezifiziert wird, d.h. von den vier möglichen Gößenangaben (NY=, NX=, NE=, NK=) werden die irrelevanten weggelassen. Die folgende Tabelle zeigt die wesentlichen Merkmale der wichtigsten Submodelle:

Tab. 3.3 Submodelle in LISREL

Typ	Spezifiziert	Modell	Parametermatrizen	Charakterisierung
1	NX, NK	$X = \Lambda_X + \delta$	$\Lambda_X, \Phi, \Theta_\delta$	Meßmodell bzw. konfirmatorische Faktorenanalyse für die X-Variablen
2	NY, NX	$Y = BY + \Gamma X + \zeta$	B, Γ, Ψ	Strukturgleichungsmodell für direkt beobachtete Variablen
3	NY, NE, NK	$\eta = B\eta + \Gamma\xi + \zeta$ $Y = \Lambda_Y\eta + \varepsilon$	$\Lambda_Y, B, \Gamma, \Phi, \Psi, \Theta_\varepsilon$	Bei $B=0$ ist dies ein faktorenanalytisches Modell 2. Ordnung.

¹ Die griechischen Thetas tauchen leider in zwei ganz verschiedenen Bedeutungen auf: Mit " θ " wird der Vektor aller freien Modellparameter bezeichnet bzw. mit " θ_v " eine Komponente aus θ , also der v-te Modellparameter. Mit " Θ_ε " bzw. " Θ_δ " wird die Kovarianzmatrix der ε - bzw. δ -Variablen bezeichnet. Dies sind zwei spezielle Parametermatrizen. Ihre Elemente werden mit $\theta_{ij}^{(\varepsilon)}$ bzw. $\theta_{ij}^{(\delta)}$ bezeichnet.

3.6 Feste, freie und restringierte Parameter

LISREL erlaubt eine flexible Kombination konfirmatorischer und exploratorischer Elemente bei der empirischen Forschung, indem es die folgenden drei Parameter-Typen unterstützt:

i) Fixierte Parameter

Ihnen wird, entweder aufgrund theoretischer Erwägungen oder aus technischen Gründen, ein fester Wert zugewiesen. In der Kovarianzgleichung treten sie als Konstanten auf. Je mehr Parameter fixiert werden, desto mehr Freiheitsgrade verbleiben für den χ^2 -Modelltest.

Wichtig: Ist ein Parameter per Voreinstellung fixiert (siehe Abschnitt 3.7), so ist er auf $\mathbf{0}$ fixiert.

ii) Restringierte Parameter

Häufig ist für einen Parameter zwar sein Wert unbekannt, aber theoretische Gründe sprechen dafür, die Gleichheit mit anderen Parametern anzunehmen. Z.B. enthält das Modell in Beispiel 1.2 die Gleichheitsrestriktion:

$$\text{Var}(\delta_1) = \text{Var}(\delta_2) = \text{Var}(\delta_3) = \text{Var}(\delta_4),$$

bzw. in der neu eingeführten Notation:

$$\theta_{11}^{(\delta)} = \theta_{22}^{(\delta)} = \theta_{33}^{(\delta)} = \theta_{44}^{(\delta)}.$$

Für jede Gruppe von r gleichgesetzten Parametern muß nur ein Wert geschätzt werden, so daß $r-1$ Freiheitsgrade für den Modelltest gewonnen werden.

Die Testbarkeit solcher Gleichheits-Restriktionen (z.B. mit Hilfe der Modifikationsindikatoren; siehe auch Abschnitt 5.2) eröffnet sehr wertvolle Möglichkeiten, zur Formulierung und Prüfung von Hypothesen, die bei Verwendung der traditionellen regressionsanalytischen Techniken nicht zur Verfügung stehen.

Durch Einführung von latenten "Phantom-Variablen" lassen sich auch andere Parameter-Restriktionen setzen, z.B. Intervall- oder Ungleichungs-Restriktionen (siehe z.B. Hayduk 1987, Kap. 7).

iii) Freie Parameter

Kann für einen Parameter weder ein fester Wert noch die Identität mit anderen Parametern vorausgesetzt werden, so wird sein Wert frei aus den Daten geschätzt.

Der Vektor θ mit allen zu schätzenden Parametern enthält alle freien Parameter sowie einen Repräsentanten für jede Gruppe gleich gesetzter Parameter (vgl. Abschnitt 3.4). Verhindern Identifikationsprobleme eine erfolgreiche Schätzung (vgl. Abschnitt 2.3), so müssen weitere Parameter fixiert oder restringiert werden.

In der Standardausgabe von LISREL wird für jedes Element der acht Parametermatrizen protokolliert, welchen Typ es hat. Dabei wird für fixierte Elemente eine 0 geschrieben. Freie Parameter bzw. Gruppen gleich gesetzter Parameter werden durch eine fortlaufende Nummer gekennzeichnet, wobei alle Mitglieder einer Gruppe dieselbe Nummer erhalten.

Im hypothetischen Beispiel 1.3, dessen Modellgleichungen in den Abschnitten 3.2 und 3.3 vollständig wiedergegeben wurden, enthält der Vektor θ 33 Parameter:

$$\begin{aligned} \theta = & (\lambda_{21}^{(Y)}, \lambda_{42}^{(Y)}, \\ & \lambda_{21}^{(X)}, \lambda_{31}^{(X)}, \lambda_{32}^{(X)}, \lambda_{52}^{(X)}, \lambda_{73}^{(X)}, \\ & \beta_{12}, \beta_{21}, \\ & \gamma_{11}, \gamma_{12}, \gamma_{21}, \gamma_{23}, \\ & \phi_{11}, \phi_{21}, \phi_{22}, \phi_{31}, \phi_{32}, \phi_{33}, \\ & \psi_{11}, \psi_{21}, \psi_{22}, \\ & \theta_{11}^{(\epsilon)}, \theta_{22}^{(\epsilon)}, \theta_{33}^{(\epsilon)}, \theta_{44}^{(\epsilon)}, \\ & \theta_{11}^{(\delta)}, \theta_{22}^{(\delta)}, \theta_{33}^{(\delta)}, \theta_{44}^{(\delta)}, \theta_{55}^{(\delta)}, \theta_{66}^{(\delta)}, \theta_{77}^{(\delta)}) \end{aligned}$$

Mit derselben Reihenfolge wie in den Gleichungen treten die Parameter in der folgenden LISREL-Ausgabe zur Parameter-Spezifikation auf:

```

HYPOTHETISCHES MODELL AUS BEISPIEL 1.3
PARAMETER SPECIFICATIONS
  LAMBDA Y
    ETA 1      ETA 2
Y1  _____ 0 _____ 0
Y2      1      0
Y3      0      0
Y4      0      2
  LAMBDA X
    KSI 1      KSI 2      KSI 3
X1  _____ 0 _____ 0 _____ 0
X2      3      0      0
X3      4      5      0
X4      0      0      0
X5      0      6      0
X6      0      0      0
X7      0      0      7
  BETA
    ETA 1      ETA 2
ETA 1 _____ 0 _____ 8
ETA 2      9      0
  GAMMA
    KSI 1      KSI 2      KSI 3
ETA 1 _____ 10 _____ 11 _____ 0
ETA 2      12      0      13
  PHI
    KSI 1      KSI 2      KSI 3
KSI 1 _____ 14 _____ _____
KSI 2      15      16
KSI 3      17      18      19
  PSI
    ETA 1      ETA 2
ETA 1 _____ 20 _____
ETA 2      21      22
  THETA EPS
    Y1      Y2      Y3      Y4
_____ 23 _____ 24 _____ 25 _____ 26
  THETA DELTA
    X1      X2      X3      X4      X5      X6
_____ 27 _____ 28 _____ 29 _____ 30 _____ 31 _____ 32 _____
    
```

Für unser Beispiel 1.2 haben wir in Abschnitt 2.2 folgenden Parametervektor ermittelt (mit $\vartheta := \text{Var}(\delta_1) = \text{Var}(\delta_2) = \text{Var}(\delta_3) = \text{Var}(\delta_4)$):

$$\theta = (\phi_{11}, \phi_{21}, \phi_{22}, \vartheta)$$

Dementsprechend protokolliert LISREL folgende Parameter-Spezifikationen:

LATENT-STATE-MODELL DER ANGST				
PARAMETER SPECIFICATIONS				
LAMBDA X				
	KSI 1	KSI 2		
X1	0	0		
X2	0	0		
X3	0	0		
X4	0	0		
PHI				
	KSI 1	KSI 2		
KSI 1	1			
KSI 2	2	3		
THETA DELTA				
	X1	X2	X3	X4
X1	4			
X2	0	4		
X3	0	0	4	
X4	0	0	0	4

3.7 Modi und Formen der Parametermatrizen

Um den Aufwand bei der Modell-Spezifikation gering zu halten, ist für jede Parametermatrix ein voreingestellter **Modus** (frei versus fixiert) ihrer Elemente festgelegt. Jedoch kann einer Matrix auch der alternative Modus zugeordnet werden. Man wählt zwischen der Voreinstellung und der Alternative so, daß für möglichst wenige Elemente ein abweichender Modus zugewiesen werden muß. Die beiden Matrixmodi:

Tab. 3.4 Mögliche Modi der Parametermatrizen

Matrixmodus	Bedeutung
FI	Alle Elemente sind fixiert.
FR	Alle Elemente sind frei.

Ebenfalls zur Vereinfachung der Modell-Spezifikation aber auch zur Steuerung des Programmablaufes ist für jede Parametermatrix eine **Form** voreingestellt, die je nach Matrix durch unterschiedliche Alternativen ersetzt werden kann. Z.B. ist für die Matrix Θ_{ϵ} die Form DI (= diagonal) voreingestellt. Dies bedeutet, daß für die Parameter $\theta_{ij}^{(\epsilon)}$, $i \neq j$, keine Speicherplätze im Rechner reserviert werden, so daß sie nicht nur auf Null fixiert, sondern auch von der Berechnung der Modifikationsindikatoren ausgeschlossen sind. Die Parameter $\theta_{ij}^{(\epsilon)}$, $i = j$, sind im Modell enthalten, wobei ihr Modus (frei oder fixiert) durch die *Matrixform* DI *nicht* definiert wird, sondern vom voreingestellten bzw. durch den Benutzer gewählten *Matrixmodus* von Θ_{ϵ} abhängt (vgl. Abschnitt 6.6.4).

Wählt der Benutzer für die Matrix B die erlaubte Form SD, dann sind die Elemente oberhalb der Hauptdiagonalen (inklusive) von allen Berechnungen ausgeschlossen und auf Null fixiert. Die restlichen Elemente sind im Modell enthalten *und frei*. Dieser Modus wird durch die Definition der Matrixform SD explizit festgelegt.

Generell ist eine LISREL-Matrixform charakterisiert durch:

- Die Menge ihrer Elemente, die (auf Null oder Eins) fixiert und von allen Berechnungen ausgeschlossen sind
- Die eventuell vorhandene Modus-Definition für die restlichen Elemente

LISREL kennt folgende Matrixformen:

Tab. 3.5 Mögliche Formen der Parametermatrizen

Matrixform	Beschreibung	Ausgeschlossene Elemente, Fixierungswert	Modus der vorhandenen Elemente
ZE	Nullmatrix, Symbol: O	Alle Elemente, fixiert auf Null	
ID	Einheitsmatrix, Symbol: I	Alle Elemente, die Hauptdiagonal-Elemente sind auf Eins, die übrigen sind auf Null fixiert.	
IZ	Partitionierte Matrix, bestehend aus I und O: [I O]	Alle Elemente, fixiert auf Null bzw. Eins (siehe Nullmatrix, Einheitsmatrix)	
ZI	Partitionierte Matrix, bestehend aus O und I: [O I]	Alle Elemente, fixiert auf Null bzw. Eins (siehe Nullmatrix, Einheitsmatrix)	
DI	Diagonalmatrix	Die Elemente außerhalb der Hauptdiagonalen, fixiert auf Null	Gemäß Modus-Voreinstellung oder Wahl des Benutzers
SD	Subdiagonalmatrix, nur erlaubt bei Matrix B	Die Elemente oberhalb der Hauptdiagonalen (<i>inklusive</i>), fixiert auf Null	Frei
SY	Symmetrische Matrix	Keine, oberes Dreieck entfällt wegen Redundanz	Gemäß Modus-Voreinstellung oder Wahl des Benutzers
ST	Symmetrische Matrix mit Einsen in der Hauptdiagonalen (Korrelationsmatrix), nur erlaubt bei Matrix Φ	Die Elemente auf der Hauptdiagonalen, fixiert auf Eins, oberes Dreieck entfällt wegen Redundanz	Frei, Modus kann <i>nicht</i> geändert werden
FU	Volle rechteckige Matrix	Keine	Gemäß Modus-Voreinstellung oder Wahl des Benutzers

Während die ausgeschlossenen Parameter je nach Matrixform auf Null oder auf Eins fixiert sein können (siehe Tabelle 3.5), werden unabhängig von der Matrixform alle vorhandenen *und* (per Voreinstellung oder Benutzerwahl) fixierten Parameter mit Null vorbesetzt. Natürlich können alternative Werte festgelegt werden (siehe Abschnitt 6.6.4.6).

Die folgende Tabelle enthält für jede der 8 LISREL-Parametermatrixen ihren voreingestellten Modus, die voreingestellte Form, die erlaubten Formen und einige Bezeichnungsvereinbarungen. Beim praktischen Arbeiten mit LISREL werden diese Informationen oft benötigt.

Tab. 3.6 Die Eigenschaften der Parametermatrixen im Überblick

Bezeichnung			Ordnung	Erlaubte Formen	Voreingestellte Form	Voreingestellter Modus
Math.	LISREL-Ausgabe	LISREL-Syntax				
Λ_Y	LAMBDA Y	LY	$NY \times NE$	ID, IZ, ZI, DI, FU	FU	FI
Λ_X	LAMBDA X	LX	$NX \times NK$	ID, IZ, ZI, DI, FU	FU	FI
B	BETA	BE	$NE \times NE$	ZE, SD, FU	ZE	FI
Γ	GAMMA	GA	$NE \times NK$	ID, IZ, ZI, DI, FU	FU	FR
Φ	PHI	PH	$NK \times NK$	ID, DI, SY, ST	SY	FR
Ψ	PSI	PS	$NE \times NE$	ZE, DI, SY	SY	FR
Θ_ϵ	THETA EPS	TE	$NY \times NY$	ZE, DI, SY	DI	FR
Θ_δ	THETA DELTA	TD	$NX \times NX$	ZE, DI, SY	DI	FR

3.8 Maßeinheiten der latenten Variablen

Ein spezielles Identifikationsproblem im LISREL-Modell besteht bzgl. der Varianzen der latenten Variablen in η und ξ . Weil diese Variablen nicht direkt beobachtet werden, ist ihre Maßeinheit (aus statistischer Sicht: ihre Varianz) zunächst beliebig (vgl. Abschnitt 2.2). Zur Sicherung der Identifikation bieten sich folgende Techniken an:

i) Fixierung einer Ladung auf 1 für jede latente Variable

Dabei wird in Λ_Y bzw. Λ_X für jede η - bzw. ξ -Variable mindestens eine Ladung auf 1 fixiert. Damit hat die latente Variable dieselbe Maßeinheit wie die (fehlerbereinigte) zugeordnete manifeste Variable. Diese von den LISREL-Urhebern empfohlene Technik wird in den Beispielen 1.2 und 1.3 angewendet. So wird etwa in Beispiel 1.3 zur Identifikation der Varianz ϕ_{11} von ξ_1 festgelegt:

$$\lambda_{11}^{(x)} = 1$$

$$\Rightarrow \text{Var}(X_1) = \Phi_{11} + \theta_{11}^{(\delta)}$$

$$\Rightarrow \Phi_{11} = \text{Var}(X_1) - \theta_{11}^{(\delta)} = \text{Var}(X_1^T), \text{ mit } X_1^T := X_1 - \delta_1$$

Man ordnet einer latenten Variablen sinnvollerweise diejenige manifeste Variable zu, durch welche sie am besten repräsentiert wird.

ii) Standardisierung der latenten Variablen

Für die ξ -Variablen läßt sich die Unbestimmtheit eliminieren, indem ihre Varianzen in der Parametermatrix Φ auf 1 fixiert werden (z.B. durch Festlegung der Form ST für Φ). Dies ist jedoch für die Variablen in η nicht möglich, weil die Kovarianzmatrix von η keine fundamentale Parametermatrix im LISREL-Modell ist, sondern von den Parametermatrizen B , Γ , Φ und Ψ impliziert wird.

3.9 Modelle mit direkt beobachteten exogenen Variablen ($\xi \equiv X$)

Gehen die kausalen Effekte auf die η -Variablen (bzw. im Submodell 2 ohne η - und ξ -Variablen: auf die Y -Variablen) unmittelbar von den X -Variablen aus, dann benötigt LISREL schwächere Verteilungsvoraussetzungen: Statt der gemeinsamen Verteilung *aller* manifester Variablen muß lediglich die bedingte Verteilung der Y -Variablen für gegebene Werte von X betrachtet werden. Dabei spielt es keine Rolle, ob X zufällig oder im Sinne des varianzanalytischen fixed-effects-Modells auf bestimmte Werte festgelegt ist (vgl. Bollen 1989, S. 126ff).

Feste X -Werte liegen z.B. vor, wenn experimentell der Einfluß einer Behandlung auf eine abhängige Variable Y durch den Vergleich einer Behandlungsgruppe mit einer Kontrollgruppe untersucht wird, wobei X die Gruppenzugehörigkeit angibt. Hier beziehen sich die Verteilungs-Voraussetzungen nur auf die Verteilung von Y innerhalb der Gruppen, d.h. für gegebene Werte von X . Konkret werden Normalität und Varianzhomogenität gefordert. Diese Voraussetzungen sind bestens bekannt aus dem Linearen Modell, das als Spezialfall des LISREL-Modells aufgefaßt werden kann.

Wegen der großen Bedeutung dieses Spezialfalls kann er vereinfacht gewählt werden durch Angabe des Schlüsselwortes "FI" (für: "FIXED-X") im LISREL-Subkommando MO zur Modell-Spezifikation (siehe Abschnitt 6.6.4.1). Dadurch werden folgende Festlegungen veranlaßt:

$$NK=NX, \Lambda_X=I, \Theta_\delta=O, \Phi=S_X \text{ (fixiert, } O = \text{Nullmatrix, } S_X = \text{Kovarianzmatrix von } X)$$

Außerdem wird die Anzahl der Freiheitsgrade korrigiert.

Bei Verwendung von Submodell 2 (Kausalmodell für direkt beobachtete Variablen, siehe Beispiel 1.1) ist die Spezifikation FI automatisch aktiv.

Übungsaufgaben

Aufgabe 3.1 (Simulationen zum Verhalten des χ^2 -Modelltests bei wachsendem N)

In Abschnitt 2.5 wurde u.a. das Verhalten der χ^2 -Prüfstatistik zur Beurteilung der Modell-Gültigkeit bei unterschiedlichen Stichprobengrößen behandelt. Dabei stellte sich folgendes Ergebnis heraus:

- i) Bei Gültigkeit des Modells folgt die Verteilung der χ^2 -Statistik mit wachsendem N immer besser der χ^2_{df} -Verteilung, wobei df die Anzahl der Freiheitsgrade im Modell angibt. Unter der Entscheidungsregel, das Modell bei einem P-level kleiner als α zu verwerfen, konvergiert also die Wahrscheinlichkeit für ein fälschliches Verwerfen des Modells bei wachsendem N gegen den gewünschten Wert α .
- ii) Bei einem (eventuell nur minimal) falschen Modell steigt hingegen die Wahrscheinlichkeit für das (streng genommen gerechtfertigte) Verwerfen des Modells mit der Stichprobengröße an. Ein brauchbares Modell "mit kleinen Fehlern zugunsten der Einfachheit" hat also bei einer kleineren Stichprobe (z.B. $N = 100$) noch gute Chancen, "davonzukommen", während bei einer großen Stichprobe (z.B. $N = 500$) seine Schwächen gnadenlos aufgedeckt werden.

Prüfen Sie dieses Verhalten des χ^2 -Anpassungstests in einer kleinen Simulationsstudie unter Verwendung des Latent-State-Modells der Angst in Beispiel 1.2 nach.

i) Verhalten bei gültigem Modell

Legen Sie für die unbekannt Parameter des Modells konkrete Werte fest, z.B. so, daß die vom Modell implizierte Kovarianzmatrix ungefähr der in Abschnitt 2.2 angegebenen empirischen Kovarianzmatrix entspricht. Ziehen Sie mit Hilfe eines SPSS-Eingabeprogramms (Kommandos INPUT PROGRAM ... END INPUT PROGRAM) Stichproben verschiedenen Umfangs (z.B. $N = 100, 200, 500, 1000, 10000$) aus der gedachten Population, in der das Latent-State-Modell exakt gilt, und berechnen Sie für diese Datensätze und das Latent-State-Modell den χ^2 -Anpassungstest mit LISREL. Über die Identifikation der Parameter brauchen Sie sich keine Gedanken zu machen, weil wir den Nachweis schon in Abschnitt 2.3 erbracht haben.

Natürlich kann von Ihnen momentan noch nicht erwartet werden, selbständig einen LISREL-Aufruf zu verfassen. Sie können jedoch anhand des folgenden, ausführlich erläuterten Lösungsvorschlags erste Gehversuche mit LISREL unternehmen:

```

title Latent-State-Modell Der Angst; MC-Stud. zum CHI**2-Test
set mxloops=1000
input program
+ loop #i=1 to 1000
+ compute #ksi=normal(3.5)
+ compute ksi1=#ksi + normal(3.5)
+ compute ksi2=#ksi + normal(3.5)
+ compute X1=ksi1 + normal(1.7)
+ compute X2=ksi1 + normal(1.7)
+ compute X3=ksi2 + normal(1.7)
+ compute X4=ksi2 + normal(1.7)
+ end case
+ end loop
+ end file
end input program

lisrel
/"Latent-State-Modell der Angst; simulierte Daten, korrektes Modell"
/da ni=6
/se
/x1 x2 x3 x4 /
/mo nx=4 nk=2
/va 1.0 lx(1,1) lx(2,1) lx(3,2) lx(4,2)
/eq td(1,1) td(2,2) td(3,3) td(4,4)
/ou

```

Erläuterungen:

- Das SPSS-Programm beginnt mit einer Titel-Zeile und der Festlegung, daß Schleifen maximal 1000 Durchläufe haben dürfen.
- Dann folgt ein Eingabeprogramm, das in einer Schleife 1000 Fälle produziert. Für jeden Fall werden durch die COMPUTE-Kommandos nach Vorschrift des Latent-State-Modells unter Verwendung des Generators für normalverteilte Zufallszahlen neben der Hilfsvariablen #KSI die permanenten Variablen KSI1, KSI2, X1, X2, X3 und X4 berechnet. Jeder so produzierte Fall kann als zufällige Ziehung aus einer Population, in der das Latent-State-Modell der Angst perfekt gilt, betrachtet werden.
- Nun wird LISREL wie eine SPSS-Prozedur aufgerufen, wobei der Kommandonamen "LISREL" gemäß der SPSS-Syntax in der ersten Spalte beginnen muß.
- Anschließend wird unter Anwendung der SPSS-Syntax für Fortsetzungszeilen (mindestens eine Spalte einrücken!) eine LISREL-interne Titelzeile vereinbart.
- Im DA(taparameters)-Subkommando wird mitgeteilt, daß die Eingabedatei, in unserem Fall also die aktive SPSS-Systemdatei, 6 Variablen enthält (Hilfsvariablen zählen nicht mit).
- Da nur die 4 manifesten Variablen mit LISREL analysiert werden sollen, müssen diese im Rahmen eines SE(lect)-Subkommandos explizit aufgelistet werden. Am Ende dieser Liste muß außerdem ein Schrägstrich stehen.
- Im MO(dellparameters)-Subkommando wird mitgeteilt, daß unser Modell 4 X- und 2 ξ -Variablen enthält.
- Mit dem VA(lue)-Subkommando werden gemäß der Diskussion in Abschnitt 2.2 alle Ladungen auf 1 gesetzt. Damit wird die Gleichheitsrestriktion $\lambda_{11} = \lambda_{21} = \lambda_{32} = \lambda_{42}$ ausgedrückt und gleichzeitig für die Identifikation der Varianzen der ξ -Variablen gesorgt.
- Mit dem EQ(ual)-Subkommando wird die Gleichheitsrestriktion $\text{Var}(\delta_1) = \text{Var}(\delta_2) = \text{Var}(\delta_3) = \text{Var}(\delta_4)$ deklariert. "TD" steht für "THETA DELTA" bzw. " Θ_δ " und dies ist die Kovarianzmatrix der δ -Variablen.
- Das OU(tputparameters)-Subkommando ist obligatorisch und erlaubt die Anforderung optionaler Ausgaben, wovon im Beispiel kein Gebrauch gemacht wird.

ii) Verhalten bei ungültigem Modell

Ändern Sie nun das Eingabeprogramm und damit die wahren Populationsverhältnisse, indem Sie eine schwache, aber modellwidrige Kovarianz zwischen den Residuen δ_2 und δ_4 einbauen. Das neue Eingabeprogramm könnte z.B. so aussehen:

```
input program
+ loop #i=1 to 1000
+ compute #ksi=normal(3.5)
+ compute ksi1=#ksi + normal(3.5)
+ compute ksi2=#ksi + normal(3.5)
+ compute ksi3=normal(1)
+ compute X1=ksi1 + normal(1.7)
+ compute X2=ksi1 + ksi3 + normal(1.4)
+ compute X3=ksi2 + normal(1.7)
+ compute X4=ksi2 + ksi3 + normal(1.4)
+ end case
+ end loop
+ end file
end input program
```

Im LISREL-Aufruf soll natürlich weiterhin das ursprüngliche, also nun nicht mehr perfekt passende Modell deklariert werden. Da KSI3 eine permanente Variable ist, muß allerdings das DA-Subkommando modifiziert werden.

4 PRELIS: Der LISREL-Preprozessor für problematische Daten

Vor einer LISREL-Analyse muß geklärt werden, ob ihre Voraussetzungen hinsichtlich Skalenqualität und Verteilung der manifesten Variablen gegeben sind. Dabei kann PRELIS mit seinen Fähigkeiten zur Daten-Exploration wertvolle Hilfe leisten.

Während frühere LISREL-Versionen im Prinzip nur für die Analyse von Kovarianzmatrizen aus intervall-skalierten, multivariat normalverteilten manifesten Variablen geeignet waren, kann die aktuelle Version auch bei problematischen Daten eingesetzt werden, z.B. bei schiefer Verteilung, ordinalem Meßniveau oder bei Zensierungen (Boden- oder Deckeneffekten). Solche Probleme werden durch die Wahl spezieller Assoziationsmaße (z.B. polychorische Korrelationen) und durch den Einsatz verteilungsfreier Schätzmethoden gelöst, wobei wiederum PRELIS entscheidende Vorleistungen erbringt.

In vielen Situationen wird man also vor dem LISREL-Aufruf die SPSS-Prozedur PRELIS einsetzen, um die Daten zu explorieren oder um aus problematischen Daten geeignete Eingabematrizen für LISREL zu erstellen.

4.1 Variablentypen und Assoziationsmaße

4.1.1 Variablentypen

PRELIS kann drei Typen von Variablen verarbeiten: kontinuierliche, ordinale und zensierte.

4.1.1.1 Kontinuierliche Variablen

Unter diesen Begriff fallen intervall- oder rationalskalierte Variablen. Für sie werden die üblichen uni- bzw. bivariaten parametrischen Statistiken berechnet.

4.1.1.2 Ordinale Variablen

Bei einer ordinalen Variablen X dürfen lediglich Größer/Kleiner - Relationen zwischen den Meßwerten interpretiert werden. Sie wird als das Ergebnis der grobschlächtigen Messung einer latenten kontinuierlichen Variablen ξ aufgefaßt, für die Normalverteilung mit Erwartungswert Null und Varianz Eins angenommen wird. Durch den Meßvorgang werden jeweils Intervalle unterschiedlicher Breite auf einen Wert abgebildet. Die folgende Abbildung 4.1 illustriert das Zustandekommen ordinaler Variablen durch vergrößerndes Messen.

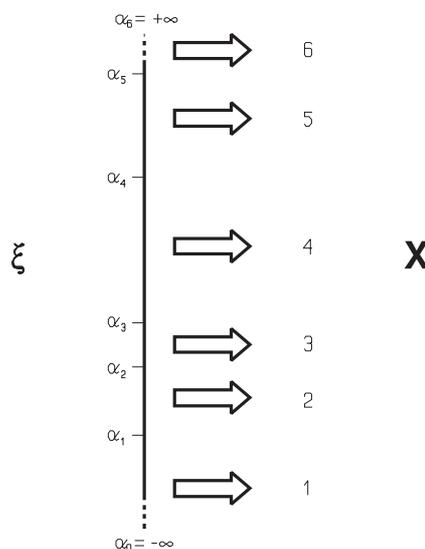


Abb. 4.1 Entstehung eines ordinalen X -Indikators aus einer kontinuierlichen ξ -Variablen

Die manifeste Variable X nimmt in diesem Beispiel also genau dann den Wert i an, wenn gilt: $\xi \in (\alpha_{i-1}, \alpha_i]$. Bei den Werten von X ist lediglich die Ordnung bedeutsam, Intervalle dürfen nicht interpretiert werden. Daher wird X als "ordinal" bezeichnet.

PRELIS kann für jedes Paar (X_1, X_2) von ordinalen Variablen die Annahme der bivariaten Normalverteilung der zugrundeliegenden latenten Variablen (ξ_1, ξ_2) testen.

Als sinnvollstes Assoziationsmaß für ordinale Variablen hat sich die polychorische Korrelation erwiesen. Wünscht ein Benutzer Produkt-Moment-Korrelationen so ersetzt PRELIS vor der Berechnung die Rohwerte durch sogenannte Normal-Scores (siehe unten).

4.1.1.3 Zensierte Variablen

Eine manifeste Variable X wird als "**unten-zensiert**" bezeichnet, wenn sie nach folgender Regel aus einer latenten Variablen ξ entstanden ist:

$$X = \begin{cases} A, & \text{wenn } \xi \leq A \\ \xi, & \text{wenn } \xi > A \end{cases}$$

Für ξ wird Normalverteilung angenommen, wobei PRELIS den Erwartungswert und die Varianz nach dem Maximum Likelihood - Prinzip aus den Daten schätzen kann.

Völlig analog wird eine manifeste Variable X als "**oben-zensiert**" bezeichnet, wenn für eine normalverteilte latente Variable ξ gilt:

$$X = \begin{cases} \xi, & \text{wenn } \xi < B \\ B, & \text{wenn } \xi \geq B \end{cases}$$

Schließlich kann PRELIS auch den Fall einer beidseitigen Zensierung behandeln.

PRELIS ersetzt bei zensierten Variablen den unteren und/oder den oberen Schwellenwert durch einen sogenannten Normal-Score (siehe unten) und behandelt die so modifizierte Variable als kontinuierlich.

4.1.2 Assoziationsmaße

4.1.2.1 Normal-Scores

Oft kann man zurecht eine ordinale oder zensierte Variable X als vergrößerte bzw. beschädigte Variante einer zugrundeliegenden, normalverteilten Variablen X_T auffassen. Zur Schätzung der Assoziationen von X_T mit anderen Variablen verwendet man dann gelegentlich anstelle der vorhandenen X -Ausprägungen die geschätzten Werte der zugrundeliegenden Variablen X_T . Weil bei der Berechnung die Normalverteilung von X_T entscheidend eingeht, werden die rekonstruierten Werte als "Normal-Scores" bezeichnet.

Zur Berechnung der Normal-Scores für eine *ordinale* Variable X müssen zunächst die Schwellenwerte $\alpha_1 < \alpha_2 < \dots < \alpha_{\kappa-1}$ geschätzt werden, wobei κ die Anzahl unterschiedlicher Ausprägungen von X in der Stichprobe ist. Wie die Abb. 4.1 zeigt, liegen die beiden äußeren "Schwellenwerte" fest: $\alpha_0 = -\infty$, $\alpha_\kappa = \infty$. Ist die j -te Ausprägungsstufe von X in der Stichprobe n_j -mal beobachtet worden, $j=1, \dots, \kappa$, so wird α_i geschätzt durch:

$$\hat{\alpha}_i := \Phi^{-1} \left(\sum_{j=1}^i \frac{n_j}{N} \right) \quad i = 1, 2, 3, \dots, \kappa-1$$

Wie das Auftreten der inversen Verteilungsfunktion Φ^{-1} der Standardnormalverteilung zeigt, erfolgt die Schätzung der $\kappa-1$ Schwellenwerte unter der Annahme, daß die zugrundeliegende Variablen ξ standardnormalverteilt ist.

Der Normal-Score z_i zur i -ten Ausprägungsstufe von X ist definiert als der mittlere ξ -Wert im Intervall $(\alpha_{i-1}, \alpha_i]$:

$$z_i := \int_{\alpha_{i-1}}^{\alpha_i} u \phi(u) du = \frac{\phi(\alpha_{i-1}) - \phi(\alpha_i)}{\Phi(\alpha_i) - \Phi(\alpha_{i-1})} \quad (4.1)$$

Dabei ist ϕ die Dichte der Standardnormalverteilung. z_i kann geschätzt werden durch:

$$\hat{z}_i := \frac{N}{n_i} [\phi(\hat{\alpha}_{i-1}) - \phi(\hat{\alpha}_i)]$$

Das arithmetische Mittel der Normal-Scores zu allen Fällen in der Stichprobe, also das mit $\frac{n_j}{N}$ gewichtete

Mittel der κ Werte \hat{z}_j , ist Null. Dies ergibt sich unmittelbar aus der Definition der \hat{z}_j : die Bildung des gewichteten Mittels führt zu einer Teleskopsumme, deren Glieder sich zu Null aufheben.

Damit haben die Normal-Scores denselben Mittelwert wie z-standardisierte Werte, denen sie ähneln, weil der unbekannte Erwartungswert der zugrundeliegenden Variable X_T auf Null gesetzt wird und analog die unbekannte Varianz auf Eins. Allerdings ist die Varianz der z_j bzw. der \hat{z}_j im allgemeinen kleiner als Eins, weil im Vergleich zu X_T der Varianzanteil "innerhalb der Kategorien" fehlt.

Bei einer unterhalb des Wertes A zensierten Variablen X ist der Normal-Score z_A für das Intervall $(-\infty, A]$ durch den Mittelwert der zugrundeliegenden Variablen ξ in diesem Intervall definiert. Die Schätzformel ergibt sich aus (4.1) mit $-\infty$ in der Rolle von α_{i-1} und $\frac{(A - \hat{\mu})}{\hat{\sigma}}$ in der Rolle von α_i , wobei $\hat{\mu}$ und $\hat{\sigma}$ die Maximum-Likelihood-Schätzungen von μ und σ sind:

$$\hat{z}_A := \hat{\mu} - \frac{\phi\left(\frac{A - \hat{\mu}}{\hat{\sigma}}\right)}{\Phi\left(\frac{A - \hat{\mu}}{\hat{\sigma}}\right)} \hat{\sigma}$$

\hat{z}_A besitzt dieselbe Metrik wie die übrigen (Roh-)Werte von X .

Bei einer Zensierung oberhalb von B ist der Normal-Score zum Intervall $[B, \infty)$ analog definiert.

4.1.2.2 Optimal-Scores

Wird für zwei ordinale Variablen eine Korrelation vom Typ OPTIMAL (siehe unten) angefordert, so werden die Rohwerte beider Variablen so zu Scores transformiert, daß die Produkt-Moment-Korrelation zwischen den Score-Variablen optimal wird, unter der Nebenbedingung, daß die Scores einen Mittelwert von Null und eine Varianz von Eins haben.

4.1.2.3 Korrelationsmaße

PRELIS kann verschiedene Korrelationsmaße berechnen, die über die Schlüsselwörter CORRELATION, OPTIMAL und POLYCHOR des TYPE-Subkommandos ausgewählt werden können, wobei außerdem die (voreingestellten oder zugewiesenen) Skalentypen der Variablen relevant sind (vgl. Abschnitt 6.5.5). Zur übersichtlichen Beschreibung der Optionen sollen nach den beteiligten Skalentypen drei Fälle unterschieden werden. Dabei ist noch zu beachten, daß zensierte Variablen nach Ersetzung der unteren und/oder oberen Schwelle durch die zugehörigen Normal-Scores wie kontinuierliche behandelt werden.

i) Beide Variablen sind kontinuierlich

In diesem Fall berechnet PRELIS bei jeder beliebigen Angabe im TYPE-Subkommando die Produktmoment-Korrelation aus den Rohwerten.

ii) Beide Variablen sind ordinal

Bei TYPE=CORRELATION berechnet PRELIS die Produktmoment-Korrelation aus den Normal-Scores. Bei TYPE=OPTIMAL wird die Produktmoment-Korrelation aus den Optimal-Scores berechnet. Bei TYPE=POLYCHOR wird keine Produktmoment-Korrelation aus irgendwelchen Scores berechnet, sondern PRELIS schätzt die Korrelation zwischen den zugrundeliegenden, normalverteilten Variablen ξ_1 und ξ_2 nach dem Maximum-Likelihood-Prinzip direkt aus der zweidimensionalen Kontingenztabelle zu X_1 und X_2 .

iii) Variable X_1 ist kontinuierlich, Variable X_2 ist ordinal

Bei TYPE=CORRELATION wird die Produktmoment-Korrelation zwischen den Rohwerten der kontinuierlichen Variablen X_1 und den Normal-Scores der ordinalen Variablen X_2 berechnet. Bei den Typen OPTIMAL und POLYCHOR wird eine polyseriale Korrelation berechnet. Dies ist die Korrelation zwischen der kontinuierlichen Variablen X_1 und der latenten Variablen ξ_2 , die der ordinalen Variablen X_2 zugrundeliegt. Dabei wird für X_1 und ξ_2 eine bivariate Normalverteilung angenommen. Bei TYPE=OPTIMAL wird ein einfaches, konsistentes Schätzverfahren benutzt, bei TYPE=POLYCHOR wird ein Maximum-Likelihood-Schätzer berechnet.

Simulationsergebnisse zur Güte verschiedener Korrelationsmaße für ordinale Variablen:

Jöreskog & Sörbom (1988) haben in zwei Monte-Carlo-Studien folgende Korrelationsmaße bei ordinalen Variablen verglichen:

- Produkt-Moment-Korrelation aus den Rohdaten (PM-R),
- Produkt-Moment-Korrelation aus den normal scores (PM-N),
- Produkt-Moment-Korrelation aus den optimal scores (PM-O),
- Spearman-Rangkorrelation (SR),
- Kendall-Rangkorrelation (KR),
- Polychorische bzw. polyseriale Korrelation (POLYCHOR)

Dabei ergab sich:

- Alle Korrelationen waren betragsmäßig nach unten verfälscht. Diese Verzerrung war am schwächsten beim Typ POLYCHOR.
- POLYCHOR, PM-N und PM-O waren relativ unabhängig von der Randverteilung der Variablen.
- POLYCHOR war durchweg der beste Schätzer hinsichtlich Erwartungstreue und Präzision. Er schien als einziger konsistent zu sein.

Wir werden in Aufgabe 5.1 eine eigene kleine Monte-Carlo-Studie zu diesem Thema durchführen.

4.1.2.4 Andere Assoziationsmaße

Außer den verschiedenen Korrelationsmatrizen kann man im PRELIS-Subkommando TYPE auch Kovarianz- und Momentenmatrizen anfordern. Bei ihrer Berechnung werden ordinale Variablen grundsätzlich durch die zugehörigen Normal-Scores ersetzt.

Varianzen und *Kovarianzen* werden nach den üblichen Formeln erwartungstreu geschätzt.

Die *Momentenmatrix* enthält die mittleren Produkte und Kreuzprodukte der Variablen. Ihr typisches Element m_{ij} ist definiert durch:

$$m_{ij} := \frac{1}{n_{ij}} \sum_{v=1}^{n_{ij}} z_{vi} z_{vj},$$

wobei z_{vi} der Wert des v -ten Falles bei der Variablen Z_i ist und n_{ij} die Anzahl der einbezogenen Fälle, die i.a. vom Kriterium für den Ausschluß fehlender Werte (fallweise versus paarweise) abhängt.

4.2 Asymptotisch verteilungsfreie Schätzer

Die in Abschnitt 2.4 vorgestellten Schätzmethoden ULS, GLS, und ML setzen voraus, daß die beobachteten X - und Y -Variablen zumindest approximativ Intervallskalen-Qualität besitzen. Die inferenzstatistischen Bewertungen der Schätzergebnisse (z.B. Standardfehler und χ^2 -Modelltest) basieren auf der weitergehenden Annahme, daß die beobachteten Variablen gemeinsam multivariat normalverteilt sind, wobei leichte Verletzungen dieser Annahme toleriert werden können.

In Abschnitt 2.4.3 wurden auch schon zwei Situationen genannt, in denen man nicht auf die Robustheit der klassischen Normalverteilungs-Methoden vertrauen sollte:

- Die Verteilung der manifesten Variablen weicht erheblich von der Normalverteilung ab.
- Einige der manifesten Variablen sind ordinalskaliert, so daß statt der Kovarianzmatrix eine Matrix mit polychorischen und polyserialen Korrelationen analysiert werden sollte. Diese Situation dürfte z.B. in den Sozial- und Wirtschaftswissenschaften sehr häufig auftreten.

Fortschritte der mathematischen Statistik in den letzten zehn Jahren (Browne 1984, Jöreskog & Sörbom 1989) führten zur Entwicklung von Schätzmethoden, die für viele vorher problematische Situationen befriedigende Lösungen bieten, insbesondere hinsichtlich der inferenzstatistischen Ergebnisse. Ihr wesentlicher Vorteil ist die asymptotische Verteilungsfreiheit, d.h. aus einer hinreichend großen Stichprobe erhält man Parameterschätzer und Inferenzstatistiken, die von der Verteilung der manifesten Variablen praktisch nicht mehr abhängen. Damit kann die etwas restriktive klassische LISREL-Annahme

multivariater Normalverteilung der manifesten Variablen überwunden werden. LISREL 7 bietet eine asymptotisch verteilungsfreie Schätzmethode unter der Bezeichnung "WLS" (Generally Weighted Least Squares) an und ist dabei auf eine Vorleistung von PRELIS angewiesen, weshalb die WLS-Methode erst in diesem Abschnitt besprochen wird.

Die eben angesprochene PRELIS-Vorleistung besteht in der Erstellung einer geschätzten Kovarianzmatrix der von LISREL zu analysierenden Stichproben-Varianzen bzw. -Kovarianzen bzw. -Korrelationen. Es soll nun erklärt werden, welche Einträge diese Kovarianzmatrix "zweiter Ordnung" enthält. Dazu greifen wir auf unser Beispiel 1.2 (Latent-State-Model der Angst) mit den manifesten Variablen X_1 bis X_4 zurück, deren Populations-Kovarianzmatrix wie üblich mit " Σ " bezeichnet wird. Führt man eine Studie mit einer Zufalls-Stichprobe der Größe N durch, so erhält man eine Stichproben-Kovarianzmatrix, die als Ergebnis eines Zufallsexperimentes i.a. von der Populations-Kovarianzmatrix Σ verschieden ist. Bei einer Wiederholung der Studie wird sich höchst wahrscheinlich eine andere Stichproben-Kovarianzmatrix ergeben. Damit ist es offenbar sinnvoll, z.B. die aus einer Stichprobe der Größe N zu ermittelnde Stichproben-Varianz von X_1 als Zufallsvariable mit der Bezeichnung $S_{11}^{(N)}$ zu betrachten, ebenso die bei Stichprobenumfang N zu ermittelte Stichproben-Kovarianz von X_1 und X_2 als Zufallsvariable mit der Bezeichnung $S_{21}^{(N)}$ u.s.w. Zur Vereinfachung der Ausdrucksweise kann man schließlich die Zufallsvariablen $S_{ij}^{(N)}$ zur Zufallsmatrix $S^{(N)}$ anordnen:

$$S^{(N)} := \begin{bmatrix} S_{11}^{(N)} & & & \\ S_{21}^{(N)} & S_{22}^{(N)} & & \\ S_{31}^{(N)} & S_{32}^{(N)} & S_{33}^{(N)} & \\ S_{41}^{(N)} & S_{42}^{(N)} & S_{43}^{(N)} & S_{44}^{(N)} \end{bmatrix}$$

Aus Sicht der Wahrscheinlichkeitstheorie sind $S_{11}^{(N)}$ und $S_{21}^{(N)}$ ganz normale Zufallsvariablen, wie z.B. auch X_1, X_2 , etc. Sie gehören lediglich zu einem etwas komplizierterem Zufallsexperiment, so daß die Gewinnung einer Realisation etwas aufwendiger ist: Während z.B. eine Realisation von X_1 durch zufällige Ziehung eines einzigen Falles aus der Population gewonnen wird, muß man N Fälle zufällig ziehen und deren Stichprobenvarianz für X_1 berechnen, um eine Realisation von $S_{11}^{(N)}$ zu erhalten.

Als wohldefinierte Zufallsvariablen haben die Elemente von $S^{(N)}$ Varianzen und Kovarianzen, die sich wie üblich zu einer Kovarianzmatrix anordnen lassen. Dies ist also die Kovarianzmatrix der Varianzen und Kovarianzen aus einer Stichprobe der Größe N und soll mit $Cov(S^{(N)})$ bezeichnet werden. In unserem Beispiel mit vier manifesten Variablen X_1 bis X_4 haben wir es mit $\frac{1}{2}4(4+1) = 10$ Zufallsvariablen $S_{ij}^{(N)}$ zu tun. Somit ergibt sich folgende 10×10 Kovarianzmatrix:

$$Cov(S^{(N)}) := \begin{bmatrix} Var(S_{11}^{(N)}) & & & & & & & & & \\ Cov(S_{21}^{(N)}, S_{11}^{(N)}) & Var(S_{21}^{(N)}) & & & & & & & & \\ Cov(S_{22}^{(N)}, S_{11}^{(N)}) & Cov(S_{22}^{(N)}, S_{21}^{(N)}) & Var(S_{22}^{(N)}) & & & & & & & \\ \vdots & \\ Cov(S_{44}^{(N)}, S_{11}^{(N)}) & Cov(S_{44}^{(N)}, S_{21}^{(N)}) & Cov(S_{44}^{(N)}, S_{22}^{(N)}) & Cov(S_{44}^{(N)}, S_{31}^{(N)}) & Cov(S_{44}^{(N)}, S_{32}^{(N)}) & Cov(S_{44}^{(N)}, S_{33}^{(N)}) & Cov(S_{44}^{(N)}, S_{41}^{(N)}) & Cov(S_{44}^{(N)}, S_{42}^{(N)}) & Cov(S_{44}^{(N)}, S_{43}^{(N)}) & Var(S_{44}^{(N)}) \end{bmatrix}$$

Eine Schätzung für $Cov(S^{(N)})$ läßt sich glücklicherweise schon aus einer einzigen Stichprobe gewinnen. Sie wird in der LISREL-Literatur üblicherweise als "geschätzte asymptotische Kovarianzmatrix" bezeichnet, weil zur Berechnung der Schätzer eine Darstellung von $Cov(S^{(N)})$ verwendet wird, die nur asymptotisch

tisch gilt, d.h. für $n \rightarrow \infty$. Gelegentlich werden wir für "asymptotische Kovarianzmatrix" die Abkürzung "ACov-Matrix" verwenden.

Nachdem wir schon einen erheblichen Aufwand betrieben haben, um die Matrix $\text{Cov}(S^{(N)})$ zu erläutern, soll auch noch gezeigt werden, wie sie von LISREL zur WLS-Parameterschätzung eingesetzt wird. Bei der WLS-Methode kommt wiederum das in Abschnitt 2.4 vorgestellte Prinzip der Parameterschätzung durch Minimierung einer Fit-Funktion zum Einsatz. Das Spezifikum der WLS-Methode liegt also in der Wahl der Fit-Funktion F_{WLS} . Wie bei der Fit-Funktion der GLS-Methode wird eine **gewichtete** Diskrepanz (daher der Name "Weighted Least-Squares") zwischen der Stichproben-Kovarianz- bzw. -Korrelations-Matrix S und der von einem Parametervektor θ^* aufgrund des Modells implizierten Matrix Σ^* minimiert:

$$F_{\text{WLS}}(S, \Sigma^*) := (s - \sigma^*)^T W^{-1} (s - \sigma^*) \quad (4.2)$$

Dabei ist $s := (s_{11}, s_{21}, s_{22}, s_{31}, \dots, s_{kk})$ ein **Vektor** mit den hintereinander angeordneten Elementen der unteren Dreiecksmatrix von S und $\sigma^* := (\sigma_{11}^*, \sigma_{21}^*, \sigma_{22}^*, \sigma_{31}^*, \dots, \sigma_{kk}^*)$ ein analog gebildeter Vektor mit den nichtredundanten Elementen von Σ^* . Für die $k \times k$ - Gewichtsmatrix W wird zunächst nur verlangt, daß sie positiv definit ist.

Der durch Minimierung von F_{WLS} bestimmte Schätzer ist bei Modellgültigkeit unter sehr allgemeinen Verteilungsvoraussetzungen **konsistent**.

Wie Browne (1984) zeigen konnte, erhält man optimale Ergebnisse durch Verwendung einer konsistenten Schätzung von $\text{Cov}(S^{(N)})$ in der Rolle von W . Dieses spezielle Verfahren ist **asymptotisch effizient** in der Klasse aller Verfahren nach dem Schema (4.2).

Ab jetzt soll der Name "WLS" für diese spezielle Wahl von W reserviert bleiben.

Ein weiterer Vorzug des (speziellen) WLS-Verfahrens liegt darin, daß **asymptotische Standardfehler** der Parameterschätzungen sowie ein χ^2 -**Modelltest** verfügbar sind.

Schließlich glänzt die WLS-Methode durch ihre **asymptotische Verteilungsfreiheit**: Die asymptotische Verteilung der Schätzer hängt nicht von der Verteilung der manifesten Variablen ab, sofern einige liberale Regularitätsbedingungen erfüllt sind. Der WLS-Schätzer wird daher auch als ADF-Schätzer bezeichnet (**A**symptotically **D**istribution **F**ree). Analoges gilt für die asymptotische Verteilung der χ^2 -Prüfgröße.

Es soll noch erwähnt werden, daß F_{GLS} ebenfalls von der Form (4.2) ist, wobei in die Gewichtsmatrix eine aus der Normalverteilungsannahme abgeleitete Schätzung für $\text{Cov}(S^{(N)})$ eingeht.

Die ML-Methode ist äquivalent zur iterativen Minimierung von Funktionen nach dem Schema (4.2), wobei jedoch in jeder Iteration eine aktualisierte Schätzung von $\text{Cov}(S^{(N)})$ verwendet wird. Diese Schätzung wird wie bei F_{GLS} unter der Normalverteilungsannahme ermittelt.

Browne (1984) hatte die Theorie der ADF-Schätzer nur für die Analyse von Kovarianzmatrizen entwickelt. Von Jöreskog & Sörbom wurde sie im Rahmen der Arbeit an PRELIS erweitert für die Analyse von Korrelationsmatrizen. Insbesondere kann PRELIS die Varianzen und Kovarianzen von geschätzten polychorischen und polyserialen Korrelationen berechnen, so daß eine statistisch befriedigende Analyse von Strukturgleichungsmodellen anhand ordinaler Daten möglich ist.

Wie schon zu Beginn dieses Abschnitts erwähnt, ist die WLS-Schätzmethode vor allem in den beiden folgenden Situationen indiziert:

- Die Verteilung der manifesten Variablen weicht erheblich von der Normalverteilung ab.
- Einige der manifesten Variablen sind ordinalskaliert, so daß statt der Kovarianzmatrix eine Matrix mit polychorischen und polyserialen Korrelationen analysiert werden sollte.

Die WLS-Methode ist trotz ihrer theoretischen Vorzüge keinesfalls in jeder Situation allen anderen Schätzmethode überlegen. Das liegt vor allem an folgenden praktischen Problemen:

- Bei k manifesten Variablen enthält die Matrix S der Stichproben-Kovarianzmatrix $\frac{1}{2}k(k+1) =: u$ nicht-redundante Elemente. Die zugehörige Acov-Matrix ist daher von der Ordnung $u \times u$ und hat $\frac{1}{2}u(u+1)$ nicht-redundante Elemente. Bei $k = 25$ müssen z.B. 52975 Schätzungen ermittelt werden, wozu in erheblichem Umfang Computer-Speicher und -Rechenzeit aufgewendet werden müssen. Eine Stichproben-Korrelationsmatrix enthält nur $\frac{1}{2}k(k-1)$ nicht-redundante Elemente, so daß die zugehörige Acov-Matrix entsprechend kleiner ist.
- Eine zuverlässige Schätzung der Acov-Matrix erfordert relativ große Stichproben. Per Voreinstellung setzt PRELIS folgende Stichprobengröße voraus (vgl. Beschreibung des CRITERIA-Subkommandos in Abschnitt 6.5.9):
 - bei $k < 12$ Variablen: 200
 - bei $k \geq 12$ Variablen: $1.5k(k+1)$.
 Bei $k = 25$ ist also z.B. ein minimales N von 975 erforderlich.
- Die WLS-Methode kann bei paarweisem Ausschluß fehlender Werte nicht verwendet werden.

Vor allem zur Bewährung der WLS-Methode bei der Analyse ordinaler Daten, die z.B. in den Wirtschafts- und Sozialwissenschaften häufig auftreten, wurden mehrere Simulationsstudien durchgeführt. Dabei schnitt die WLS-Methode zwar in der Regel besser ab als alternative, einfachere Schätzverfahren, jedoch waren die Unterschiede nicht dramatisch (vgl. Abschnitt 4.3; Kühnel, 1993).

Beim Einsatz der WLS-Schätzmethode geht man folgendermaßen vor:

- Nach dem Einlesen der Daten läßt man mit einem PRELIS-Kommando daraus neben der Stichproben-Kovarianz- bzw. -Korrelationsmatrix S der manifesten Variablen die geschätzte asymptotische Acov-Matrix der Stichproben-Kovarianzen bzw. -Korrelationen berechnen. Erstere wird dabei in eine SPSS-Matrix-Systemdatei geschrieben, letztere in eine Textdatei.
- Anschließend läßt man unter Verwendung der beiden Dateien die LISREL-Analyse mit der WLS-Schätzmethode durchführen.

Ein kommentiertes Beispiel für dieses Vorgehen folgt im nächsten Abschnitt.

Für den Fall, daß die Anforderungen der WLS-Methode an Rechenzeit und Computer-Speicher zu groß werden, steht in LISREL 7 die weniger aufwendige **DWLS-Methode** zur Verfügung (**D**iaagonally **W**eighted **L**east **S**quares). Hier werden statt der gesamten Matrix $\text{Cov}(S^{(N)})$ nur die Varianzen der Elemente von S berücksichtigt. Die zugehörige Fitfunktion F_{DWLS} ist wieder von der allgemeinen Form (4.2), wobei als Gewichtsmatrix W eine Diagonalmatrix mit der Hauptdiagonalen aus $\text{Cov}(S^{(N)})$ verwendet wird. Wenn mit w_{ij} die geschätzte asymptotische Varianz von S_{ij} bezeichnet wird, so kann man F_{DWLS} wie folgt schreiben:

$$F_{\text{WLS}}(S, \Sigma^*) = \sum_{i=1}^k \sum_{j=1}^k \frac{1}{w_{ij}} (s_{ij} - \sigma_{ij}^*)^2$$

F_{DWLS} hat zwar nicht mehr die Optimalitätseigenschaften von F_{WLS} , ist aber als asymptotisch verteilungsfreie Methode bei problematischen Daten und hinreichend großer Stichprobe sicher eine bessere Wahl als F_{ULS} .

Für die praktische Anwendung der DWLS-Methode gilt analog die obige Beschreibung zur WLS-Methode.

4.3 Simulationsstudie: Verhalten verschiedener Schätzmethoden bei ordinalen Indikatoren

In der sozialwissenschaftlichen Forschungspraxis kommt es sehr häufig vor, daß die verfügbaren Indikatoren teilweise ordinalen Charakter im Sinne von Abschnitt 4.1.1.2 haben. Vielfach ist ferner die Annahme gerechtfertigt, daß die ordinalen Indikatoren durch Vergrößerung von normalverteilten Variablen entstanden sind. In dieser Situation kann aus den verfügbaren Indikatoren zwar keine Kovarianzmatrix der interessierenden Variablen geschätzt werden, jedoch eine Matrix mit polychorischen bzw. polyserialen Korrelationen. Diese ist ein befriedigender Ersatz für die Kovarianzmatrix, wenn die drei folgenden Bedingungen erfüllt sind (vgl. Jöreskog & Sörbom 1989, S. 47):

- Das zu analysierende Strukturgleichungsmodell ist **skaleninvariant** ist (vgl. Abschnitte 2.4.1 und 2.4.3).
- Das Modell kann die Varianzen der manifesten Variablen perfekt reproduzieren, d.h.:

$$\text{diag}(\hat{\Sigma}) = \text{diag}(S)$$

- Es wird und eine skaleninvariante Schätzmethode (ML, GLS, WLS, DWLS) verwendet.

Dann ändern sich nämlich beim Wechsel der Maßeinheiten für manifeste Variablen die Parameterschätzungen nur entsprechend den verwendeten Skalierungsfaktoren. In dieser Situation erbringt also die Korrelationsmatrix im wesentlichen dieselben Parameter-Schätzungen wie die Kovarianzmatrix.

Leider sind viele interessante Strukturgleichungsmodelle nicht skaleninvariant. Ein Beispiel ist unser Latent-State-Modell der Angst, wie nun kurz gezeigt werden soll. Dazu wiederholen wir zunächst die Definition der Skaleninvarianz: Ein Strukturgleichungsmodell mit der Fundamental-Hypothese $\Sigma = f(\theta)$ heißt skaleninvariant, wenn für jede Diagonalmatrix D mit positiven Einträgen und für jeden Parametervektor θ ein anderer Parametervektor $\tilde{\theta}$ existiert mit:

$$f(\tilde{\theta}) = Df(\theta)D$$

In Abschnitt 2.3 haben wir die folgende Fundamental-Hypothese $\Sigma = f(\theta)$ für das Latent-State-Modell der Angst hergeleitet:

$$\begin{bmatrix} \text{Var}(X_1) & & & & \\ \text{Cov}(X_2, X_1) & \text{Var}(X_2) & & & \\ \text{Cov}(X_3, X_1) & \text{Cov}(X_3, X_2) & \text{Var}(X_3) & & \\ \text{Cov}(X_4, X_1) & \text{Cov}(X_4, X_2) & \text{Cov}(X_4, X_3) & \text{Var}(X_4) & \end{bmatrix} = \begin{bmatrix} \phi_{11} + \vartheta & & & & \\ \phi_{11} & \phi_{11} + \vartheta & & & \\ \phi_{21} & \phi_{21} & \phi_{22} + \vartheta & & \\ \phi_{21} & \phi_{21} & \phi_{22} & \phi_{22} + \vartheta & \end{bmatrix}$$

Es wird also z.B. für die Diagonalmatrix

$$D = \text{diag}(1, 2, 1, 1)$$

ein Parametervektor $\tilde{\theta} = (\tilde{\phi}_{11}, \tilde{\phi}_{21}, \tilde{\phi}_{22}, \tilde{\vartheta})$ gesucht mit:

$$\begin{bmatrix} \tilde{\phi}_{11} + \tilde{\vartheta} & & & \\ \tilde{\phi}_{11} & \tilde{\phi}_{11} + \tilde{\vartheta} & & \\ \tilde{\phi}_{21} & \tilde{\phi}_{21} & \tilde{\phi}_{22} + \tilde{\vartheta} & \\ \tilde{\phi}_{21} & \tilde{\phi}_{21} & \tilde{\phi}_{22} & \tilde{\phi}_{22} + \tilde{\vartheta} \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} \phi_{11} + \vartheta & & & \\ \phi_{11} & \phi_{11} + \vartheta & & \\ \phi_{21} & \phi_{21} & \phi_{22} + \vartheta & \\ \phi_{21} & \phi_{21} & \phi_{22} & \phi_{22} + \vartheta \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 1 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} \phi_{11} + \vartheta & & & \\ 2\phi_{11} & 4(\phi_{11} + \vartheta) & & \\ \phi_{21} & 2\phi_{21} & \phi_{22} + \vartheta & \\ \phi_{21} & 2\phi_{21} & \phi_{22} & \phi_{22} + \vartheta \end{bmatrix}$$

Übung:

Zeigen Sie, daß kein Parametervektor $\tilde{\theta} = (\tilde{\phi}_{11}, \tilde{\phi}_{21}, \tilde{\phi}_{22}, \tilde{\vartheta})$ mit den geforderten Eigenschaften existiert, und folglich das Modell nicht skaleninvariant ist.

Um die Analyse von Strukturgleichungsmodellen anhand von polychorischen Korrelationen und der WLS-Schätzmethode üben zu können, soll das Latent-State-Modell zu einem skaleninvarianten Modell umgeformt werden. Dazu müssen wir die beiden Gleichheitsrestriktionen:

$$\lambda_{11} = \lambda_{21} = \lambda_{32} = \lambda_{42}$$

$$\text{Var}(\delta_1) = \text{Var}(\delta_2) = \text{Var}(\delta_3) = \text{Var}(\delta_4)$$

aufheben. Die übrigen Modellgleichungen bleiben unverändert:

$$\begin{array}{rcl} X_1 & = & \lambda_{11}\xi_1 & + & \delta_1 \\ X_2 & = & \lambda_{21}\xi_1 & + & \delta_2 \\ X_3 & = & & \lambda_{32}\xi_2 & + & \delta_3 \\ X_4 & = & & \lambda_{42}\xi_2 & + & \delta_4 \end{array}$$

Übung:

Zeigen Sie für das modifizierte Modell, daß es skaleninvariant ist und die Varianzen der manifesten Variablen perfekt reproduziert.

Wir wollen nun mit dem umgeformten Modell eine Simulationsstudie zum Verhalten des WLS-Schätzers aus polychorischen Korrelationen durchführen. Dabei legen wir für ϕ_{11} , ϕ_{21} , ϕ_{22} , λ_{11} , λ_{21} , λ_{32} , λ_{42} , $\text{Var}(\delta_1)$, $\text{Var}(\delta_2)$, $\text{Var}(\delta_3)$ und $\text{Var}(\delta_4)$ wahre Parameterwerte fest. Zur Festlegung der Maßeinheiten für die latenten Variablen soll hier einmal die in Abschnitt 3.8 aufgeführte alternative Strategie eingesetzt

werden: wir standardisieren also ξ_1 und ξ_2 auf eine Varianz von Eins. Insgesamt sollen folgende Parameterwerte in der Simulationsstudie verwendet werden:

$$\begin{aligned}\phi_{11} &= \phi_{22} = 1.0 \\ \phi_{12} &= 0.64 \\ \lambda_{11} &= \lambda_{32} = \sqrt{2} \\ \lambda_{21} &= \lambda_{42} = 1.0 \\ \text{Var}(\delta_1) &= \text{Var}(\delta_2) = \text{Var}(\delta_3) = \text{Var}(\delta_4) = 1.0\end{aligned}$$

Dies sind auch die Erwartungswerte für die aus einer Stichproben-Kovarianzmatrix gewonnenen Parameterschätzer. Bei der Analyse einer Stichproben-Korrelationsmatrix ergeben sich jedoch andere Erwartungswerte für die Schätzer, wie folgende Überlegung zeigt. In unserer künstlichen Population gilt für X_1 :

$$X_1 = \lambda_{11}\xi_1 + \delta_1$$

Bei der Analyse einer Korrelationsmatrix betrachten wir statt X_1 jedoch:

$$\begin{aligned}\tilde{X}_1 &:= \frac{1}{\sqrt{\text{Var}(X_1)}} X \\ &= \frac{1}{\sqrt{\text{Var}(X_1)}} (\lambda_{11}\xi_1 + \delta_1) \\ &= \tilde{\lambda}_{11}\xi_1 + \tilde{\delta}_1, \quad \text{mit } \tilde{\lambda}_{11} := \frac{\lambda_{11}}{\sqrt{\text{Var}(X_1)}} \quad \text{und } \tilde{\delta}_1 := \frac{1}{\sqrt{\text{Var}(X_1)}} \delta_1\end{aligned}$$

Der Erwartungswert für den Schätzer zum Ladungsparameter $\tilde{\lambda}_{11}$ (bei Analyse einer Korrelationsmatrix) ist also:

$$\frac{\lambda_{11}}{\sqrt{\text{Var}(X_1)}}$$

Damit lauten die Erwartungswerte für die geplante Simulationsstudie:

Parameter	Erwartungswert
$\tilde{\lambda}_{11}, \tilde{\lambda}_{32}$	$\sqrt{\frac{1}{3}} \approx 0.57735$
$\tilde{\lambda}_{21}, \tilde{\lambda}_{42}$	$\sqrt{\frac{1}{2}} \approx 0.7071$
ϕ_{21}	0.64

Das Generieren der simulierten Stichprobendaten soll in zwei Schritten erfolgen:

- Im ersten Schritt werden intervallskalierte (normalverteilte) Indikatoren produziert. Dazu dienen die folgenden SPSS-Kommandos:

```
set seed=812833
set mxloops=500
input program
+ loop #i=1 to 500
+ compute #ksi=normal(1)
+ compute #ksi1=0.8*#ksi + normal(0.6)
+ compute #ksi2=0.8*#ksi + normal(0.6)
+ compute X1=0.7071*#ksi1 + normal(1)
+ compute X2=#ksi1 + normal(1)
+ compute X3=0.7071*#ksi2 + normal(1)
+ compute X4=#ksi2 + normal(1)
+ end case
+ end loop
+ end file
end input program
```

- Im zweiten Schritt werden die intervallskalierten Indikatoren durch willkürlich gewählte Schnittpunkte zu ordinalen Indikatoren mit fünf bzw. sechs Stufen vergrößert. Dazu dienen die folgenden SPSS-Kommandos:

```
recode x1 x3 (lo thru -2 = 1)
           (-2 thru -1.5 = 2)
           (-1.5 thru 0 = 3)
           ( 0 thru 2 = 4)
           ( 2 thru hi = 5) /
x2 x4 (lo thru -2.5 = 1)
        (-2.5 thru -1 = 2)
        ( -1 thru 1 = 3)
        ( 1 thru 1.5 = 4)
        ( 1.5 thru 2.5 = 5)
        ( 2.5 thru hi = 6)
```

Anhand der so erzeugten Daten sollen folgende Schätzmethoden verglichen werden:

- A ML-Schätzer, berechnet aus den Produkt-Moment-Korrelationen intervallskalierter (normalverteilter) Indikatoren
Da unser modifiziertes Modell skaleninvariant ist und alle Verteilungsvoraussetzungen perfekt erfüllt sind, ist dies ein zulässiges Vorgehen. Man erhält in der Tat exakt dieselbe χ^2 -Prüfstatistik für den Modelltest wie bei der Analyse der Kovarianzmatrix.
- B ML-Schätzer, berechnet aus den Produkt-Moment-Korrelationen ordinaler Indikatoren
Hierbei handelt es sich um ein unsauberes Vorgehen, das aber in der Forschungspraxis häufig anzutreffen ist.
- C ML-Schätzer, berechnet aus den polychorischen Korrelationen ordinaler Indikatoren
Die polychorischen Korrelationen sind zwar den in Methode B verwendeten Korrelationsmaßen überlegen, aber sie besitzen ein anderes Verteilungsverhalten, als es in der ML-Theorie bei der Berechnung der Standardfehler und der χ^2 -Prüfstatistik in Rechnung gestellt wird.
- D WLS-Schätzer, berechnet aus den polychorischen Korrelationen ordinaler Variablen sowie den geschätzten Varianzen und Kovarianzen der polychorischen Korrelationen
Dieses Vorgehen sollte zu unverzerrten und konsistenten Schätzungen sowie zu korrekten inferenzstatistischen Ergebnissen (Standardfehler und χ^2 -Prüfstatistik) führen, da wir den Stichprobenumfang bei unserer Simulationsstudie mit $N = 500$ hinreichend groß wählen können.

Um einen Eindruck vom Verteilungsverhalten der verschiedenen Schätzer zu erlangen, sollen insgesamt 30 simulierte Stichproben der Größe $N = 500$ analysiert werden. Die Auswertung erfolgt natürlich mit PRELIS und LISREL. Weil uns in diesem Zusammenhang erstmals der kombinierte Einsatz von PRELIS und LISREL begegnet, der später (siehe z.B. Abschnitt 6.2) als Standardverfahren empfohlen wird, sollen beispielhaft die zur Anwendung von Schätzmethode D verwendeten Kommandos wiedergegeben und kurz kommentiert werden:

```
prelis
/variables=x1 to x4 (ordinal)
/type=polychor
/matrix=out (*)
/write=acov 'presi.acp'

lisrel
/"Analyse der Poly.-Kor.-Matr. aus ordinalen Indikatoren, WLS"
/da ni=4 ma=pm
/matrix=in (*)
/ac file='presi.acp'
/mo nx=4 nk=2 ph=st
/fr lx(1,1) lx(2,1) lx(3,2) lx(4,2)
/ou
```

PRELIS wird hier wie jede beliebige SPSS-Prozedur aufgerufen. Die Subkommandos haben folgende Bedeutung:

- Im VARIABLES-Subkommando wird angegeben, daß die Variablen X1 bis X4 aus der aktiven SPSS-Systemdatei analysiert werden sollen. Durch das Schlüsselwort ORDINAL werden diese als ordinalskaliert deklariert.
- Im TYPE-Subkommando wird festgelegt, daß eine Matrix mit polychorischen Korrelationen erstellt werden soll.
- Im MATRIX-Subkommando wird veranlaßt, daß die Matrix mit den polychorischen Korrelationen als SPSS-Matrix-Systemdatei die aktive SPSS-Systemdatei ersetzen soll.
- Im WRITE-Subkommando wird verlangt, daß die geschätzten Varianzen und Kovarianzen der polychorischen Korrelationen in die Textdatei PRESI.ACP geschrieben werden sollen.

Das LISREL-Kommando enthält gegenüber der in Aufgabe 3.1 erläuterten Version folgende Änderungen:

- IM DA-Subkommando wird mit "MA=PM" angegeben, daß eine Matrix mit polychorischen Korrelationen analysiert werden soll.
- Mit dem AC-Subkommando wird das Einlesen der geschätzten Kovarianzmatrix der geschätzten polychorischen Korrelationen angeordnet. Dadurch wird automatisch auch die Schätzmethode WLS gewählt.
- Im MO-Subkommando wird durch "PH=ST" festgelegt, daß Φ eine symmetrische Matrix mit fixierten Einsen in der Hauptdiagonalen (also eine Korrelationsmatrix) sein soll. Damit sind insbesondere, wie am Anfang dieses Abschnitts besprochen, die Varianzen von ξ_1 und ξ_2 auf Eins fixiert. Dementsprechend ist die Fixierung der Ladungen auf Eins entfallen. Schließlich wurden aus den oben angegebenen Gründen die Gleichheitsrestriktionen für die Ladungsparameter und für die Residualvarianzen weggelassen.

Nun sollen endlich die Resultate der Simulationsstudie vorgestellt werden. Die folgende Tabelle enthält für die oben beschriebenen Schätzverfahren A - D jeweils zu den Parametern λ_{11} , λ_{21} , λ_{32} , λ_{42} und ϕ_{21} sowie zur χ^2 -Prüfstatistik aus den 30 Durchgängen des Experimentes den Mittelwert des Schätzers bzw. der Prüfgröße und den mittleren Fehler. Letzterer ist definiert als die mittlere Abweichung der Stichprobenwerte von ihrem (theoretischen) Erwartungswert, der in den Spaltenüberschriften der Tabelle nochmals angegeben wird. Der Erwartungswert der χ^2 -Prüfstatistik ergibt sich einfach daraus, daß diese

bei Gültigkeit der Nullhypothese und einem Freiheitsgrad der χ^2 -Verteilung mit $df=1$ folgt und somit den Erwartungswert Eins hat.

Schätz- methode	λ_{11} (0.5773)	λ_{21} (0.7071)	λ_{32} (0.5773)	λ_{42} (0.7071)	ϕ_{21} (0.64)	χ^2 (1.0)
	Mittel mittl. Fehler	Mittel mittl. Fehler	Mittel mittl. Fehler	Mittel mittl. Fehler	Mittel mittl. Fehler	Mittel mittl. Fehler
A	0.587 0.010	0.695 -0.012	0.574 -0.003	0.717 0.010	0.641 0.001	1.737 0.737
B	0.525 -0.052	0.643 -0.064	0.515 -0.063	0.663 -0.044	0.667 0.027	1.537 0.573
C	0.586 0.009	0.691 -0.016	0.571 -0.007	0.718 0.010	0.656 0.016	2.831 1.831
D	0.588 0.010	0.689 -0.018	0.577 0.000	0.717 0.010	0.659 0.019	0.827 -0.173

Die Ergebnisse legen u.a. folgende Schlüsse auf Eigenschaften der Schätzmethoden nahe:

- Bei Methode B sind die geschätzten Ladungen infolge der Verwendung ordinaler Indikatoren systematisch gemindert.
- Bei Methode C, die ebenfalls von ordinalskalierten Indikatoren ausgeht, sind die Parameterschätzungen praktisch ebenso präzise wie bei Methode A, die mit den intervallskalierten Indikatoren arbeitet. Allerdings scheint die χ^2 -Statistik systematisch überhöht zu sein.
- Die Methode D (WLS) erbringt sehr ähnliche Schätzer wie Methode C und außerdem eine korrekt verteilte χ^2 -Statistik. Deren Stichprobenmittelwert liegt knapp unter dem Erwartungswert Eins, ihr Stichprobenmedian liegt sogar fast exakt beim theoretischen Wert von ca. 0.46. Damit hat die aufwendigste Schätzmethode in unserer kleinen Simulationsstudie den besten Eindruck hinterlassen. Sie scheint tatsächlich die beste Wahl zu sein, wenn bei einer echten Untersuchung für die manifesten Variablen nur krude Indikatoren (auf Ordinalniveau) vorliegen.

Kühnel (1993) hat eine ähnliche Simulationsstudie durchgeführt und dabei vergleichbare Ergebnisse erzielt:

- Beim Vorgehen wie in unserer Methode B (Ignorieren des ordinalen Meßniveaus) konnten korrekte von fehlspezifizierten linearen Modellen gut unterschieden werden.
- Bei expliziter Berücksichtigung des ordinalen Meßniveaus verbesserte sich die Genauigkeit der Parameterschätzungen.

Diese Ergebnisse traten auch dann auf, wenn bei der Vergrößerung der intervallskalierten Variablen die Schnittpunkte nicht konstant waren (wie bei unserem Vorgehen), sondern zufällig variierten, um individuell verschiedene Übersetzungsmechanismen zu simulieren.

5 LISREL-Ausgaben zur Modelldiagnose und zur Modellmodifikation

Wir haben schon in Abschnitt 2 wichtige LISREL-Ausgaben zur Modelldiagnose und -modifikation kennengelernt, die dort zur Veranschaulichung von zentralen Begriffen und Verfahren der Kovarianzstrukturanalyse dienen. Nun werden ergänzend weitere Informationen diskutiert, die LISREL über ein Modell liefern kann.

Bei der Ergebnisinspektion steht natürlich zunächst die Frage nach der Modell-Gültigkeit im Vordergrund. Zu ihrer Klärung sind neben den globalen Fit-Indikatoren (vgl. Abschnitt 2.5) auch Detail-Ergebnisse heranzuziehen (z.B. Parameterschätzungen, Residuen, Determinationskoeffizienten).

Bei positivem Ausgang kann die gefundene Lösung interpretiert werden, wobei LISREL den Forscher durch zahlreiche diagnostische Informationen unterstützt (z.B. standardisierte Lösung, Matrizen mit totalen und indirekten Effekten, Stabilitätsindex).

Bei negativem Ausgang kommt eine Modifikation des Modells in Betracht, wobei LISREL mit genauen Hinweisen zur Lokalisation der Schwachstellen hilft.

5.1 Modelldiagnose

Zur Erläuterung der von LISREL angebotenen Ergebnisse zur detaillierten Beschreibung einer Lösung eignet sich das konstruierte Beispiel 1.3 sehr gut, weil es sehr viele Möglichkeiten der LISREL-Modellierung enthält. Jöreskog & Sörbom (1986, Kapitel III.9) haben durch Wahl konkreter Werte für die Parameter eine künstliche Population mit folgendem wahren Modell (bei gültiger Normalverteilungsannahme) festgelegt:

i) Strukturmodell

$$\begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} = \begin{bmatrix} 0.000 & 0.493 \\ 0.595 & 0.000 \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} + \begin{bmatrix} 0.399 & 0.399 & 0.000 \\ -1.000 & 0.000 & 1.098 \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{bmatrix} + \begin{bmatrix} \zeta_1 \\ \zeta_2 \end{bmatrix}$$

ii) Meßmodell

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{bmatrix} = \begin{bmatrix} 1.000 & 0.0 \\ 0.902 & 0.0 \\ 0.0 & 1.000 \\ 0.0 & 1.095 \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{bmatrix}$$

$$\begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \\ X_6 \\ X_7 \end{bmatrix} = \begin{bmatrix} 1.000 & 0.0 & 0.0 \\ 1.300 & 0.0 & 0.0 \\ 0.901 & 1.200 & 0.0 \\ 0.0 & 1.000 & 0.0 \\ 0.0 & 1.098 & 0.0 \\ 0.0 & 0.0 & 1.000 \\ 0.0 & 0.0 & 1.400 \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{bmatrix} + \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \\ \delta_6 \\ \delta_7 \end{bmatrix}$$

- Mit dem CM-Subkommando wird das Einlesen einer Kovarianzmatrix angefordert. Der Steuer-Parameter SY besagt, daß die Matrix in unterer Dreiecksgestalt vorliegt und daß jede Matrixzeile in einer neuen Dateizeile beginnt. Leider muß zugunsten der Konsistenz mit der Originaldokumentation von LISREL der bereits belegte Ausdruck "Parameter" hier in völlig anderer Bedeutung nochmals benutzt werden.

In der FI-Spezifikation wird mitgeteilt, aus welcher Textdatei die Kovarianzmatrix gelesen werden soll, und "FO" signalisiert, daß die Zeile mit dem FORTRAN-Format zur Steuerung des Einlesens *nicht* unmittelbar vor den Daten in der externen Textdatei liegt, sondern dem CM-Subkommando im Programm folgt. Ohne den Hinweis "FO" würde LISREL sinnloserweise versuchen, die Formatangabe "(11f5.3)" als Subkommando zu interpretieren.

- Mit dem MO-Subkommando beginnt die Modell-Spezifikation. Hier erfährt LISREL, wie viele Y-, X-, η - und ξ -Variablen im Modell enthalten sind. Da unser Modell die freien Parameter β_{21} und β_{12} enthält, muß für die Parametermatrix B die voreingestellte Form ZE (Null-Matrix) auf FU (volle Rechtecksmatrix) abgeändert werden (vgl. Tabelle 3.6). Die voreingestellte Fixierung der β -Parameter auf Null bleibt zwar durch diese Form-Erweiterung zunächst unberührt, doch darf nun der Modus der β -Parameter in folgenden Subkommandos geändert werden. Welche weiteren Spezifikationen nötig sind, hängt vom Modell und von den voreingestellten Modi der Parametermatrizen ab (vgl. Tabelle 3.5).
 - Im FR-Subkommando werden Parameter frei gesetzt, die im Modell frei schätzbar sind, aber per LISREL-Voreinstellung fixiert. Neben den eben angesprochenen β -Parametern sind dies in unserem Beispiel die modellgemäß frei schätzbaren Ladungsparameter aus den per Voreinstellung fixierten Matrizen Λ_X und Λ_Y .
 - Γ hat per Voreinstellung die Form FU(II) und den Modus FR(ee). Folglich müssen die im Modell fixierten γ -Parameter in einem FI-Subkommando aufgelistet werden.
 - Von den modellgemäß fixierten Ladungen müssen einige auf den Wert Eins statt auf den bei LISREL voreingestellten Wert Null gesetzt werden. Dies geschieht in einem VA(lue)-Subkommando.
- Alle anderen Parameter-Spezifikationen unseres Modells entsprechen den LISREL-Voreinstellungen, z.B. ist die Kovarianzmatrix Φ der ξ -Variablen per Voreinstellung und modellgemäß vom Modus FR(ee). Damit ist die Modell-Deklaration beendet.
- Im OU(tput)-Subkommando werden folgende Ausgaben angefordert: Standardfehler und T-Werte zu den Parameter-Schätzungen, Modifikationsindikatoren zu den fixierten oder restringierten Parametern, Informationen über die Residuen, Matrizen mit den totalen und indirekten Effekten die vollständig standardisierte Lösung und verschiedene Kovarianzen.

Am Anfang des Listings erfolgt eine Kontrollausgabe der zu analysierenden Matrix:

HYPOTHETISCHES MODELL AUS BEISPIEL 1.3 COVARIANCE MATRIX TO BE ANALYZED											
	Y1	Y2	Y3	Y4	X1	X2	X3	X4	X5	X6	X7
Y1	3.204										
Y2	2.722	2.629									
Y3	3.198	2.875	4.855								
Y4	3.545	3.202	5.373	6.315							
X1	0.329	0.371	-0.357	-0.471	1.363						
X2	0.559	0.592	-0.316	-0.335	1.271	1.960					
X3	1.006	1.019	-0.489	-0.591	1.742	2.276	3.803				
X4	0.468	0.456	-0.438	-0.539	0.788	1.043	1.953	1.376			
X5	0.502	0.539	-0.363	-0.425	0.838	1.070	2.090	1.189	1.741		
X6	1.050	0.960	1.416	1.714	0.474	0.694	0.655	0.071	0.104	1.422	
X7	1.260	1.154	1.923	2.309	0.686	0.907	0.917	0.136	0.162	1.688	2.684

Unter den gegebenen idealen Voraussetzungen der multivariaten Normalität ist die ML-Schätzmethode wegen ihrer asymptotischen Effizienz die beste Wahl. Damit kann der Likelihood-Quotienten - χ^2 -Test zur Fit-Beurteilung herangezogen werden.

5.1.1 Fit-Beurteilung

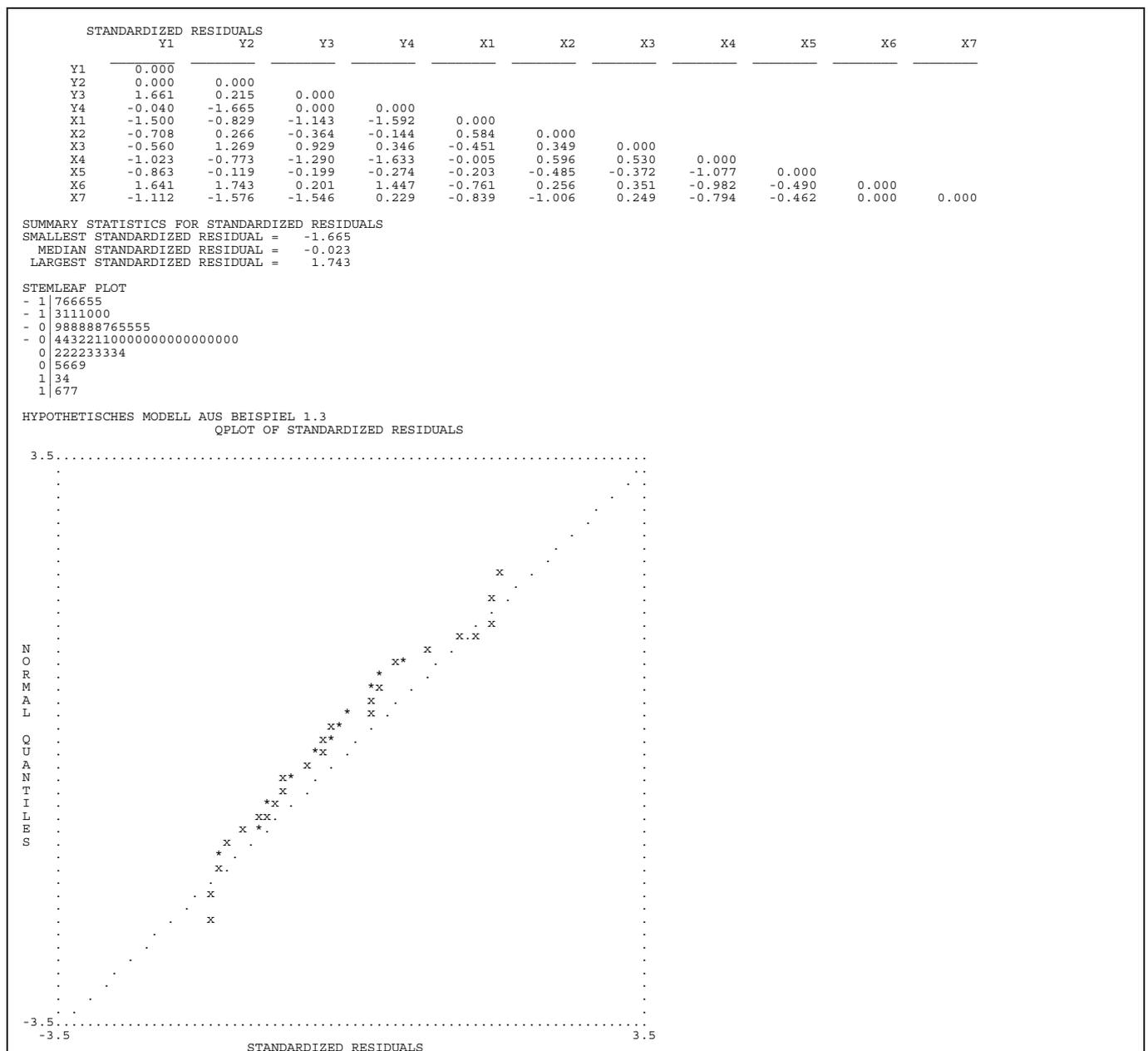
Ein Blick auf die LISREL-Ergebnisse zur **globalen Fit-Beurteilung** kann einen Eindruck vom Verhalten dieser Maße bei Gültigkeit des Modells vermitteln:

```

CHI-SQUARE WITH 33 DEGREES OF FREEDOM = 29.10 (P = .662)
GOODNESS OF FIT INDEX =0.953
ADJUSTED GOODNESS OF FIT INDEX =0.906
ROOT MEAN SQUARE RESIDUAL = 0.065
  
```

Diese Ergebnisse sprechen klar für die Interpretierbarkeit der Parameter-Schätzer und ihrer Standardfehler.

Zur weiteren Überprüfung der Modell-Gültigkeit ist eine Inspektion der Residuen unbedingt erforderlich. Wie in Abschnitt 2.5.4 ausgeführt, müssen die standardisierten Residuen annähernd normalverteilt sein, was für Beispiel 1.3 durch den Stem-and-Leaf - Plot und durch den Q-Plot bestätigt wird:



5.1.2 Quadrierte multiple Korrelationen und Determinationskoeffizienten

LISREL gibt für jede Indikatorvariable eine Schätzung ihrer quadrierten multiplen Korrelation mit allen bedingenden latenten Variablen aus. Sie ist z.B. für Y_i definiert durch:

$$1 - \frac{\hat{\theta}_{ii}^{(\varepsilon)}}{\hat{\sigma}_{ii}^{(Y)}}$$

wobei $\hat{\theta}_{ii}^{(\varepsilon)}$ die geschätzte Fehlervarianz, d.h. die geschätzte Varianz von ε_i , und $\hat{\sigma}_{ii}^{(Y)}$ die vom Modell implizierte Varianz von Y_i ist. Dieser Koeffizient entspricht der Kommunalität h_i^2 , die z.B. bei einer exploratorischen Faktorenanalyse üblicherweise berechnet wird.

Die Verallgemeinerung dieser quadrierten multiplen Korrelation beschreibt, wie gut die Kovarianzmatrix der Y - bzw. X -Variablen durch die zugehörigen latenten Variablen aufgeklärt wird. Für die Y -Variablen ist dieser **Determinationskoeffizient** definiert durch:

$$1 - \frac{|\hat{\Theta}_\varepsilon|}{|\hat{\Sigma}_Y|}$$

wobei $\hat{\Sigma}_Y$ die vom Modell implizierte Kovarianzmatrix der Y -Variablen ist (also das "Y-Segment" von $f(\theta)$). Für die X -Variablen ist der Determinationskoeffizient analog definiert. In der Ergebnisliste ist die Ausgabe überschrieben durch: "TOTAL COEFFICIENT OF DETERMINATION".

Es ist zu beachten, daß hier nicht die Fähigkeit des Modells zur Erklärung von S beurteilt wird, sondern die Güte der Meßinstrumente für die latenten Variablen. In unserem Beispiel 1.3 geht es also im konkreten Fall der Indikatorvariablen Y_1 nicht darum, wie gut ihre Stichprobenvarianz s_{11} durch das Modell *inklusive* ε_1 erklärt werden kann, sondern es wird beurteilt, wie groß der durch η_1 determinierte Varianzanteil von Y_1 ist.

Die Koeffizienten liegen in der Regel zwischen 0 und 1, wobei ein hoher Wert für reliable Messungen spricht. Bei der Interpretation ist zu beachten, daß die Determinationskoeffizienten für die X - bzw. Y -Variablen oft unrealistisch hoch ausfallen (vgl. Bollen 1989, S. 119).

Zur Beschreibung der Effektstärken im Strukturmodell liefert LISREL für jedes η_i folgenden Schätzer des erklärten Varianzanteils ("quadrierte multiple Korrelation"):

$$1 - \frac{\hat{\text{Var}}(\zeta_i)}{\hat{\text{Var}}(\eta_i)}$$

und für alle Strukturgleichungen gemeinsam den Determinationskoeffizienten:

$$1 - \frac{|\hat{\Psi}|}{|\hat{\text{Cov}}(\eta)|}$$

der in der LISREL-Ausgabe als "TOTAL COEFFICIENT OF DETERMINATION FOR STRUCTURAL EQUATIONS" bezeichnet wird.

Für das konstruierte Beispiel 1.3 ergeben sich sehr günstige Werte, die bei echten sozialwissenschaftlichen Anwendungen selten auftreten dürften:

SQUARED MULTIPLE CORRELATIONS FOR Y - VARIABLES						
Y1	Y2	Y3	Y4			
0.923	0.953	0.972	0.969			
TOTAL COEFFICIENT OF DETERMINATION FOR Y - VARIABLES IS 0.999						
SQUARED MULTIPLE CORRELATIONS FOR X - VARIABLES						
X1	X2	X3	X4	X5	X6	X7
0.715	0.829	0.983	0.812	0.747	0.826	0.903
TOTAL COEFFICIENT OF DETERMINATION FOR X - VARIABLES IS 1.000						
SQUARED MULTIPLE CORRELATIONS FOR STRUCTURAL EQUATIONS						
ETA 1	ETA 2					
0.836	0.972					
TOTAL COEFFICIENT OF DETERMINATION FOR STRUCTURAL EQUATIONS IS 0.986						

Die Determinationskoeffizienten bewegen sich im zulässigen Wertebereich. Damit hat das Modell einen weiteren Gültigkeits-Test bestanden und wir dürfen die Parameter-Schätzungen und deren Standardfehler interpretieren.

5.1.3 ML-Schätzer und T-Werte

LISREL ermittelt für das konstruierte Beispiel folgende ML-Schätzungen und T-Werte:

HYPOTHETISCHES MODELL AUS BEISPIEL 1.3							
LISREL ESTIMATES (MAXIMUM LIKELIHOOD)							
LAMBDA Y							
	ETA 1	ETA 2					
Y1	1.000	0.000					
Y2	0.921	0.000					
Y3	0.000	1.000					
Y4	0.000	1.139					
LAMBDA X							
	KSI 1	KSI 2	KSI 3				
X1	1.000	0.000	0.000				
X2	1.291	0.000	0.000				
X3	0.920	1.092	0.000				
X4	0.000	1.000	0.000				
X5	0.000	1.079	0.000				
X6	0.000	0.000	1.000				
X7	0.000	0.000	1.437				
BETA							
	ETA 1	ETA 2					
ETA 1	0.000	0.538					
ETA 2	0.937	0.000					
GAMMA							
	KSI 1	KSI 2	KSI 3				
ETA 1	0.213	0.495	0.000				
ETA 2	-1.223	0.000	0.996				
COVARIANCE MATRIX OF ETA AND KSI							
	ETA 1	ETA 2	KSI 1	KSI 2	KSI 3		
ETA 1	2.957						
ETA 2	3.115	4.719					
KSI 1	0.482	-0.217	0.974				
KSI 2	0.554	-0.312	0.788	1.117			
KSI 3	0.932	1.402	0.525	0.133	1.175		
PSI							
	ETA 1	ETA 2					
ETA 1	0.486						
ETA 2	-0.069	0.133					
THETA EPS							
	Y1	Y2	Y3	Y4			
	0.247	0.123	0.136	0.197			
THETA DELTA							
	X1	X2	X3	X4	X5	X6	X7
	0.389	0.336	0.063	0.259	0.440	0.247	0.259

HYPOTHETISCHES MODELL AUS BEISPIEL 1.3							
T-VALUES							
LAMBDA Y							
	ETA 1	ETA 2					
Y1	0.000	0.000					
Y2	25.993	0.000					
Y3	0.000	0.000					
Y4	0.000	38.739					
LAMBDA X							
	KSI 1	KSI 2	KSI 3				
X1	0.000	0.000	0.000				
X2	12.335	0.000	0.000				
X3	7.462	9.318	0.000				
X4	0.000	0.000	0.000				
X5	0.000	12.798	0.000				
X6	0.000	0.000	0.000				
X7	0.000	0.000	15.479				
BETA							
	ETA 1	ETA 2					
ETA 1	0.000	9.525					
ETA 2	5.255	0.000					
GAMMA							
	KSI 1	KSI 2	KSI 3				
ETA 1	1.389	3.351	0.000				
ETA 2	-10.052	0.000	6.566				
PHI							
	KSI 1	KSI 2	KSI 3				
KSI 1	5.170						
KSI 2	5.212	5.725					
KSI 3	3.927	1.059	5.784				
PSI							
	ETA 1	ETA 2					
ETA 1	3.834						
ETA 2	-0.408	1.698					
THETA EPS							
	Y1	Y2	Y3	Y4			
	4.663	3.240	3.325	3.603			
THETA DELTA							
	X1	X2	X3	X4	X5	X6	X7
	6.094	5.001	1.229	5.136	5.875	4.596	2.827

Die meisten Schätzer liegen sehr nahe an den wahren Werten und die zugehörigen Signifikanz-Tests lehnen zurecht die Nullhypothesen ab. Merkliche Diskrepanzen ergeben sich allerdings bei β_{21} , ψ_{21} und ψ_{22} . Möglicherweise treten im Zusammenhang mit der wechselseitigen Abhängigkeit von η_1 und η_2 größere Schätzprobleme auf.

5.1.4 Direkte, indirekte und totale Effekte, Stabilitätsindex

Der Strukturkoeffizient β_{21} besagt, daß bei einer Erhöhung von η_1 um eine Einheit für η_2 eine Erhöhung um β_{21} Einheiten zu erwarten ist. Damit gibt β_{21} den **direkten** Effekt von η_1 auf η_2 an. Allerdings löst dieser Effekt aufgrund der non-rekursiven Struktur unseres Modells eine Kettenreaktion aus:

- η_2 revanchiert sich nämlich wegen des direkten Effektes β_{12} sofort bei η_1 mit $\beta_{12}\beta_{21}$, was nun zu einem Nachschlag von $\beta_{21}\beta_{12}\beta_{21}$ von η_1 nach η_2 führt, macht insgesamt einen "übertragenen Effekt" von $\beta_{21} + \beta_{21}\beta_{12}\beta_{21}$.
- η_2 beantwortet die zweite Lieferung prompt mit $\beta_{12}\beta_{21}\beta_{12}\beta_{21}$, was unweigerlich eine dritte Lieferung $\beta_{21}\beta_{12}\beta_{21}\beta_{12}\beta_{21}$ nach sich zieht und die Summe aller Effekte von η_1 nach η_2 auf $\beta_{21} + \beta_{21}\beta_{12}\beta_{21} + \beta_{21}\beta_{12}\beta_{21}\beta_{12}\beta_{21}$ steigert.
- Wir erhalten insgesamt eine unendliche Summe, die unter der Stabilitätsbedingung $|\beta_{12}\beta_{21}| < 1$ nach dem mathematischen Satz über geometrische Reihen gegen eine reelle Zahl konvergiert. Dies ist der **totale Effekt** von η_1 auf η_2 :

$$\beta_{21} \sum_{v=0}^{\infty} (\beta_{12}\beta_{21})^v = \frac{\beta_{21}}{1 - \beta_{12}\beta_{21}}, \text{ falls } |\beta_{12}\beta_{21}| < 1$$

Im Beispiel erhalten wird mit $\beta_{21} = 0.937$ und $\beta_{12} = 0.538$ (siehe oben) und damit $|\beta_{12}\beta_{21}| < 1$ als totalen Effekt von η_1 auf η_2 gerade 1.8895.

Der **indirekte Effekt** ist einfach als Differenz aus dem totalen und dem direkten Effekt definiert.

Zur Erläuterung des totalen Effektes ist das konstruierte Modell wegen seiner non-rekursiven Struktur leider kein geeignetes Einstiegsbeispiel. Eine umfangreiche und didaktisch gelungene Darstellung findet sich in Hayduk (1987, Kap. 8).

Aus einer Generalisierung obiger Überlegungen ergeben sich die folgenden von LISREL verwendeten Definitionen für direkte, indirekte und totale Effekte von ξ bzw. η auf η und Y (Jöreskog & Sörbom 1989, S. 35):

Tab. 5.1 Direkte, indirekte und totale Effekte

	$\xi \rightarrow \eta$	$\eta \rightarrow \eta$	$\xi \rightarrow Y$	$\eta \rightarrow Y$
Direkt	Γ	B	O	Λ_Y
Indirekt	$(I - B)^{-1}\Gamma - \Gamma$	$(I - B)^{-1} - I - B$	$\Lambda_Y(I - B)^{-1}\Gamma$	$\Lambda_Y(I - B)^{-1} - \Lambda_Y$
Total	$(I - B)^{-1}\Gamma$	$(I - B)^{-1} - I$	$\Lambda_Y(I - B)^{-1}\Gamma$	$\Lambda_Y(I - B)^{-1}$

Eine Verallgemeinerung der obigen Bedingung $|\beta_{12}\beta_{21}| < 1$ ist der von LISREL bestimmte **Stabilitätsindex**. Weil die Konvergenz der unendlichen Summen für die totalen Effekte in direktem Zusammenhang steht mit der Stabilität eines non-rekursiven Systems, ist der Stabilitätsindex hier von großer Bedeutung: Ein non-rekursives System ist genau dann stabil, wenn der Stabilitätsindex kleiner als Eins ist.

Die LISREL-Ausgabe für das Beispiel 1.3 bestätigt u.a. die obige Rechnung und belegt die Stabilität des Modells:

```

HYPOTHETISCHES MODELL AUS BEISPIEL 1.3
TOTAL AND INDIRECT EFFECTS
  TOTAL EFFECTS OF KSI ON  ETA
    KSI 1      KSI 2      KSI 3
ETA 1  -0.895   0.999   1.080
ETA 2  -2.062   0.936   2.008
  STANDARD ERRORS FOR TOTAL EFFECTS OF KSI ON  ETA
    KSI 1      KSI 2      KSI 3
ETA 1  0.428   0.342   0.224
ETA 2  0.517   0.403   0.268
  INDIRECT EFFECTS OF KSI ON  ETA
    KSI 1      KSI 2      KSI 3
ETA 1  -1.109   0.503   1.080
ETA 2  -0.839   0.936   1.012
  STANDARD ERRORS FOR INDIRECT EFFECTS OF KSI ON  ETA
    KSI 1      KSI 2      KSI 3
ETA 1  0.340   0.240   0.224
ETA 2  0.466   0.403   0.309
  TOTAL EFFECTS OF ETA ON  ETA
    ETA 1      ETA 2
ETA 1  1.016   1.084
ETA 2  1.890   1.016
LARGEST EIGENVALUE OF B*B' (STABILITY INDEX) IS  0.879
  STANDARD ERRORS FOR TOTAL EFFECTS OF ETA ON  ETA
    ETA 1      ETA 2
ETA 1  0.410   0.282
ETA 2  0.717   0.410
  INDIRECT EFFECTS OF ETA ON  ETA
    ETA 1      ETA 2
ETA 1  1.016   0.546
ETA 2  0.953   1.016
  STANDARD ERRORS FOR INDIRECT EFFECTS OF ETA ON  ETA
    ETA 1      ETA 2
ETA 1  0.410   0.247
ETA 2  0.548   0.410
  TOTAL EFFECTS OF ETA ON  Y
    ETA 1      ETA 2
Y1  2.016   1.084
Y2  1.856   0.998
Y3  1.890   2.016
Y4  2.152   2.296
  STANDARD ERRORS FOR TOTAL EFFECTS OF ETA ON  Y
    ETA 1      ETA 2
Y1  0.410   0.282
Y2  0.384   0.259
Y3  0.717   0.410
Y4  0.817   0.471
  INDIRECT EFFECTS OF ETA ON  Y
    ETA 1      ETA 2
Y1  1.016   1.084
Y2  0.936   0.998
Y3  1.890   1.016
Y4  2.152   1.157
  STANDARD ERRORS FOR INDIRECT EFFECTS OF ETA ON  Y
    ETA 1      ETA 2
Y1  0.410   0.282
Y2  0.380   0.259
Y3  0.717   0.410
Y4  0.817   0.468
  TOTAL EFFECTS OF KSI ON  Y
    KSI 1      KSI 2      KSI 3
Y1  -0.895   0.999   1.080
Y2  -0.824   0.919   0.994
Y3  -2.062   0.936   2.008
Y4  -2.348   1.066   2.287
  STANDARD ERRORS FOR TOTAL EFFECTS OF KSI ON  Y
    KSI 1      KSI 2      KSI 3
Y1  0.428   0.342   0.224
Y2  0.394   0.315   0.206
Y3  0.517   0.403   0.268
Y4  0.589   0.459   0.306

```

5.1.5 Varianzen und Kovarianzen

Aus den Modellgleichungen (3.1) und (3.2) sowie den in ihrem Kontext formulierten LISREL-Annahmen läßt sich die gesamte gemeinsame Kovarianzmatrix von Y , X , η und ξ herleiten. Z.B. läßt sich die Kovarianzmatrix von η , die in obigen Definitionen der Determinationskoeffizienten benötigt wird, wie folgt berechnen:

Wegen (3.1) gilt:

$$\eta = (I - B)^{-1}\Gamma\xi + (I - B)^{-1}\zeta$$

Daraus folgt wegen der Unkorreliertheit von ξ und ζ mit den Rechenregeln aus Abschnitt 2.1:

$$\text{Cov}(\eta) = (I - B)^{-1}\Gamma\Phi[(I - B)^{-1}\Gamma]^T + (I - B)^{-1}\Psi[(I - B)^{-1}]^T$$

Mit Hilfe der geschätzten Parametermatrizen erhält man also sofort eine Schätzung für $\text{Cov}(\eta)$. In der Ausgabe von LISREL 7 findet man diese Matrix im Abschnitt LISREL ESTIMATES. Die noch nicht anderswo ausgegebenen Segmente der gemeinsamen Kovarianzmatrix von Y , X , η und ξ erscheinen bei Anforderung durch den Steuerparameter MR des OU-Subkommandos im Abschnitt "COVARIANCES" der Ausgabe. Für Beispiel 1.3 erhalten wir:

HYPOTHETISCHES MODELL AUS BEISPIEL 1.3							
COVARIANCES							
Y - ETA							
	Y1	Y2	Y3	Y4			
ETA 1	2.957	2.722	3.115	3.547			
ETA 2	3.115	2.868	4.719	5.373			
Y - KSI							
	Y1	Y2	Y3	Y4			
KSI 1	0.482	0.444	-0.217	-0.247			
KSI 2	0.554	0.510	-0.312	-0.355			
KSI 3	0.932	0.858	1.402	1.596			
X - ETA							
	X1	X2	X3	X4	X5	X6	X7
ETA 1	0.482	0.622	1.048	0.554	0.598	0.932	1.339
ETA 2	-0.217	-0.280	-0.540	-0.312	-0.336	1.402	2.014
X - KSI							
	X1	X2	X3	X4	X5	X6	X7
KSI 1	0.974	1.258	1.757	0.788	0.851	0.525	0.754
KSI 2	0.788	1.018	1.945	1.117	1.206	0.133	0.192
KSI 3	0.525	0.678	0.629	0.133	0.144	1.175	1.688

5.1.6 Standardisierte Lösung

Bei der aus statistischen Gründen zu bevorzugenden Analyse einer Kovarianzmatrix sind die latenten Variablen im allgemeinen nicht standardisiert. LISREL kann aber als Interpretationshilfe aus der ursprünglichen Lösung eine standardisierte Lösung berechnen und zusätzlich ausgeben (siehe Steuerparameter SS im OU-Subkommando). Dabei werden die latenten η - und ξ -Variablen standardisiert zu η^* und ξ^* , während die manifesten Variablen ihre Original-Metrik behalten. Man erhält das Modell:

$$\eta^* = B^* \eta^* + \Gamma^* \gamma^* + \zeta^* \quad \Leftrightarrow \quad (I - B^*) \eta^* = \Gamma^* \gamma^* + \zeta^*$$

$$Y = \Lambda_Y^* \eta^* + \varepsilon$$

$$X = \Lambda_X^* \zeta^* + \delta$$

Während z.B. der unstandardisierte Strukturkoeffizient γ_{ij} angibt, um wieviel Einheiten sich η_i erwartungsgemäß verändert, wenn ξ_j um eine Einheit erhöht wird, enthält der standardisierte Strukturkoeffizient γ_{ij}^* die oft bequemere Information darüber, um wieviel Standardabweichungseinheiten sich η_i verändert, wenn ξ_j um eine Standardabweichung erhöht wird. Vor allem der Vergleich von Effekten verschiedener Variablen wird so erleichtert. Allerdings kann ein Vergleich verschiedener Populationen mit unterschiedlicher Variabilität erschwert werden.

Zur Berechnung von γ_{ij}^* aus γ_{ij} werden nur die Varianzen von η_i und ξ_j benötigt. Einer Erhöhung von ξ_j^* um eine Einheit entspricht eine Erhöhung von ξ_j um eine Standardabweichung. Folglich besagt $\gamma_{ij} \cdot \phi_{jj}^{1/2}$, um wieviel Einheiten (in der Original-Metrik) sich η_i bei Erhöhung von ξ_j^* um eine Einheit verändert. Schließlich gibt dann:

$$\gamma_{ij}^* = \sqrt{\frac{\phi_{jj}}{\text{Var}(\eta_i)}} \gamma_{ij}$$

an, um wieviel Einheiten sich η_i^* bei Erhöhung von ξ_j^* um eine Einheit verändert.

Mit analogen Argumenten lassen sich leicht die anderen zu standardisierten η^* - und ξ^* -Variablen gehörigen Parametermatrizen bestimmen. Bei der Interpretation dieser Matrizen ist zu beachten:

- Die manifesten Variablen bleiben in der ursprünglichen Metrik.
- Die ζ -Variablen gehen über in ζ^* -Variablen, jedoch sind letztere nicht standardisiert, weshalb ihre Kovarianzmatrix Ψ^* keine Korrelationsmatrix ist.
- Gleichheitsrestriktionen für Modellparameter gelten im allgemeinen in der standardisierten Lösung nicht mehr.

Eine vollständig standardisierte Lösung, bei der auch die manifesten Variablen auf eine Varianz von Eins normiert sind, kann mit dem Steuerparameter SC des OU-Subkommandos angefordert werden.

Es ist zu beachten, daß die standardisierten Lösungen *nach* der erfolgreichen Schätzung der Original-Lösung aus dieser berechnet werden, wobei die Varianzen der manifesten Variablen nicht aus S , sondern aus $\hat{\Sigma}$ übernommen werden. Daher kann die durch Analyse einer empirischen Kovarianzmatrix gewonnene (vollständig) standardisierte Lösung nur dann direkt aus der empirischen Korrelationsmatrix bestimmt werden, wenn für das Modell gilt:

$$\text{diag}(\hat{\Sigma}) = \text{diag}(S)$$

Zusammen mit der (vollständig) standardisierten Lösung gibt LISREL noch die gemeinsame Korrelationsmatrix der η - und ξ -Variablen sowie die Koeffizienten-Matrix für die Regression von η^* auf ξ^* aus. Dies ist gerade die Matrix der Koeffizienten aus der reduzierten Form für die standardisierten Variablen (siehe obige Modellgleichung):

$$(I - \hat{B}^*)^{-1} \hat{\Gamma}^*$$

Für unser Beispiel ermittelt LISREL die folgende standardisierte Lösung:

HYPOTHETISCHES MODELL AUS BEISPIEL 1.3
STANDARDIZED SOLUTION

LAMBDA Y		
	ETA 1	ETA 2
Y1	1.719	0.000
Y2	1.583	0.000
Y3	0.000	2.172
Y4	0.000	2.474

LAMBDA X			
	KSI 1	KSI 2	KSI 3
X1	0.987	0.000	0.000
X2	1.274	0.000	0.000
X3	0.908	1.154	0.000
X4	0.000	1.057	0.000
X5	0.000	1.141	0.000
X6	0.000	0.000	1.084
X7	0.000	0.000	1.557

BETA		
	ETA 1	ETA 2
ETA 1	0.000	0.679
ETA 2	0.742	0.000

GAMMA			
	KSI 1	KSI 2	KSI 3
ETA 1	0.123	0.304	0.000
ETA 2	-0.556	0.000	0.497

CORRELATION MATRIX OF ETA AND KSI					
	ETA 1	ETA 2	KSI 1	KSI 2	KSI 3
ETA 1	1.000				
ETA 2	0.834	1.000			
KSI 1	0.284	-0.101	1.000		
KSI 2	0.305	-0.136	0.756	1.000	
KSI 3	0.500	0.595	0.491	0.117	1.000

PSI		
	ETA 1	ETA 2
ETA 1	0.164	
ETA 2	-0.018	0.028

REGRESSION MATRIX ETA ON KSI (STANDARDIZED)			
	KSI 1	KSI 2	KSI 3
ETA 1	-0.514	0.614	0.681
ETA 2	-0.937	0.455	1.002

5.2 Modellmodifikation

Es ist natürlich trotz aller Vorzüge der konfirmatorischen Forschungsstrategie legitim, aus der Empirie zu lernen, d.h. Daten exploratorisch (hypothesenbildend) zu analysieren. Beim Arbeiten mit LISREL konfrontiert man ein apriori-Modell mit einer empirischen Kovarianzmatrix und erhält dabei eine Fülle diagnostischer Informationen, darunter u.a. auch Hinweise zu möglichen Ansatzpunkten für eine Verbesserung des Modells. Vor einer Diskussion von Techniken zur Verbesserung der Fit-Beurteilung muß jedoch nachdrücklich daran erinnert werden, daß ein mit Hilfe diagnostischer Informationen aus einer bestimmten Stichprobe verfeinertes Modell nur an einer völlig unabhängigen Stichprobe geprüft werden kann!

Wir haben schon in Abschnitt 2 den sogenannten Modifikationsindex $mi(\omega)$ für einen fixierten oder restringierten Parameter ω kennengelernt. Er gibt ungefähr die Reduktion der χ^2 -Prüfgröße des Modelltests bei Freisetzung von ω an. Die Aussage aus Abschnitt 2, daß $mi(\omega)$ daher approximativ einer χ^2 -Verteilung mit einem Freiheitsgrad folge, soll hier näher begründet werden. Dies soll im Rahmen der Beschreibung eines allgemeinen Prinzips zur Konstruktion von Hypothesentests mit LISREL geschehen.

M_u sei ein **gültiges** Kovarianzstrukturmodell mit dem Parametervektor θ_u und df_u Freiheitsgraden. Eine prüfbare Nullhypothese H_0 kann aus konjunktiv verknüpften Teilhypothesen H_0^i bestehen, die jeweils eine der beiden folgenden Behauptungen aufstellen:

- Gewisse Komponenten von θ_u sind gleich bestimmten Zahlen (Fixierung).
- Gewisse Komponenten von θ_u sind untereinander gleich (Gleichheitsrestriktion).

H_0 führt zu einem Modell M_e mit $df_e > df_u$ Freiheitsgraden. Wesentlich ist, daß M_e (eingeschränktes Modell) durch zusätzliche Forderungen an M_u (uneingeschränktes Modell) entsteht. Man sagt auch, M_e sei in M_u geschachtelt. Ist χ_u^2 die χ^2 -Fitstatistik für M_u und χ_e^2 die χ^2 -Fitstatistik für M_e , dann ist:

$$D^2 := \chi_e^2 - \chi_u^2$$

unter der H_0 approximativ χ^2 -verteilt mit $d := df_e - df_u$ Freiheitsgraden. Folglich ist die H_0 bei einem Test zum Niveau α genau dann abzulehnen, wenn D^2 größer als das $(1-\alpha)$ -Fraktile der χ^2 -Verteilung mit d Freiheitsgraden ist.

Für die Abhängigkeit der Teststärke des D^2 -Tests von der Stichprobengröße gelten die Erläuterungen in Abschnitt 2.5.

Die Voraussetzungen für den D^2 -Test bei Verwendung der ML- oder GLS-Schätzmethode:

- Die manifesten Variablen sind multivariat normalverteilt.
- Es wurde die Kovarianzmatrix analysiert.
- Das Modell M_u ist gültig.
- Die Stichprobe ist hinreichend groß

Beim Einsatz der WLS-Schätzmethode sind die beiden ersten Voraussetzungen durch die Forderung nach Verwendung einer korrekten Gewichtsmatrix W^{-1} zu ersetzen, und der erforderliche Stichprobenumfang ist deutlich größer.

Wie man leicht einsehen wird, ist z.B. auch der normale χ^2 -Modelltest von LISREL ein Spezialfall für das D^2 -Prinzip. Das uneingeschränkte Modell erlaubt in diesem Fall für jede empirische Kovarianz einen eigenen Parameter und hat daher Null Freiheitsgrade.

Da ein Modifikationsindex annähernd der Differenz der χ^2 -Prüfgrößen zu den Modellen mit und ohne die fragliche Fixierung bzw. Gleichheitsrestriktion entspricht und weil beim Aufheben der Fixierung bzw. Restriktion gerade ein Freiheitsgrad verloren geht, liegt unter Annahme der Gültigkeit des liberalisierten Modelles gerade ein Spezialfall für obiges D^2 -Prinzip mit $d = 1$ vor. Der Modifikationsindex hat im vorliegenden Spezialfall den Vorteil, daß er von LISREL in einem einzelnen Lauf (mit dem Modell M_e) ausgegeben werden kann, während ein D^2 -Test zwei LISREL-Läufe (mit den Modellen M_u und M_e) erfordert.

Es ist von Interesse, in welcher Beziehung die folgenden drei Statistiken zueinander stehen:

- der bei einem LISREL-Lauf mit dem Modell M_u berechnete T-Wert für den frei geschätzten Parameter ω ,
- der Modifikationsindex für den fixierten Parameter ω aus dem LISREL-Lauf mit dem Modell M_e und
- die D^2 -Teststatistik $\chi_e^2 - \chi_u^2$ aus den beiden beteiligten LISREL-Läufen.

Erfreulicherweise gilt asymptotisch (also wieder für $N \rightarrow \infty$) für den *quadrierten* T-Wert, den Modifikationsindex und die D^2 -Teststatistik:

$$T^2 = \text{Modifikationsindex} = D^2$$

Das Freisetzen eines Parameters ist natürlich nur dann sinnvoll, wenn dieser anschließend auch identifiziert ist. Allerdings muß der Anwender hier keine lästigen Kontrollen vornehmen: Wäre ein Parameter nach Freisetzung nicht identifiziert, so erhält er als Modifikationsindex eine 0. Auf diesem Wege kann man im übrigen den Identifikationsstatus für einen Parameter von LISREL prüfen lassen.

Bei der Verwendung der Modifikationsindikatoren sind folgende Punkte zu beachten:

- Modifikationsindizes werden Üblicherweise als bedeutsam betrachtet, wenn sie größer als 7 sind ($\approx 99\%$ - Fraktile der χ^2 -Verteilung mit $df = 1$).
- Pro Modifikationsschritt sollte nur ein Parameter freigesetzt werden. Das so erweiterte Modell produziert für die verbliebenen Fixierungen bzw. Restriktionen neue Modifikationsindikatoren, welche die besten Hinweise für weitere Liberalisierungen geben.
- Man sollte nur solche Parameter in das Modell aufnehmen, die theoretisch begründet werden können.

Eine Verbesserung der Fit-Beurteilung kann nicht nur durch Aufnahme zusätzlicher Parameter erreicht werden, sondern auch durch Fixieren bzw. Restringieren von bislang freien Parametern. Während durch eine solche Maßnahme die χ^2 -Statistik (hoffentlich nur wenig) ansteigt, wird ein Freiheitsgrad gewonnen, so daß die neue Überschreitungswahrscheinlichkeit des χ^2 -Wertes durchaus günstiger sein kann. Allerdings sollte jede eingeführte Fixierung oder Restriktion theoretisch begründbar sein.

Übungsaufgaben

Aufgabe 5.1 (Konstruiertes Beispiel)

Prüfen Sie das konstruierte Modell in Beispiel 1.3 anhand der in Abschnitt 5.1 enthaltenen Kovarianzmatrix. Dabei dürfen Sie die Identifikation des Modells ohne Beweis voraussetzen.

a) Führen Sie einen Programmlauf mit Verwendung der ULS-Schätzmethode und einen Lauf mit Verwendung der ML-Schätzmethode durch. Während die ML-Methode voreingestellt ist, müssen Sie die ULS-Methode durch die Spezifikation ME=UL im OU-Subkommando explizit anfordern (siehe Abschnitt 6.6.5). Vergleichen Sie Ihre Ergebnisse mit den in Abschnitt 5.1 wiedergegebenen wahren Werten des Modells, das die simulierten Daten erzeugt hat. Welche Schätzmethode ist überlegen? Wie ist das Ergebnis zu erklären?

b) Wenn Sie das Grundmodell akzeptieren können, prüfen Sie bitte die weitergehende Hypothese:

$$H_0: \gamma_{11} = \gamma_{12}$$

Hinweis: Sie können gemäß Abschnitt 5.2 eine χ^2 -Prüfgröße mit einem Freiheitsgrad durch Vergleich zweier geschachtelter Modelle gewinnen.

6 Beschreibung der SPSS-Prozeduren PRELIS und LISREL

6.1 Zur Integration von PRELIS und LISREL in das SPSS-System

PRELIS und LISREL 7 sind eigentlich zwei selbständige Computer-Programme, die von Karl Jöreskog und Dag Sörbom geschrieben wurden und von ihnen auch ständig weiterentwickelt werden. SPSS enthält zwei gleichnamige Prozeduren, welche die Benutzung der Originalprogramme über eine SPSS-konforme Kommandosprache und mit Zugriff auf die aktive SPSS-Systemdatei erlauben.

Für die Prozedur LISREL wurde dabei die Original-Syntax wegen ihrer weiten Verbreitung in der Literatur weitgehend übernommen, so daß die Subkommandos der SPSS-Prozedur LISREL genau den Kommandos bzw. Kommando-Zeilen der Original-Syntax entsprechen und sich infolgedessen stilistisch etwas von typischen SPSS-Subkommandos unterscheiden. Allerdings sind in der SPSS-Umgebung einige im Original-Handbuch (Jöreskog & Sörbom 1989) angegebene Syntax-Regeln obsolet (z.B. das "C" zur Ankündigung einer Fortsetzungszeile). Benutzer von SPSS-LISREL sollten sich daher bei Syntaxfragen am Handbuch "SPSS LISREL 7 and PRELIS" (SPSS 1990) orientieren; dieses Buch wurde in mehreren Exemplaren für die Lehrbuchsammlung beschafft. Andererseits ist das Original-Handbuch von Jöreskog und Sörbom als vollständige Dokumentation der statistischen Möglichkeiten von LISREL 7 auch für Benutzer der SPSS-Version sehr zu empfehlen.

Für die SPSS-Prozedur PRELIS wurde eine neue Benutzerschnittstelle mit Subkommandos im SPSS-Stil entworfen. Außerdem fehlen hier einige Optionen der PRELIS-Original-Version, die mit Hilfe des SPSS-Basissystems bequemer realisiert werden können (z.B. Variablentransformationen). Das eben zitierte Handbuch "SPSS LISREL 7 and PRELIS" (SPSS 1990) enthält zu PRELIS neben einer vollständigen Syntax-Beschreibung auch ausreichende Erläuterungen der statistischen Optionen.

Bei der Ausführung eines SPSS-Programms mit einem PRELIS- oder LISREL-Kommando erstellt SPSS aus den vom Benutzer eingegebenen Subkommandos eine Anweisungsfolge für PRELIS bzw. LISREL und übergibt diese zur Ausführung an das jeweilige Originalprogramm. Der Benutzer bemerkt diese Zusammenarbeit erst bei der Inspektion der Ergebnis-Liste: Hier findet er eine Begrüßungsmeldung des Originalprogramms, eine Auflistung der von SPSS produzierten Eingabezeilen für das Originalprogramm (z.B. überschrieben mit: "THE FOLLOWING PRELIS CONTROL LINES HAVE BEEN READ :") und natürlich die vom Originalprogramm produzierten Ergebnisse. Z.B. entsteht bei Ausführung des SPSS-Programms:

```
get  file='mc.sys'  
  
prelis  
  /variables=x1 to x4 (ordinal)  
  /type=polychor  
  /matrix=out('presi.pm')  
  /write=acov 'presi.acp'
```

die folgende Ausgabe:

```

16 Aug 93 SPSS-X RELEASE 4.1 FOR SIEMENS BS2000          Page  1
11:25:57 Version 4.1 (10.03.92)          SIEMENS H90-D    BS2000 V10

For BS2000 V10          Version 4.1 (10.03.92)          License Number 18079
This software is functional through November 30, 1993.

Try the new SPSS Release 4 features:
* LOGISTIC REGRESSION procedure          * CATEGORIES Option:
* EXAMINE procedure to explore data      *   conjoint analysis
* FLIP to transpose data files          *   correspondence analysis
* MATRIX Transformations Language       *   New LISREL and PRELIS Options
* GRAPH interface to SPSS Graphics      *
See the new SPSS documentation for more information on these new features.

  2  get file='mc.sys'
  3
File :B:$RZSPSS.MC.SYS
Created: 16 AUG 93 11:25:19 - 4 variables

  4  prelis
  5  /variables=x1 to x4 (ordinal)
  6  /type=polychor
  7  /matrix=out('presi.pm')
  8  /write=acov 'presi.acp'
Time stamp on saved file: 16 AUG 93 11:26:27
File contains 6 variables, 48 bytes per case before compression
There are 1,021,440 bytes of memory available.
The largest contiguous area has 1,020,840 bytes.

          P R E L I S      A PREPROCESSOR FOR LISREL

          PROGRAM VERSION 1.12          DISTRIBUTED BY

          SCIENTIFIC SOFTWARE, INC.
          1369 NEITZEL ROAD
          MOORESVILLE, INDIANA 46158
          (317) 831-6336

          THIS COPY AUTHORIZED FOR USE IN SPSS-X
          PROGRAM COPYRIGHT 1987-89 BY SCIENTIFIC SOFTWARE, INC.,
          (A MICHIGAN CORPORATION).
          DISTRIBUTION OR USE UNAUTHORIZED BY SCIENTIFIC SOFTWARE, INC. IS PROHIBITED.
          MVS - P R E L I S 1.12
          BY
          KARL G JORESKOG AND DAG SORBOM

THE FOLLOWING PRELIS CONTROL LINES HAVE BEEN READ :

SPSS-X RELEASE 4.1 FOR SIEMENS BS2000;
DA NI=4 NO=0 MI= -0.989898D+37 MC=1 TR=LI
LA
  X1      X2      X3      X4
RA FI=#T.1319.#RAW
OR X1
OR X2
OR X3
OR X4
OU MA=PM SM=#T.1319.#MAT          SA=PRESI.ACP

TOTAL SAMPLE SIZE = 200
CONVERSION OF ORIGINAL VALUES TO CATEGORIES
CATEGORY
VARIABLE          1          2          3          4          5          6
-----
X1          1.00    2.00    3.00    4.00    5.00
X2          1.00    2.00    3.00    4.00    5.00    6.00
X3          1.00    2.00    3.00    4.00    5.00
X4          1.00    2.00    3.00    4.00    5.00    6.00
          .
          .
          .

          ESTIMATED CORRELATION MATRIX
          X1          X2          X3          X4
-----
X1          1.000
X2          0.244          1.000
X3          0.132          0.231          1.000
X4          0.300          0.285          0.346          1.000
          THE PROBLEM USED          35464 BYTES (= 3.5% OF AVAILABLE WORKSPACE)

16 Aug 93 SPSS-X RELEASE 4.1 FOR SIEMENS BS2000          Page  2
11:26:30 Version 4.1 (10.03.92)          SIEMENS H90-D    BS2000 V10

Preceding task required 1.68 seconds CPU time; 4.00 seconds elapsed.

```

Die SPSS-Prozedur LISREL setzt wie jede andere SPSS-Prozedur voraus, daß durch ein vorangegangenes Datei-Definitions-Kommando eine aktive SPSS-Systemdatei definiert worden ist. Dies gilt auch dann, wenn LISREL seine Eingabedaten **nicht** aus der aktiven SPSS-Systemdatei übernimmt (vgl. Abschnitt 6.2). Nötigenfalls kann diese ohne tatsächliches Einlesen oder Generieren von Daten mit einem einfachen SPSS-Eingabeprogramm erstellt werden:

```
input program
numeric dummy
end file
end input program

lisrel
/Hypothetisches Modell aus Beispiel 1.3
/da ni=11 no=100
/matrix=in('const.cov')
.
.
.
/ou se tv mi rs ef mr
```

Für die Benutzer des SIEMENS-Rechners H90-D unter dem Betriebssystem BS2000 an der Universität Trier gilt, daß auch beim Arbeiten mit PRELIS und LISREL die integrierte Programmierumgebung aus SPSS und EDT mit Funktionstastenbedienung den größten Komfort bietet. Sie ist z.B. in der URT-Benutzereinführung Nr. 16 (Einführung in SPSS 4.1 unter BS2000) beschrieben.

6.2 Datenaustausch zwischen dem SPSS-Basissystem, PRELIS und LISREL

Die Integration von PRELIS und LISREL in das SPSS-System bringt u.a. den großen Vorteil, daß beide Programme auf die aktive SPSS-(Matrix-)Systemdatei zugreifen können. Damit wird die Übergabe von Daten an PRELIS bzw. LISREL erheblich einfacher und flexibler, weil viele Datensätze bereits als SPSS-Dateien vorliegen und weil die exzellenten Möglichkeiten des SPSS-Basissystems zum Einlesen und Modifizieren von Daten genutzt werden können.

Für die Übergabe der Daten an LISREL ist in der Regel folgende Strategie zu empfehlen:

- Einlesen der Daten, am besten aus einer passiven SPSS-Systemdatei unter Verwendung des GET-Kommandos. Dabei entsteht die aktive SPSS-Systemdatei des Programmlaufs. (Zu SPSS siehe z.B. Baltes-Götz 1993)
- Nötigenfalls Modifikation oder Neuberechnung von Variablen durch SPSS-Transformationskommandos (z.B. COMPUTE, RECODE)
- Erstellung der zu analysierenden Kovarianz-, Korrelations- oder Momentenmatrix durch PRELIS, Abspeichern als aktive SPSS-Matrix-Systemdatei
Falls eine der LISREL-Schätzmethoden WLS oder DWLS eingesetzt werden soll, muß zusätzlich die zugehörige Varianz-(Kovarianz-)Matrix in eine Textdatei geschrieben werden.
- Einlesen und Verarbeiten der PRELIS-Ausgaben durch LISREL

Diese Strategie wird in folgendem Programm demonstriert. Es handelt sich im wesentlichen um das Beispiel aus Abschnitt 4.3, wobei die dort verwendeten Anweisungen zum Erzeugen von Simulationsdaten durch ein einfaches GET-Kommando zum Einlesen einer SPSS-Systemdatei ersetzt wurden:

```

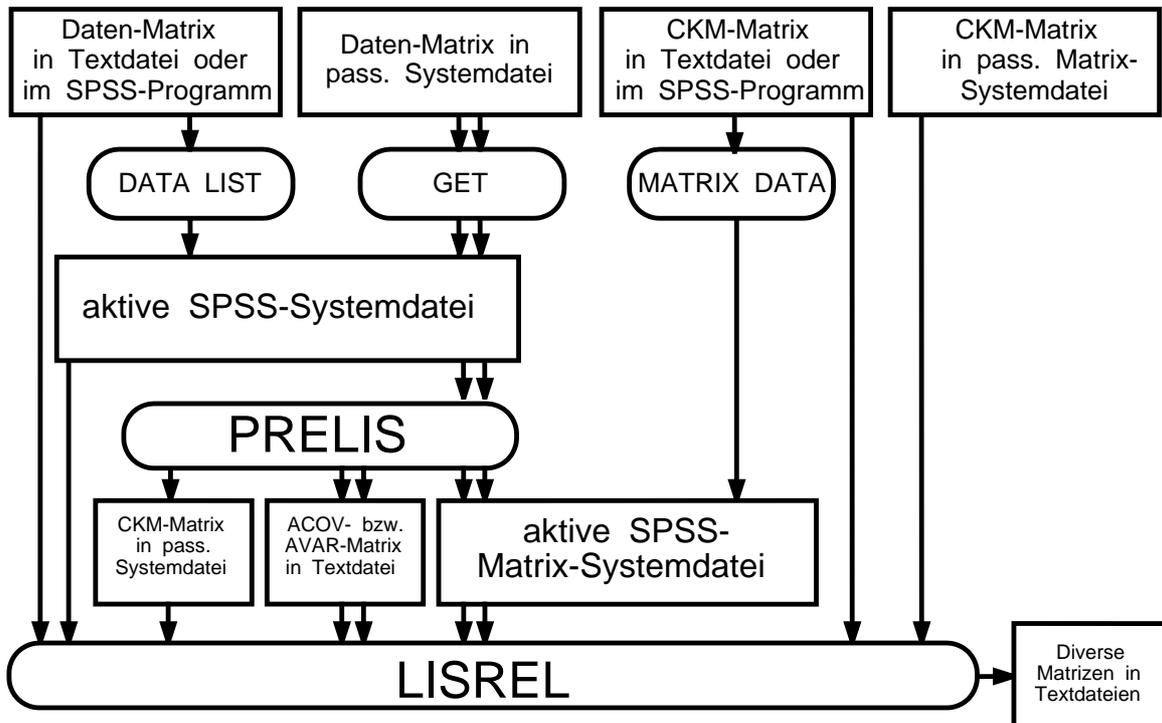
get file='presi.sys'

prelis
/variables=x1 to x4 (ordinal)
/type=polychor
/matrix=out(*)
/write=acov 'presi.acp'

lisrel
/"Analyse der Poly.-Kor.-Matr. aus ordinalen Indikatoren, WLS"
/da ni=4 ma=pm
/matrix=in(*)
/ac file='presi.acp'
/mo nx=4 nk=2 ph=st
/fr lx(1,1) lx(2,1) lx(3,2) lx(4,2)
/ou
    
```

Aus Gründen der Kompatibilität stehen auch die LISREL-eigenen Verfahren zum Einlesen von Daten zur Verfügung. Bei ihrer Anwendung muß man allerdings auf die Möglichkeiten des Preprozessors PRELIS (z.B. polychorische bzw. polyserielle Korrelationen) verzichten.

Die folgende Abbildung zeigt mögliche Transport- bzw. Verarbeitungssequenzen für Daten, wobei die oben empfohlene Strategie durch Doppelpfeile gekennzeichnet ist. Als "Daten-Matrix" wird eine rechteckige Matrix bezeichnet, deren Zeilen jeweils einem Fall und deren Spalten jeweils einer Variablen entsprechen. Die Abkürzung "CKM-Matrix" steht für "Kovarianz-, Korrelations- oder Momentenmatrix, eventuell ergänzt durch Mittelwerts- und/oder Standardabweichungs-Vektoren".



6.3 Allgemeine Syntax-Regeln für die SPSS-Prozeduren PRELIS und LISREL

Die folgenden allgemeinen Syntax-Regeln für die Prozedur-Kommandos PRELIS und LISREL stimmen weitgehend mit den generellen Regeln der SPSS-Kommandosprache überein:

- Ein Kommando besteht aus seinem Namen ("PRELIS" oder "LISREL") und seinen Spezifikationen:

kommandoname	spezifikationen
--------------	-----------------

Die Spezifikationen können enthalten:

- Subkommandonamen und andere Schlüsselwörter (z.B. "PAIRWISE" bei PRELIS, "FR" bei LISREL). Sie werden bei der Syntaxdarstellung groß geschrieben (siehe Abschnitt 6.4).
- Variablennamen
- Matrixnamen (bei LISREL)
- Zeichenfolgen (für Variablenetiketten und Dateinamen)
- Zahlen
- spezielle Begrenzungszeichen: / () = ' "

Zwischen diesen Elementen ist mindestens ein Leerzeichen erforderlich.

Ausnahme: Die speziellen Begrenzungszeichen sind selbstbegrenzend, d.h. vor und nach ihnen sind keine Leerzeichen nötig (aber erlaubt).

Statt eines Leerzeichens darf man meist auch mehrere Leerzeichen oder ein Komma verwenden.

- Ein Kommando muß in der ersten Spalte beginnen.
- Max. Länge einer Programmzeile: 80 Spalten
- Ein Kommando kann sich über beliebig viele Fortsetzungszeilen erstrecken. Diese müssen um mindestens eine Spalte eingerückt werden.
- Ein Kommando ist beendet, sobald eine neue Zeile in der ersten Spalte beschrieben ist. Folglich können nicht mehrere Kommandos in einer Zeile stehen.
- Die Verwendung von Groß- oder Kleinbuchstaben ist beliebig.
- Die Spezifikationen sind in Subkommandos unterteilt. Diese beginnen mit einem Subkommando-Namen und sind durch Schrägstriche voneinander getrennt. Schrägstriche vor dem ersten und nach dem letzten Subkommando sind überflüssig, werden aber toleriert.
Es ist der Übersichtlichkeit halber empfehlenswert, aber nicht vorgeschrieben, für jedes Subkommando eine neue Zeile zu beginnen.
- Bei PRELIS dürfen Schlüsselwörter bis auf die ersten drei Zeichen abgekürzt werden.
Beispiel: "con" für "continuous"
- Bei LISREL-Schlüsselwörtern sind nur die beiden ersten Buchstaben signifikant, so daß z.B. die folgenden Schreibweisen des Schlüsselwortes MODELPARAMETERS äquivalent sind:
MODELPARAMETERS MODEL mo model Model motel
Abweichend von der allgemeinen Regel sind drei Buchstaben signifikant:
 - beim Subkommandonamen "MATRIX",
 - beim Schlüsselwort "ALL" in den Subkommandos VALUE und START.

Für die LISREL-Subkommandos haben sich die Zwei-Buchstaben-Abkürzungen als Namen eingebürgert, man spricht als z.B. vom "MO-Subkommando" oder von der "MO-Zeile". In diesem Manuskript wird bei der Nennung von Subkommandonamen im Fließtext gelegentlich als mnemotechnische Hilfe in Klammern eine Namenserweiterung angegeben, z.B. bei der Erläuterung des MO(dellparameters)-Subkommandos.

6.4 Konventionen zur Beschreibung der PRELIS- und LISREL-Subkommandos

Zur Steuerung der SPSS-Prozeduren PRELIS und LISREL steht jeweils eine Menge von Subkommandos zur Verfügung. Dabei handelt es sich um Sätze einer künstlichen Sprache mit strengen syntaktischen Regeln. Wir müssen vereinbaren, wie diese Regeln zur Bildung zulässiger Sätze kompakt und eindeutig niedergeschrieben werden sollen. Etwas hochtrabend ausgedrückt vereinbaren wir dabei eine Metasprache zur Kommunikation über die Kommandosprache. In der folgenden Tabelle sind die in diesem Manuskript verwendeten Konventionen zur Beschreibung der PRELIS- und LISREL-Subkommandos enthalten:

(Meta)zeichen	Erläuterung	Beispiel
GROSS- BUCHSTABEN und Sonderzeichen (z.B. "=", "/")	Großbuchstaben und Sonderzeichen bezeichnen Konstanten, die in dieser Form vom Benutzer eingegeben werden müssen. Bei konkreten Kommandos ist allerdings Großschreibung i.a. nicht erforderlich.	Allgemeines Format: /VARIABLES=variablen-liste
kleinbuchstaben	Kleinbuchstaben bezeichnen Platzhalter, die bei der Eingabe vom Benutzer durch aktuelle Werte ersetzt werden müssen. Welche Werte zulässig sind, ist für jeden Platzhalter genau festgelegt.	Einzugeben ist z.B.: /variables=x1 to x4
[]	Spezifikationen in eckigen Klammern sind optional.	Allgemeines Format: /[VARIABLES=] variablen-liste Einzugeben ist z.B.: /variables=x1 to x4 oder: /x1 to x4
{ } ____	Alternativen, von denen genau eine gewählt werden muß, werden durch senkrechte Striche voneinander getrennt und insgesamt in geschweifte Klammern eingeschlossen. Bei einer optionalen Spezifikation mit Auswahlmöglichkeit ist diejenige Alternative unterstrichen, die bei Weglassung der Spezifikation in Kraft ist.	Allgemeines Format: [/MISSING={ <u>LISTWISE</u> PAIRWISE}] Einzugeben ist z.B.: /missing=pairwise oder: /missing=listwise Wegen der Voreinstellung ist die zweite Angabe allerdings redundant.
...	Drei Punkte bedeuten eine Wiederholung: mache weiter wie nach dem letzten Auftreten des Wortes bzw. Zeichens vor den Punkten	

6.5 Beschreibung der PRELIS-Subkommandos

6.5.1 Die PRELIS-Syntax im Überblick

```

PRELIS
[ /VARIABLES =] varliste [(CONTINUOUS | ORDINAL | CABOVE | CBELOW | CENSORED)]
    [varliste ...] ]
[ /MISSING = {LISTWISE | PAIRWISE} ] [{EXCLUDE | INCLUDE}] ]
[ /MAXCAT = {0 | max} ]
[ /TYPE = {NONE | AUGMENTED | CORRELATION | COVARIANCE | MOMENT | OPTIMAL | POLYCHOR} ]
[ /MATRIX = {OUT('PRLS#MA.SPSSFILE') | OUT(* | 'dateiname')} | NONE] ]
[ /WRITE = {NONE | ACOV [{'PRLS#ASY.DATA'} | 'dateiname']} | AVAR 'dateiname'} ]
[ /PRINT = {NONE | ACOV | AVAR} ] [KURTOSIS] [XBIVARIATE] [XTEST] ]
[ /CRITERIA = [ASIZE (n)] [DEFAULT] ]

```

Regel: Die Reihenfolge der Subkommandos ist beliebig.
Ausnahme: Falls der Subkommando-Name VARIABLES weggelassen wird, muß die Variablenliste unmittelbar hinter dem Kommandonamen PRELIS stehen.

Hinweis: Die einzelnen Subkommandos werden anschließend in eigenen Abschnitten erläutert.

Beispiel:

```

prelis
  /variables=x1 to x4 (ordinal)
  /type=polychor
  /matrix=out(*)
  /write=acov 'presi.acp'

```

Dieses Beispiel wurde in Abschnitt 4.3 erläutert.

6.5.2 Das VARIABLES-Subkommando

Mit diesem Subkommando kann man:

- Variablen der aktiven SPSS-Systemdatei für den PRELIS-Lauf auswählen.
- Die Reihenfolge der Variablen ändern.
Dies ist relevant, wenn PRELIS-Ergebnisse später von LISREL verarbeitet werden sollen, weil LISREL die Y-Variablen vor den X-Variablen erwartet.
- Den Skalentyp von Variablen festlegen.

```

/[VARIABLES =] varliste [(CONTINUOUS | ORDINAL | CABOVE | CBELOW | CENSORED)]
    [varliste ...]

```

varliste Es können numerische Variablen aus der aktiven SPSS-Systemdatei nach den Regeln der SPSS-Syntax für Variablenlisten angegeben werden. Insbesondere ist also die aufrufende TO-Konvention erlaubt.
Ist eine Weiterverarbeitung der PRELIS-Ergebnisse durch LISREL geplant, dann sollten die Y- vor den X-Variablen aufgelistet werden.

CONTINUOUS	Die Variablen der vorangegangenen Liste werden als intervall- bzw. rationalskaliert deklariert.
ORDINAL	Die Variablen der vorangegangenen Liste werden als ordinalskaliert deklariert.
CABOVE, CBELOW, CENSORED	Die Variablen der vorangegangenen Liste werden als oben-zensiert und/oder unten-zensiert deklariert.

Regeln:

- Wird das VARIABLES-Subkommando weggelassen, dann werden alle numerischen Variablen der aktiven Systemdatei einbezogen.
Steht das VARIABLES-Subkommando am Anfang des PRELIS-Aufrufs, dann darf "VARIABLES =" weggelassen werden.
- Die Zuweisung eines Skalentyps bezieht sich auf alle Elemente der vorangegangenen Variablenliste.
- Wird einer Variablen kein Skalentyp explizit zugewiesen, so erfolgt eine automatische Klassifikation in Abhängigkeit von der Anzahl v ihrer verschiedenen Werte und von der aktuellen MAXCAT-Einstellung (max) (siehe PRELIS-Subkommando MAXCAT):

$v > max \Rightarrow$ Variable wird als kontinuierlich behandelt

$v \leq max \Rightarrow$ Variable wird als ordinal behandelt

Ein zensierter Typ (CABOVE, CBELOW, CENSORED) kann also *nicht* automatisch zugewiesen werden.

Die Festlegung des Skalentyps im VARIABLES-Subkommando hat Priorität gegenüber der automatischen Zuweisung. Allerdings werden Variablen mit mehr als 15 verschiedenen Werten auf jeden Fall als kontinuierlich behandelt.

Hinweis:

Nähere Hinweise zu den verschiedenen Skalentypen und zu den jeweils voreingestellten bzw. wählbaren Assoziationsmaßen finden sich in den Abschnitten 4.1 und 6.5.5.

Beispiel:

```
/var = income (con) ses1 to ses4 (ord) iq
```

Falls bei den Variablen SES1 bis SES4 weniger, bei der Variablen IQ hingegen mehr als 15 verschiedene Werte auftreten, führt dieses Subkommando zu folgender Typ-Zuordnung:

INCOME, IQ	:	kontinuierlich
SES1, SES2, SES3, SES4	:	ordinal

6.5.3 Das MISSING-Subkommando

Dieses Subkommando regelt die Behandlung von Missing-Data-Indikatoren (MD-Indikatoren).

```
/MISSING = [{LISTWISE | PAIRWISE}] [{EXCLUDE | INCLUDE}]
```

<u>LISTWISE</u>	Fallweiser Ausschluß fehlender Werte: In die Analyse werden nur Fälle mit validen Werten für <i>alle</i> im VARIABLES-Subkommando aufgeführten Variablen einbezogen.
<u>PAIRWISE</u>	<p>Paarweiser Ausschluß fehlender Werte: Bei der Berechnung einer Kovarianz oder eines anderen Assoziationsmaßes werden alle Fälle mit validen Werten für <i>die beiden beteiligten</i> Variablen einbezogen.</p> <p>Damit basieren die Elemente einer Assoziations-Matrix im allgemeinen auf unterschiedlichen Teilstichproben, so daß die Matrix eventuell nicht positiv-semidefinit ist. Dies kann zu Problemen bei der Parameterschätzung mit LISREL führen.</p>
<u>EXCLUDE</u>	Ausschluß des automatischen MD-Indikators (SYSMIS) und der benutzerdefinierten MD-Indikatoren von allen Berechnungen
<u>INCLUDE</u>	<p>Die benutzerdefinierten MD-Indikatoren werden wie valide Werte in die Berechnungen einbezogen, der automatische MD-Indikator (SYSMIS) bleibt ausgeschlossen.</p> <p>Vorsicht: Wenn z.B. die Variable GEW das Körpergewicht von Menschen beinhaltet und fehlende Werte mit 999.9 kodiert wurden, dann bewirkt "/missing = include", daß der Wert 999.9 bei der Variablen GEW wie jede echte Gewichtsangabe verrechnet wird, was kaum jemals sinnvoll sein dürfte.</p>

Beispiel: /missing pairwise
 Es wird eine paarweise Behandlung fehlender Werte angefordert.

6.5.4 Das MAXCAT-Subkommando

Wird einer Variablen im VARIABLES-Subkommando kein Skalentyp explizit zugewiesen, so erfolgt eine automatische Klassifikation in Abhängigkeit von der Anzahl v ihrer verschiedenen Werte und der aktuellen MAXCAT-Einstellung (max) von PRELIS:

$v > max \Rightarrow$ Variable wird als kontinuierlich behandelt
 $v \leq max \Rightarrow$ Variable wird als ordinal behandelt

Die Festlegung des Skalentyps im VARIABLES-Subkommando hat Priorität gegenüber der automatischen Zuweisung. Allerdings werden Variablen mit mehr als 15 verschiedenen Werten auf jeden Fall als kontinuierlich behandelt.

```
/MAXCAT = {0 | max}
```

max Erlaubte Werte: $0 \leq max \leq 15$

Beispiel: /maxcat=9

6.5.5 Das TYPE-Subkommando

Dieses Subkommando regelt den Typ der zu erstellenden Assoziations-Matrix. Das bei einem konkreten Variablenpaar verwendete Berechnungsverfahren hängt von den beiden Variablentypen (kontinuierlich, ordinal, zensiert) ab (vgl. Abschnitt 4.1). Zur Vereinfachung der folgenden Darstellung soll schon hier daran erinnert werden, daß zensierte Variablen (CABOVE, CBELOW, CENSORED) nach Ersetzung der Ränder durch Normal-Scores wie kontinuierliche Variablen behandelt werden (vgl. Abschnitte 4.1.1.3 und 4.1.2.1).

```
/TYPE = {CORRELATION | OPTIMAL | POLYCHOR | COVARIANCE | MOMENT | AUGMENTED | NONE}
```

CORRELATION Es wird eine Matrix mit Produkt-Moment-Korrelationen berechnet, wobei jedoch zuvor die Werte ordinal-deklarerter Variablen durch Normal-Scores (vgl. Abschnitt 4.1.2.1) ersetzt werden.

OPTIMAL Es wird eine Matrix mit Produkt-Moment-Korrelationen berechnet, wobei jedoch zuvor die Werte ordinal-deklarerter Variablen durch Korrelations-maximierende Optimal-Scores (vgl. Abschnitt 4.1.2.2) ersetzt werden. Welcher Koeffizient für ein konkretes Variablenpaar (Z_1, Z_2) berechnet wird, zeigt folgende Tabelle:

Deklariertes Typ von Z_1	Deklariertes Typ von Z_2	
	kontinuierlich	ordinal
kontinuierlich	Produkt-Moment-Korrelation der Rohdaten	einfache, konsistente Schätzung der polyserialen Korrelation unter Voraussetzung bivariater Normalverteilung der zugrundeliegenden Variablen
ordinal	einfache, konsistente Schätzung der polyserialen Korrelation unter Voraussetzung bivariater Normalverteilung der zugrundeliegenden Variablen	Produkt-Moment-Korrelation der optimalen Scores zu beiden Variablen

POLYCHOR Es wird eine Korrelationsmatrix mit polychorischen bzw. polyserialen Korrelationen bei Beteiligung ordinalskalierten Variablen berechnet. Welcher Koeffizient für ein konkretes Variablenpaar (Z_1, Z_2) berechnet wird, zeigt folgende Tabelle:

Deklariertes Typ von Z_1	Deklariertes Typ von Z_2	
	kontinuierlich	ordinal
kontinuierlich	Produkt-Moment-Korrelation der Rohdaten	Polyseriale Korrelation, nach dem ML-Prinzip geschätzt unter Voraussetzung bivariater Normalverteilung der zugrundeliegenden Variablen
ordinal	Polyseriale Korrelation, nach dem ML-Prinzip geschätzt unter Voraussetzung bivariater Normalverteilung der zugrundeliegenden Variablen	Polychorische Korrelation, nach dem ML-Prinzip geschätzt unter Voraussetzung bivariater Normalverteilung der zugrundeliegenden Variablen

COVARIANCE	Es wird eine Kovarianzmatrix berechnet, wobei jedoch zuvor die Werte ordinal-deklarerter Variablen durch Normal-Scores (vgl. Abschnitt 4.1.2.1) ersetzt werden.
MOMENT	Es wird eine Matrix der mittleren Produkte und Kreuzprodukte (vgl. Abschnitt 4.1.2.4) berechnet, wobei jedoch zuvor die Werte ordinal-deklarerter Variablen durch Normal-Scores (vgl. Abschnitt 4.1.2.1) ersetzt werden.
AUGMENTED	Erweiterte Momentenmatrix (augmented moment matrix) Sie ist definiert als die Momentenmatrix zu der rechtsseitig um die Spalte $(1,1,\dots,1)^T$ erweiterten Datenmatrix.
NONE	PRELIS berechnet keine Assoziationsmaße, sondern beschränkt sich auf eine Analyse der univariaten Verteilungen.
Hinweis:	Die verschiedenen Assoziationsmaße werden in Abschnitt 4.1 erläutert.
Beispiel:	<code>/type=polychor</code>

6.5.6 Das MATRIX-Subkommando

Mit diesem Subkommando wird festgelegt, wohin eine entsprechend der TYPE-Spezifikation erstellte Korrelations- bzw. Kovarianz- bzw. Momentenmatrix geschrieben werden soll. PRELIS erzeugt dabei eine SPSS-Matrix-Systemdatei, die durch ein anschließendes LISREL-Kommando sehr bequem eingelesen werden kann.

```
/MATRIX = {OUT('PRLS#MA.SPSSFILE') | OUT(* | 'dateiname' ) | NONE}
```

OUT(...)	Es wird eine SPSS-Matrix-Systemdatei erstellt. Diese wird per Voreinstellung auf der Festplatte unter einem Standardnamen abgelegt, der vom Betriebssystem abhängt ¹ . Alternativ können folgende Ausgabeziele gewählt werden:
*	Die aktive Systemdatei des aktuellen SPSS-Laufs. Deren vorheriger Inhalt wird also ersetzt.
dateiname	Eine beliebige Benutzerdatei auf der Festplatte
NONE	Es wird keine Matrix-Systemdatei erstellt.
Hinweis:	Bei paarweisem Ausschluß fehlender Werte enthält die Matrix-Systemdatei eine Fallzahl für jeden einzelnen Koeffizienten. Bei Verarbeitung mit LISREL wird im DA(taparameters)-Subkommando der NO(bservations)-Wert automatisch auf die kleinste auftretende bivariate Fallzahl eingestellt.
Beispiel:	<code>/matrix=out('presi.pm')</code>

¹ Unter dem Betriebssystem BS2000 wird per Voreinstellung in die Datei PRLS#MA.SPSSFILE geschrieben.

6.5.7 Das WRITE-Subkommando

Mit diesem Subkommando kann PRELIS angewiesen werden, die Matrix mit den asymptotischen (Ko)varianzen der Stichproben-Assoziationen ((Ko)varianzen bzw. -Korrelationen der manifesten Variablen) zu schätzen und in eine Textdatei zu schreiben. Diese ist vorgesehen für die anschließende Weiterverarbeitung durch LISREL und ermöglicht dann die Anwendung der asymptotisch verteilungsfreien LISREL-Schätzmethoden WLS bzw. DWLS (vgl. Abschnitt 4.2).

Eine Schätzung der asymptotischen (Ko)varianzen der Assoziations-Statistiken ist nur unter folgenden Voraussetzungen möglich:

- Fallweiser Ausschluß fehlender Werte
- TYPE ist nicht gleich OPTIMAL oder MOMENT oder AUGMENTED
- Die Stichprobe ist hinreichend groß (siehe PRELIS-Subkommando CRITERIA)

```
/WRITE = {NONE | ACOV [{'PRLS#ASY.DATA' | 'dateiname'}] | AVAR 'dateiname' }
```

<u>NONE</u>	Per Voreinstellung wird <i>keine</i> asymptotische (Ko)varianzmatrix der Assoziations-Statistiken erzeugt.
ACOV	Es wird eine Matrix mit den geschätzten asymptotischen Varianzen und Kovarianzen der Assoziations-Statistiken erzeugt. Diese wird per Voreinstellung auf der Festplatte unter einem Standardnamen abgelegt, der vom Betriebssystem abhängt ¹
dateiname	Die ACOV-Matrix wird in die angegebene Benutzerdatei geschrieben.
AVAR	Es wird eine Matrix mit den geschätzten asymptotischen Varianzen der Assoziations-Statistiken erzeugt.
dateiname	Die AVAR-Matrix wird in die angegebene Benutzerdatei geschrieben. Fehlt die Angabe eines Dateinamens, so erzeugt PRELIS statt einer AVAR-Matrix eine ACOV-Matrix und legt diese unter dem Standardnamen für ACOV-Matrizen ab (siehe oben).
Hinweis:	Die erzeugte Datei ist <i>keine</i> SPSS-Matrix-Systemdatei (vgl. MATRIX-Subkommando), sondern eine konventionelle Textdatei.
Beispiel:	<code>/write acov 'presi.acp'</code>

¹ Unter dem Betriebssystem BS2000 wird per Voreinstellung in die Datei PRLS#ASY.DATA geschrieben.

6.5.8 Das PRINT-Subkommando

Mit diesem Subkommando können optionale Ausgaben angefordert (z.B. ein Maß für die multivariate Schiefe) oder Standardausgaben unterdrückt werden (z.B. die bivariaten Kontingenztabelle für Paare ordinaler Variablen).

```
/PRINT = [{ACOV | AVAR | NONE}] [KURTOSIS] [XBIVARIATE] [XTEST]
```

ACOV	Zusätzliche Ausgabe einer Matrix mit den asymptotischen Varianzen und Kovarianzen der Assoziations-Statistiken (vgl. Abschnitt 4.2).
AVAR	Zusätzliche Ausgabe einer Matrix mit den asymptotischen Varianzen der Assoziations-Statistiken (vgl. Abschnitt 4.2).
NONE	Verkürzte Ausgabe. Z.B. gibt PRELIS bei ordinalen Variablen keine bivariaten Kontingenztabelle sowie keine Teststatistiken für polychorische und polyseriale Korrelationen aus.
KURTOSIS	Zusätzliche Ausgabe eines Maßes für die multivariate Schiefe der gemeinsamen Verteilung aller Variablen (vgl. Mardia 1970). Es ist nur verfügbar bei TYPE=COVARIANCE und MISSING=LISTWISE.
XBIVARIATE	Bei ordinalen Variablen wird die Ausgabe der bivariaten Kontingenztabelle unterdrückt.
XTEST	Die Ausgabe von Teststatistiken für polychorische und polyseriale Korrelationen wird unterdrückt.
Regeln:	<ul style="list-style-type: none"> - Eine ACOV- oder AVAR-Matrix kann nur unter folgenden Voraussetzungen erstellt werden: <ul style="list-style-type: none"> - Fallweiser Ausschluß fehlender Werte - TYPE ist nicht gleich OPTIMAL oder MOMENT oder AUGMENTED - Die Stichprobe ist hinreichend groß (siehe Subkommando CRITERIA) - Die ACOV-Matrix kann nur dann ausgegeben werden, wenn im PRELIS-Aufruf gleichzeitig ein WRITE-Subkommando mit ACOV-Anforderung vorhanden ist. Analoges gilt für die Ausgabe der AVAR-Matrix. <p>Bei einem inkompatiblen WRITE-Kommando unterbleibt die gewünschte ACOV- bzw.- AVAR-Ausgabe. Ein fehlendes WRITE-Kommando wird von SPSS automatisch passend erstellt.</p>

Beispiel:

```
prelis /var = x1 x2 (ord) x3 x4 (con)
      /type=polychor /write = acov ('presi.acv')
      /print acov kurt
```

6.5.9 Das CRITERIA-Subkommando

PRELIS berechnet asymptotische (Ko)varianzen der geschätzten Variablen-Korrelationen bzw. -Kovarianzen nur, wenn folgende minimale Fallzahl gegeben ist (bei MISSING=LISTWISE):

- bei $k < 12$ Variablen: 200
- bei $k \geq 12$ Variablen: $1.5k(k+1)$.

Mit dem CRITERIA-Subkommando kann diese Restriktion gelockert werden, wobei aber Vorsicht geboten ist, weil asymptotische (Ko)varianzen aus kleinen Stichproben potentiell unzuverlässig sind. In dieser Situation kann der Verzicht auf die asymptotisch verteilungsfreien LISREL-Schätzmethoden WLS und DWLS das kleinere Übel sein.

`/CRITERIA = ASIZE (n)`

ASIZE Erlaubt die Festsetzung der minimalen Fallzahl für die Berechnung asymptotischer (Ko)varianzen für die Assoziations-Statistiken.

n Voreinstellung: siehe oben

Beispiel: `/crit asize(197)`

6.6 Beschreibung der LISREL-Subkommandos

6.6.1 Die LISREL-Syntax im Überblick

Ein vollständiges LISREL-Kommando besteht aus vier Blöcken von Subkommandos. Dies wird im folgenden unter Verwendung des Beispiels aus Abschnitt 4.3 verdeutlicht:

Block	Beispiel
Titelzeile(n)	lisrel /"Analyse der Poly.-Kor.-Matr. aus ordinalen Indikatoren, WLS"
⇓	
Daten-Spezifikation	/da ni=4 ma=pm /matrix=in(*) /ac file='presi.acp'
⇓	
Modell-Spezifikation	/mo nx=4 nk=2 ph=st /fr lx(1,1) lx(2,1) lx(3,2) lx(4,2)
⇓	
Spezifikation der Ausgabe und der Schätzmethode	/ou

Allgemeine Regeln für die Anordnung der Subkommandos:

- Die eben dargestellte Abfolge der vier Subkommando-Blöcke muß eingehalten werden.
- Der erste Block (mit den Titelzeilen) ist optional, alle übrigen Blöcke müssen vorhanden sein.
- Der Daten-Spezifikations-Block muß mit dem DA(ta)-Subkommando (s.u.) beginnen.
- Der Modell-Spezifikationsblock muß mit dem MO(dell)-Subkommando (s.u.) beginnen.
- Der letzte Block besteht nur aus dem OU(tput)-Subkommando.
- Innerhalb des Daten- bzw. Modell-Spezifikations-Blocks können **nach** dem eröffnenden DA- bzw. MO-Subkommando die übrigen Subkommandos beliebig angeordnet und auch wiederholt werden, wobei im Falle konkurrierender Bestimmungen spätere Subkommandos die vorangegangenen dominieren.

Allgemeine Regeln für die Struktur eines Subkommandos:

- Subkommandos bestehen aus Steuerparametern, die vom Subkommando-Namen und untereinander durch mindestens ein Leerzeichen zu trennen sind. In obigem Beispiel für ein DA-Subkommando treten z.B. die Steuer-Parameter NI und MA auf:
/da ni=4 ma=pm
- Die meisten Steuerparameter haben Voreinstellungen und sind daher optional.

- Es gibt zwei Typen von Steuerparametern:
 - Wert-Parameter
Sie haben einen Parameter-Namen und eine Liste erlaubter Parameter-Werte. Die Parameterspezifikation besteht in der Zuweisung eines Wertes zum Parameternamen mit Hilfe des Gleichheitszeichens. In obigem Beispielprogramm wird im DA-Subkommando die Anzahl der manifesten Variablen vereinbart, indem der Wert 4 dem Parameter NI zugewiesen wird.
 - Logische Parameter
Sie können nur zwei Werte annehmen. Daher erfolgt hier keine Wertzuweisung per Gleichheitszeichen, sondern bei Weglassen des Parameter-Namens gilt der voreingestellte Wert und bei Auftreten des Namens die Alternative. Im OU(tput)-Subkommando kann z.B. die Ausgabe von Standardfehlern zu den geschätzten Parametern angefordert werden durch Angabe des betreffenden Parameter-Namens:
/OU SE
Per Voreinstellung werden keine Standardfehler ausgegeben.
- Die Reihenfolge der Steuerparameter ist beliebig.

6.6.2 Titelzeilen

Man kann einem LISREL-Aufruf beliebig viele Titelzeilen voranstellen. Sie erscheinen am Anfang der Ergebnisausgabe. Die erste Titelzeile erscheint zusätzlich im Kopf jeder Ausgabeseite.

```
/"titelzeile"  
/...
```

- Regel:** Titelzeilen dürfen *nicht* mit einer der folgenden Zeichenfolgen beginnen:
"DA", "Da", "Da", "da"
Dies muß auch nach Streichung führender Leerzeichen gelten. Verboten sind also
z.B.: - /"Daten aus Diplomarbeit"
- /" Datum: 1.01.94"
- Hinweis:** Wenn die Anführungszeichen weggelassen werden, übernimmt LISREL beim Einlesen der Daten aus einer per MATRIX DATA erstellten aktiven SPSS-Matrix-Systemdatei (vgl. Abschnitt 6.6.3.2) *nicht* die SPSS-Variablenamen als Etiketten für die manifesten Variablen.
- Beispiel:**
- ```
lisrel
/"Konfirmatorische Faktorenanalyse mit acht Intelligenzmaßen"
/"und drei Kreativitaetstests"
/"Schaetzmethode: ML"
/da ni=11
.
.
```

### 6.6.3 Daten-Spezifikation

Im Daten-Spezifikations-Block können folgende Subkommandos zum Beschreiben und Einlesen der Daten verwendet werden:

|        |                                                                                                                             |
|--------|-----------------------------------------------------------------------------------------------------------------------------|
| DA     | Deklaration globaler Merkmale der Daten<br>Das Subkommando leitet den Daten-Spezifikations-Block ein und ist obligatorisch. |
| MATRIX | Einlesen von Daten aus einer SPSS-Matrix-Systemdatei                                                                        |
| RA     | Einlesen von Rohdaten aus einer externen Textdatei oder aus dem Eingabestrom                                                |
| CM     | Einlesen einer Kovarianzmatrix                                                                                              |
| KM     | Einlesen einer Korrelationsmatrix                                                                                           |
| OM     | Einlesen einer Korrelationsmatrix aus Optimal-Scores                                                                        |
| PM     | Einlesen einer Matrix mit polychorischen und/oder polyserialen Korrelationen                                                |
| MM     | Einlesen einer Momentenmatrix                                                                                               |
| ME     | Einlesen von Mittelwerten                                                                                                   |
| SD     | Einlesen von Standardabweichungen                                                                                           |
| AC     | Einlesen der asymptotischen Kovarianzen zu den Assoziations-Statistiken                                                     |
| AV     | Einlesen der asymptotischen Varianzen zu den Assoziations-Statistiken                                                       |
| LA     | Einlesen von Etiketten für die manifesten Variablen                                                                         |
| SE     | Auswählen und Umordnen von Variablen für die Analyse                                                                        |

Regeln für den Daten-Spezifikations-Block:

- **Wichtig:** LISREL geht per Voreinstellung beim Lesen der Eingabedaten davon aus, daß die Variablen in folgender Reihenfolge vorliegen:
  1. Y-Variable, 2. Y-Variable, ..., letzte Y-Variable,
  1. X-Variable, 2. X-Variable, ..., letzte X-Variable.
 Dies gilt sowohl beim Lesen von Rohdaten (Variablen-Werten) als auch beim Lesen einer Assoziationsmatrix.  
 Mit Hilfe des SE(lect)-Subkommandos (s.u.) kann man diese Voreinstellung abändern, wenn anders angeordnete Daten gelesen werden sollen: Man kann Variablen von der Analyse ausschließen und die Aufteilung in Y- und X-Variablen neu festlegen, so daß eine Reorganisation von ungeeignet angeordneten Eingabedaten *nicht* erforderlich ist.
- Innerhalb des Daten-Spezifikations-Blocks können nach dem einleitenden und obligatorischen DA-Subkommando die erforderlichen weiteren Subkommandos beliebig angeordnet werden. Welche Subkommandos benötigt werden, hängt von den Angaben im DA-Subkommando (z.B. Analyse einer Kovarianz- oder einer Korrelationsmatrix) und von der Organisation der Eingabedaten (z.B. Variablen-Werte versus Kovarianzmatrix) ab.  
 Wenn der Daten-Spezifikations-Block lediglich aus dem DA-Subkommando besteht, also kein Subkommando zum Einlesen von Daten enthält, so werden die  $k$  ersten numerischen Variablen in der aktiven SPSS-Systemdatei herangezogen. Dabei kann es sich um eine "normale", Variablen-orientierte Systemdatei handeln oder um eine, z.B. von PRELIS erstellte, Matrix-Systemdatei. Hinsichtlich der Variablen-Reihenfolge ist der erste Hinweis zu beachten.

### 6.6.3.1 Das DA(ta)-Subkommando

Das DA(ta)-Subkommando ist obligatorisch und leitet den Daten-Spezifikations-Block ein. Hier werden globale Merkmale der Daten deklariert.

```
/DA [NG={1 | g}] NI=k [NO={0 | n}] [MA={CM | KM | MM | AM | OM | PM}] [XM=md]
```

|    |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |
|----|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| NG | <p>Anzahl der Gruppen (number of groups)<br/>Die Möglichkeiten von LISREL zur Analyse mehrerer Gruppen (z.B. Vergleich von Faktorladungsmatrizen) werden in diesem Manuskript nicht behandelt.</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
| NI | <p>Anzahl der Variablen im einzulesenden Datensatz (number of input variables)<br/>Für die Analyse kann später eine Teilmenge der Eingabevariablen ausgewählt werden (vgl. SE-Subkommando).</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |
| k  | <p>Der erlaubte Wertebereich ist nur durch die Rechnerkapazität beschränkt.</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |
| NO | <p>Anzahl der Fälle (number of observations)</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |
| n  | <p>Die Voreinstellung 0 kann beibehalten werden, wenn die Daten aus der (aktiven) SPSS-(Matrix)-Systemdatei oder aus einer externen Textdatei eingelesen werden, wobei LISREL die tatsächliche Fallzahl übernimmt bzw. ermittelt.<br/>Eine Angabe ist nur erforderlich, wenn</p> <ul style="list-style-type: none"> <li>- Rohdaten (Variablenwerte) eingelesen werden sollen, die innerhalb des LISREL-Aufrufs abgelegt sind,</li> <li>- eine Matrix mit Assoziationsmaßen eingelesen werden soll, die innerhalb des LISREL-Aufrufs oder in einer externen Textdatei abgelegt ist.</li> </ul> <p>Beim Einlesen von Variablenwerten aus der aktiven SPSS-Systemdatei oder aus einer externen Textdatei kann man mit der NO-Angabe dafür sorgen, daß nur die ersten n Fälle in die Analyse einbezogen werden. Sind die Rohdaten innerhalb des LISREL-Aufrufs abgelegt, dann muß die Anzahl der vorhandenen Fälle angegeben werden.</p> |
| MA | <p>Typ der <i>zu analysierenden</i> Matrix<br/>Die eingelesene und die zu analysierende Matrix können von verschiedenem Typ sein. Z.B. kann man eine Korrelationsmatrix und den Vektor mit den zugehörigen Standardabweichungen einlesen, um eine Kovarianzmatrix zu analysieren.<br/>Da LISREL als Mittelwert bzw. Standardabweichung für eine manifeste Variable gegebenenfalls die voreingestellten Werte 0 bzw. 1 benutzt, kann aus jeder eingelesenen Matrix mit Assoziationsmaßen jede zu analysierende Matrix gebildet werden. Falls die voreingestellten Mittelwerte und Standardabweichungen nicht verwendet werden sollen, müssen empirische Werte eingelesen werden.<br/>Die folgende Tabelle listet mögliche Pakete von Eingabedaten in Abhängigkeit vom Typ der zu analysierenden Matrix auf:</p>                                                                                                                       |

| Zu analysierende Matrix | Sinnvolle Pakete von Eingabedaten                                                                                                                                                                                                                                                        |
|-------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| CM                      | - Kovarianzmatrix (MATRIX, CM)<br>- Korrelationsmatrix (MATRIX, KM) + Vektor der Standardabweichungen (MATRIX, SD)<br>- Rohdaten (RA)                                                                                                                                                    |
| KM                      | - Korrelationsmatrix (MATRIX, KM)<br>- Rohdaten (RA)                                                                                                                                                                                                                                     |
| OM, PM                  | - Korrelationsmatrix (MATRIX, OM, PM)                                                                                                                                                                                                                                                    |
| MM, AM                  | - Kovarianzmatrix (MATRIX, CM) + Vektor der Mittelwerte (MATRIX, ME)<br>- Momentenmatrix (MATRIX, MM) + Vektor der Mittelwerte (MATRIX, ME)<br>- Korrelationsmatrix (MATRIX, KM) + Vektor der Standardabweichungen (MATRIX, SD) + Vektor der Mittelwerte (MATRIX, ME)<br>- Rohdaten (RA) |

Hinweise: - In der rechten Spalte der Tabelle ist zu den Eingabedaten jeweils in Klammern angegeben, mit welchen LISREL-Subkommandos des Daten-Spezifikations-Blocks sie eingelesen werden können.  
- Bei Einsatz der LISREL-Schätzmethoden WLS oder DWLS müssen zusätzlich zu den in der Tabelle angegebenen Daten asymptotische (Ko)varianzen der Assoziations-Statistiken eingelesen werden (vgl. Subkommandos AC und AV).

CM

Kovarianzmatrix

KM

Matrix mit Produkt-Moment-Korrelationen, basierend auf Rohwerten oder Normal-Scores

OM

Korrelationsmatrix, basierend auf Optimal-Scores

PM

Matrix mit polychorischen und/oder polyserialen Korrelationen

MM, AM

Momentenmatrix (moment matrix) und erweiterte Momentenmatrix (augmented moment matrix)

Diese Optionen sind wegen der Erweiterung des LISREL 7 - Modells um Parametervektoren für Ordinatenabschnitte und Erwartungswerte latenter Variablen überflüssig und nur aus Gründen der Kompatibilität zu LISREL VI weiter verfügbar. LISREL-Modelle mit Ordinatenabschnitten und Mittelwerten für latente Variablen werden in diesem Manuskript nicht behandelt. Der Vollständigkeit halber sollen jedoch die Optionen MM und AM erläutert werden:

MM

Momentenmatrix (moment matrix)

Dies ist die Matrix der mittleren Produkte und Kreuzprodukte, definiert durch:

$$1/N Z^T Z,$$

wobei  $Z$  die  $(N \times (p+q))$ -Matrix der Beobachtungswerte ist (zu den Bezeichnungen siehe Abschnitt 3.1).

Wenn ein LISREL-VI-Modell Ordinatenabschnitte oder Erwartungswertparameter für die latenten Variablen enthält, muß die Momentenmatrix der manifesten Variablen analysiert werden. Ferner muß eine der manifesten Variablen für alle Fälle den Wert 1 haben. Man kann LISREL auffordern, diese Hilfsvariable automatisch zu generieren und die Anzahl NI der Eingabevariablen entsprechend um 1 zu erhöhen, indem man AM als Wert für den Parameter MA spezifiziert:

|           |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |
|-----------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| AM        | Erweiterte Momentenmatrix (augmented moment matrix)<br>Sie ist definiert als die Momentenmatrix zu der rechtsseitig um die Spalte $(1,1,\dots,1)^T$ erweiterten Matrix $Z$ (siehe Beschreibung von MM).                                                                                                                                                                                                                                                                                                                                                                                                                             |
| XM        | Globaler Indikator für fehlende Werte<br>Der angegebene Wert wird bei allen Variablen als MD-Indikator aufgefaßt. Mit Hilfe des Steuerparameters XM kann also nur ein fallweiser Ausschluß fehlender Werte realisiert werden. Wer eine paarweise Behandlung wünscht, muß PRELIS als vorgeschaltetes Programm benutzen (vgl. PRELIS-Subkommando MISSING).<br>Beim Lesen der Variablen-Werte aus der aktiven SPSS-Systemdatei (vgl. Abschnitt 6.6.3.3) ist der Steuerparameter XM überflüssig: Hierbei werden der allgemeine MD-Indikator (SYSMIS) und benutzerspezifizierte MD-Indikatoren automatisch berücksichtigt <sup>1</sup> . |
| Beispiel: | <code>/da ni=4</code><br>Oft muß im DA-Subkommando nur die Anzahl der manifesten Variablen spezifiziert werden. LISREL analysiert dann eine Kovarianzmatrix und bestimmt die Fallzahl selbst.                                                                                                                                                                                                                                                                                                                                                                                                                                       |

### 6.6.3.2 Lesen aus einer SPSS-Matrix-Systemdatei (Subkommando MATRIX)

Die folgende Strategie zum Einlesen von Daten durch LISREL wurde in Abschnitt 6.2 ausdrücklich empfohlen:

- Einlesen der Daten, am besten aus einer passiven SPSS-Systemdatei unter Verwendung des GET-Kommandos. Dabei entsteht die aktive SPSS-Systemdatei des Programmlaufs. (Zu SPSS siehe z.B. Balthes-Götz 1993)
- Gegebenenfalls Modifikation der Daten mit SPSS-Transformationskommandos (z.B. COMPUTE, RECODE)
- Erstellung einer SPSS-Matrix-Systemdatei mit PRELIS

Diese enthält eine Matrix mit Assoziationsmaßen (z.B. Kovarianzen, Korrelationen, Momente, in Abhängigkeit vom PRELIS-Subkommando TYPE, vgl. Abschnitt 6.5.5), die Stichprobengröße und eventuell zusätzlich den Vektor der Mittelwerte.

Die vom Beispielprogramm in Abschnitt 4.3 produzierte Matrix-Systemdatei hat z.B. folgenden Inhalt:

| ROWTYPE_ | VARNAME_ | X1           | X2           | X3           | X4           |
|----------|----------|--------------|--------------|--------------|--------------|
| CORR     | X1       | 1.000000E+00 | 3.972100E-01 | 2.527460E-01 | 2.927070E-01 |
| CORR     | X2       | 3.972100E-01 | 1.000000E+00 | 2.599760E-01 | 3.008240E-01 |
| CORR     | X3       | 2.527460E-01 | 2.599760E-01 | 1.000000E+00 | 4.330990E-01 |
| CORR     | X4       | 2.927070E-01 | 3.008240E-01 | 4.330990E-01 | 1.000000E+00 |
| N        |          | 5.000000E+02 | 5.000000E+02 | 5.000000E+02 | 5.000000E+02 |

<sup>1</sup> Leider funktioniert in der Version 4.1 von SPSS unter BS2000 die automatische Übergabe der MD-Indikatoren *nicht*. Das Programm stürzt ab, wenn in der aktiven Systemdatei bei den von LISREL zu verarbeitenden Variablen MD-Indikatoren vorliegen. Dies gilt für SYSMIS wie für benutzerdefinierte MD-Indikatoren. Das Problem läßt sich beheben durch explizite Angabe des globalen MD-Indikators -0.989898D+37 im DA-Subkommando.

Beispiel: `/da ni=4 xm=-0.989898d+37`

In der Spalte ROWTYPE\_ ist für jede Zeile angegeben, von welchem Typ die darin enthaltenen Statistiken sind.

Falls eine der LISREL-Schätzmethoden WLS oder DWLS eingesetzt werden soll, muß zusätzlich die zugehörige Matrix mit asymptotischen (Ko)varianzen der Assoziations-Statistiken in eine Textdatei geschrieben werden.

- Einlesen und Verarbeiten der PRELIS-Ausgaben durch LISREL

Im LISREL-Aufruf kommt dabei das MATRIX-Subkommando zum Einsatz:

```
/MATRIX={IN(*) | IN('dateiname')}
```

|           |                                                        |
|-----------|--------------------------------------------------------|
| IN(...)   | Es soll eine SPSS-Matrix-Systemdatei gelesen werden.   |
| *         | Die aktive Matrix-Systemdatei des aktuellen SPSS-Laufs |
| dateiname | Die angegebene passive Matrix-Systemdatei              |

- Regeln:
- Die in der Syntax-Box angegebene Voreinstellung ist folgendermaßen zu verstehen: Wenn der Daten-Spezifikations-Block weder das MATRIX- noch ein anderes Subkommando zum Einlesen von Daten enthält, so werden die  $k$  ersten numerischen Variablen in der aktiven SPSS-(Matrix-)Systemdatei herangezogen. Nähere Erläuterungen hierzu finden sich am Anfang von Abschnitt 6.6.3.
  - Gemeinsam mit dem Subkommando MATRIX dürfen außer AC und AV keine weiteren Subkommandos zum Einlesen von Daten verwendet werden. (Verboten sind also die Subkommandos RA, CM, KM, OM, PM, MM, SD und ME.) Allerdings ist die ausschließliche Verwendung von MATRIX, AC und AV nicht einschränkend und wird von SPSS empfohlen.
  - LISREL übernimmt aus der aktiven SPSS-Matrix-Systemdatei neben den Daten auch die Stichprobengröße und die Variablennamen (als Labels für die manifesten Variablen). Man braucht also im DA-Subkommando zum Steuerparameter NO (s.o) keine Angaben zu machen und benötigt kein LA(bels)-Subkommando (s.u.). Die Übernahme der Variablennamen als Etiketten mißlingt merkwürdigerweise allerdings, wenn die aktive Matrix-Systemdatei mit dem SPSS-Kommando MATRIX DATA erstellt wurde und im Titel zum LISREL-Aufruf die Anführungszeichen fehlen (vgl. Abschnitt 6.6.2).
  - Bei der Verarbeitung einer *nicht* von PRELIS produzierten Matrix-Systemdatei ist zu beachten, daß vom LISREL-Subkommando MATRIX nur die folgenden Zeilen-Typen (siehe ROWTYPE\_ - Spalte in obigem Beispiel) akzeptiert werden: CORR, COV, MOMNT, MEAN, N.

Beispiel: `/matrix=in('presi.pm')`

### 6.6.3.3 Einlesen von Variablen aus der aktiven SPSS-Systemdatei

Für den direkten Zugriff auf die numerischen Variablen der aktiven SPSS-Systemdatei verwendet man einen Daten-Spezifikations-Block, der lediglich aus dem DA-Subkommando besteht. Dann werden die ersten NI (gemeint ist der Steuerparameter aus dem DA-Subkommando) numerischen Variablen von LISREL eingelesen. Per Voreinstellung wird die erste numerische Variable der aktiven Systemdatei als erste Y-Variable (im Sinne des LISREL-Modells, vgl. Abschnitt 3.1) verwendet, die zweite numerische Variable als zweite Y-Variable, u.s.w. bis zur  $k$ -ten numerischen Variablen, die als letzte X-Variable verwendet wird ( $k$  gibt die Anzahl der manifesten Variablen im LISREL-Modell an). Diese Zuordnung kann allerdings mit dem LISREL-Subkommando SE(lect) (siehe Abschnitt 6.6.3.9) abgeändert werden.

- Hinweise:
- LISREL übernimmt aus der aktiven Systemdatei neben den Variablenwerten auch die Stichprobengröße, die MD(missing data)-Deklarationen<sup>1</sup> und die Variablennamen (als Labels für die manifesten Variablen). Man braucht also im DA-Subkommando zu den Parametern NO und XM keine Angaben zu machen und benötigt kein LA(bels)-Subkommando (siehe Abschnitt 6.6.3.8).
  - Beim Einlesen von Variablen aus der aktiven SPSS-Systemdatei verzichtet man auf die Vorteile des LISREL-Preprozessors PRELIS, u.a.:
    - LISREL kann im Unterschied zu PRELIS keine polychorischen und polyserialen Korrelationen berechnen.
    - Man kann die LISREL-Schätzmethoden WLS und DWLS nicht verwenden, weil sie Vorleistungen von PRELIS benötigen.
    - Während LISREL 7 nur eine fallweise Behandlung fehlender Werte erlaubt, kann PRELIS auch einen paarweisen Ausschluß vornehmen.

Beispiel:

```
get file='last.sys'
lisrel
/da ni=4 xm=-0.989898d+37
/mo nx=4 nk=2 ph=st
/fr lx(1,1) lx(2,1) lx(3,2) lx(4,2)
/ou
```

### 6.6.3.4 Einlesen von Rohdaten aus einer Textdatei (Subkommando RA(w data))

Mit dem RA(w data)-Subkommando kann man für eine LISREL-Analyse Rohdaten (Variablen-Ausprägungen) einlesen, die sich entweder in einer externen Textdatei oder innerhalb des Programms (hinter dem RA-Subkommando) befinden. Das RA-Subkommando ist **nicht** geeignet zum Einlesen von Variablen aus der aktiven SPSS-Systemdatei.

Für das Lesen von Rohdaten ist oft eine Format-Angabe im FORTRAN-Stil erforderlich, die meist hinter dem RA-Subkommando in einer eigenen Programmzeile abgelegt wird. Obwohl es sich hierbei nicht um ein Subkommando handelt, **muß die Formatzeile mit einem Schrägstrich beginnen**. Dasselbe gilt für **Datenzeilen**, die im Programm hinter dem RA-Subkommando und der Formatzeile stehen.

---

<sup>1</sup> Leider funktioniert in der Version 4.1 von SPSS unter BS2000 die automatische Übergabe der MD-Indikatoren *nicht*. Das Programm stürzt ab, wenn in der aktiven Systemdatei bei den von LISREL zu verarbeitenden Variablen MD-Indikatoren vorliegen. Dies gilt für SYSMIS wie für benutzerdefinierte MD-Indikatoren. Das Problem läßt sich beheben durch explizite Angabe des globalen MD-Indikators -0.989898D+37 im DA-Subkommando.

Beispiel: /da ni=4 xm=-0.989898d+37

```

Beispiel: lisrel
 /da ni=4 no=219
 /ra
 /(4f3.2)
 /321 23178248
 /561121332665
 /419 82210410
 /
 /
 /219109198359
 /mo nx=4 nk=2 ph=st
 /fr lx(1,1) lx(2,1) lx(3,2) lx(4,2)
 /ou

```

In diesem Beispiel sollten die Syntaxregeln für Format- und Datenzeilen innerhalb von LISREL-Aufrufen demonstriert werden. Es wird jedoch *nicht* zur Nachahmung empfohlen, weil die hier verwendete Methode zur Dateneingabe als unpraktisch und veraltet bezeichnet werden muß. Wenn überhaupt Rohdaten mit LISREL eingelesen werden, dann sollten sich diese in einer externen Textdatei auf der Festplatte befinden. Meist ist es jedoch sinnvoller, die Daten aus einer SPSS-(Matrix-)Systemdatei zu übernehmen.

|                                  |
|----------------------------------|
| /RA [ FI='dateiname' [FO] [RE] ] |
| [(fortran-format)]               |
| [/ erste datenzeile]             |
| [/ zweite datenzeile]            |
| [/ . . .]                        |

- FI** Bestimmt, wo die Rohdaten abgelegt sind  
Voreinstellung (ohne FI-Spezifikation): Die Daten folgen, eventuell eingeleitet durch eine Formatzeile, unmittelbar hinter dem RA-Subkommando (siehe obiges Beispiel).
- dateiname** Benutzereigene Textdatei mit den Rohdaten
- FO** LISREL liest die Daten (NI Werte pro Fall) per Voreinstellung im **Freiformat**, d.h. zwischen zwei aufeinanderfolgenden Zahlen muß ein zulässiger Trenner (Leerzeichen oder Komma) stehen (Zu Abkürzungsmöglichkeiten siehe Jöreskog & Sörbom, 1989, S. 57).  
Das voreingestellte Freiformat kann durch ein **festes Einleseformat** ersetzt werden, indem den Daten eine Formatspezifikation im FORTRAN-Stil vorangestellt wird. Die FORTRAN-Formatangabe muß mit "(" beginnen, mit ")" enden und maximal 5 Zeilen à 80 Spalten lang sein.  
Der Steuer-Parameter FO ist nur relevant, wenn Daten aus einer externen Textdatei gelesen werden. Dann erwartet LISREL eine eventuelle Formatspezifikation zur Aufhebung des voreingestellten Freiformats in der ersten Zeile dieser externen Datei. Mit dem logischen Steuerparameter FO kann man LISREL veranlassen, statt dessen die Formatangabe aus der nächsten Programmzeile zu lesen.  
Die Anwendung von FORTRAN-Formaten wird später an Beispielen demonstriert.

RE

Dieser optionale Steuerparameter ist nur unter folgenden Bedingungen relevant:

- Es sollen in *einem* LISREL-Aufruf mehrere LISREL-Probleme unter Verwendung derselben Daten bearbeitet werden, d.h. nach dem OU-Subkommando, das die Beschreibung eines LISREL-Problems abschließt, folgen weitere Subkommandos zum selben LISREL-Kommando, in denen weitere LISREL-Probleme beschrieben werden.

Beispiel: `lisrel`  
`/"Problem 1, ML"`  
`/da ni=4`  
`/ra fi='presi.dat' re`  
`/mo nx=4 nk=2 ph=st`  
`/fr lx(1,1) lx(2,1) lx(3,2) lx(4,2)`  
`/ou`  
`/"Problem 2, GLS"`  
`/da ni=4`  
`/ra fi='presi.dat'`  
`/mo nx=4 nk=2 ph=st`  
`/fr lx(1,1) lx(2,1) lx(3,2) lx(4,2)`  
`/ou me=glS`

- Die bei allen LISREL-Problemen zu verwendenden Daten sollen aus einer externen Textdatei eingelesen werden.

In einem solchen Fall muß nach dem Lesen der Daten für ein LISREL-Problem die Eingabedatei mit Hilfe des RE-Steuerparameters zurückgespult werden, damit die Daten beim nächsten Teilauftrag erneut gelesen werden können.

Üblicherweise verwendet man jedoch für jedes LISREL-Problem einen eigenen LISREL-Aufruf und benötigt dabei den RE-Steuerparameter *nicht*.

Regeln:

- Auch beim Einlesen von Rohdaten mit dem RA-Subkommando muß eine aktive SPSS-Systemdatei definiert sein. Nötigenfalls kann diese ohne tatsächliches Einlesen oder Generieren von Daten mit einem einfachen SPSS-Eingabeprogramm erstellt werden (siehe Beispiel unten).
- Wenn die Rohdaten innerhalb des Programms vorliegen, muß im DA-Subkommando die tatsächliche Fallzahl angegeben werden. Beim Lesen aus einer externen Textdatei, kann LISREL die Fallzahl selbst bestimmen (vgl. Abschnitt 6.6.3.1).

Hinweis:

Wie beim direkten Zugriff auf die Variablen der aktiven SPSS-Systemdatei verzichtet man auch beim Einlesen von Rohdaten auf die Vorteile von PRELIS (vgl. die Hinweise in Abschnitt 6.6.3.3).

Beispiel:

```
input program
numeric dummy
end file
end input program

lisrel
/da ni=4
/ra fi='last.dat' fo
/(4f10.4)
/mo nx=4 nk=2 ph=st
/fr lx(1,1) lx(2,1) lx(3,2) lx(4,2)
/ou
```

Zunächst wird mit einem einfachen SPSS-Eingabeprogramm eine aktive Systemdatei definiert. Dies ist Voraussetzung für jede SPSS-Prozedur, also auch für LISREL. Weil im DA-Subkommando dem Steuerparameter NO kein Wert zugewiesen wird, ermittelt LISREL die tatsächliche Fallzahl beim Lesen. Im RA-Subkommando wird angegeben, daß die Daten in der Textdatei LAST.DAT gespeichert sind und daß eine Formatangabe zum Einlesen als nächste Programmzeile folgt.

### 6.6.3.5 Einlesen einer Assoziationsmatrix aus einer Textdatei (Subkommandos CM, KM, MM, OM und PM)

Das Einlesen einer Assoziationsmatrix vom Typ CM, KM, MM, OM oder PM (vgl. DA-Subkommando), die sich entweder in einer externen Textdatei oder innerhalb des Programms befindet, wird durch das gleichnamige Subkommando angefordert. Da sich die Subkommandos syntaktisch nur durch ihren Namen unterscheiden, können sie in einer gemeinsamen Syntax-Box dargestellt werden. Ferner haben sie folgende Steuer-Parameter mit dem RA-Subkommando gemeinsam, so daß alle diesbezüglichen Erläuterungen aus Abschnitt 6.6.3.4 übertragen werden können:

- FI     Wo ist Matrix gespeichert?  
Voreinstellung: Innerhalb des LISREL-Aufrufs
- FO     In welchem Format ist die Matrix geschrieben?  
Voreinstellung: Freiformat
- RE     Bei Speicherung in einer externen Datei: Zurückspulen nach dem Lesen?  
Voreinstellung: Nein

Weiterhin gilt analog zum RA-Subkommando, daß die SPSS-Prozedur LISREL auch beim Einlesen einer Assoziationsmatrix aus einer externen Textdatei bzw. aus dem Kommandostrom eine aktive SPSS-Systemdatei voraussetzt. Diese kann gegebenenfalls ohne tatsächliches Einlesen oder Generieren von Daten mit einem kurzen SPSS-Eingabeprogramm erstellt werden (siehe Beispiel in Abschnitt 6.6.3.4).

Für das Lesen aus einer aktiven oder passiven SPSS-Matrix-Systemdatei sind die in diesem Abschnitt beschriebenen Subkommandos **nicht** geeignet. Hierzu dient das weiter oben beschriebene MATRIX-Subkommando (siehe Abschnitt 6.6.3.2).

|                                                                                                   |
|---------------------------------------------------------------------------------------------------|
| <code>/{CM   KM   MM   OM   PM} [{SY   FU}] [ FI='dateiname' [FO] [RE] ]</code>                   |
| <code>[/ (fortran-format)]</code>                                                                 |
| <code>[/ erste datenzeile]</code><br><code>[/ zweite datenzeile]</code><br><code>[/ . . .]</code> |

|            |                                                                                                                                  |
|------------|----------------------------------------------------------------------------------------------------------------------------------|
| CM         | Das Subkommando CM dient zum Einlesen einer Kovarianzmatrix.                                                                     |
| KM         | Das Subkommando KM dient zum Einlesen einer Matrix mit Produkt-Moment-Korrelationen, basierend auf Rohwerten oder Normal-Scores. |
| MM         | Das Subkommando MM dient zum Einlesen einer Momentenmatrix.                                                                      |
| OM         | Das Subkommando OM dient zum Einlesen einer Korrelationsmatrix aus Optimal-Scores.                                               |
| PM         | Das Subkommando PM dient zum Einlesen einer Matrix mit polychorischen und/oder polyserialen Korrelationen.                       |
| FI, FO, RE | Siehe Subkommando RA.                                                                                                            |

SY, FU

Per Voreinstellung erwartet LISREL eine *untere Dreiecksmatrix inklusive Hauptdiagonale in Gestalt eines "langen Vektors"* (zeilenweise geordnet).  
Zu der Kovarianzmatrix aus Beispiel 1.2:

```
24.670 21.895 10.353 11.665
21.895 25.135 10.624 12.636
10.353 10.624 27.239 25.258
11.665 12.636 25.258 28.683
```

ergibt sich z.B. folgende untere Dreiecksmatrix in Gestalt eines "langen Vektors":

```
24.670 21.895 25.135 10.353 10.624 27.239 11.665 12.636 25.258 28.683
```

Der "lange Vektor" zu einer Matrix muß in einer Datei nicht unbedingt in einer einzigen, langen Zeile untergebracht sein. Er darf vielmehr nach praktischen Gesichtspunkten in mehrere Dateizeilen aufgeteilt werden. Wo die Elemente des Vektors von LISREL nacheinander in der Datei gesucht werden, legt das Einleseformat fest.

SY

Die Matrix liegt in unterer Dreiecksgestalt vor. Im Vergleich zur voreingestellten Vektorgestalt ist jedoch eine zusätzliche Eigenschaft gegeben: Jede Matrixzeile beginnt in einer neuen Dateizeile.

Als SY-Matrix sehen die Beispieldaten aus Beispiel 1.2 so aus:

```
24.670
21.895 25.135
10.353 10.624 27.239
11.665 12.636 25.258 28.683
```

Bei einer größeren Matrix darf sich eine Matrixzeile durchaus auch über mehrere Dateizeilen erstrecken.

FU

Die Matrix liegt in quadratischer Gestalt vor. Jede Matrixzeile beginnt in einer neuen Dateizeile. In dieser Gestalt wurde die Matrix aus Beispiel 1.2 schon oben wiedergegeben.

Natürlich gilt auch hier, daß eine sich eine Matrixzeile auch über mehrere Dateizeilen erstrecken darf.

Regel:

Obwohl es sich bei Format- oder Datenzeilen innerhalb des Programms nicht um Subkommandos handelt, müssen diese eingerückt und mit einem Schrägstrich eingeleitet werden.

Einige der vielen Möglichkeiten zum Einlesen einer Matrix sollen anhand der Daten aus Beispiel 1.2 demonstriert werden:

- i) Kovarianzmatrix innerhalb des LISREL-Aufrufs, unteres Dreieck wird als "langer Vektor" gelesen, FORTRAN-Format:

```
/cm
/(10f7.3)
/ 24.670 21.895 25.135 10.353 10.624 27.239 11.665 12.636 25.258 28.683
```

- ii) Kovarianzmatrix innerhalb des LISREL-Aufrufs, in unterer Dreiecksgestalt, FORTRAN-Format

```
/cm sy
/(4f7.3)
/ 24.670
/ 21.895 25.135
/ 10.353 10.624 27.239
/ 11.665 12.636 25.258 28.683
```

- iii) Kovarianzmatrix innerhalb des LISREL-Aufrufs, Matrix in quadratischer Gestalt, FORTRAN-Format

```
/cm fu
/(4f7.3)
/ 24.670 21.895 10.353 11.665
/ 21.895 25.135 10.624 12.636
/ 10.353 10.624 27.239 25.258
/ 11.665 12.636 25.258 28.683
```

- iv) Kovarianzmatrix innerhalb des LISREL-Kommandos, in unterer Dreiecksgestalt, freies Format

```
/cm
/24.670
/21.895 25.135
/10.353 10.624 27.239
/11.665 12.636 25.258 28.683
```

- v) Kovarianzmatrix in Textdatei LAST.DAT, in unterer Dreiecksgestalt, Format im LISREL-Kommando

```
/cm sy fi='last.dat' fo
/(4f7.3)
```

### 6.6.3.6 Einlesen von Mittelwerten und Standardabweichungen aus einer Textdatei (Subkommandos ME(ans)- und SD(avs))

Zur Analyse einer Kovarianzmatrix kann es sich anbieten, die zugehörige Korrelationsmatrix und den Vektor der Standardabweichungen einzulesen. Das ME(ans)-Subkommando zum Einlesen des Vektors mit den Mittelwerten der manifesten Variablen braucht man nur bei erweiterten LISREL-Modellen mit Mittelwertsstruktur, die in diesem Manuskript nicht behandelt werden.

Die gemeinsame Syntax der Subkommandos ME(ans) und SD(avs) unterscheidet sich von der Syntax der Subkommandos zum Einlesen einer Assoziationsmatrix nur durch das Fehlen der Steuerparameter SY und FU. Alle Regeln zu Dateien und Eingabeformaten gelten hier ebenfalls.

|                                                                                                   |
|---------------------------------------------------------------------------------------------------|
| <code>/{ME   SD} [ FI='dateiname' [FO] [RE] ]</code>                                              |
| <code>/(fortran-format)</code>                                                                    |
| <code>[/ erste datenzeile]</code><br><code>[/ zweite datenzeile]</code><br><code>[/ . . .]</code> |

**ME** Das Subkommando ME dient zum Einlesen eines Vektors mit den Mittelwerten der manifesten Variablen.

|            |                                                                                                                                                                           |
|------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| SD         | Das Subkommando SD dient zum Einlesen eines Vektors mit den Standardabweichungen der manifesten Variablen.                                                                |
| FI, FO, RE | Siehe Subkommando RA.                                                                                                                                                     |
| Regel:     | Obwohl es sich bei Format- oder Datenzeilen innerhalb des Programms nicht um Subkommandos handelt, müssen diese eingerückt und mit einem Schrägstrich eingeleitet werden. |
| Beispiel:  | <pre> /sd /4.967 5.013 5.219 5.356 </pre> <p>Hier werden die Standardabweichungen zu den Variablen aus Beispiel 1.2 im Freiformat eingelesen.</p>                         |

### 6.6.3.7 Einlesen von asymptotischen (Ko-)Varianzen der Assoziations-Statistiken (Subkommandos AC und AV)

Für die Anwendung der WLS-Schätzmethode muß mit dem Subkommando AC die in Abschnitt 4.2 beschriebene Matrix mit den geschätzten asymptotischen Kovarianzen der zu analysierenden Assoziationsmaße eingelesen werden. Analog muß beim Einsatz der DWLS-Methode mit dem Subkommando AV der Vektor mit den geschätzten asymptotischen Varianzen der Assoziationsmaße gelesen werden. Asymptotische (Ko)varianzen können nur aus externen Textdateien gelesen werden. Normalerweise gelangen sie durch einen vorgeschalteten PRELIS-Aufruf dorthin (vgl. Abschnitt 6.5.7).

```
/{AC | AV} FI='dateiname'
```

|         |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
|---------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| AC      | Das Subkommando AC dient zum Einlesen einer geschätzten asymptotischen Kovarianzmatrix der zu analysierenden Assoziationsmaße.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
| AV      | Das Subkommando AV dient zum Einlesen eines Vektors mit geschätzten asymptotischen Varianzen der zu analysierenden Assoziationsmaße.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
| FI      | Benutzereigene Textdatei mit den asymptotischen (Ko)varianzen                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
| Regeln: | <ul style="list-style-type: none"> <li>- Beim Einlesen eines PRELIS-Produktes (siehe Beispiel unten) muß im AC- bzw. AV-Kommando lediglich der richtige Dateiname angegeben werden. Ist die einzulesende Datei kein Original-PRELIS-Produkt und kann sie nicht freiformatig gelesen werden, so muß sie in der ersten Zeile das Leseformat im FORTRAN-Stil enthalten.</li> <li>- Das Subkommando AC erwartet die Eingabematrix in unterer Dreiecksgestalt (inklusive Hauptdiagonale).</li> <li>- AC und AV setzen voraus, daß die zu analysierende Matrix vom Typ CM, KM oder PM ist (vgl. Steuerparameter MA im Subkommando DA).</li> <li>- Bei Auftreten des Subkommandos AC bzw. AV ändert sich automatisch die vor-eingestellte Schätzmethode von ML auf WLS bzw. DWLS.</li> </ul> |

```

Beispiel: prelis
 /variables=x1 to x4 (ordinal)
 /type=polychor
 /matrix=out(*)
 /write=acov 'presi.acp'

 lisrel
 /"Poly.-Kor.-Matr. aus ordinalen Indikatoren, WLS"
 /da ni=4 ma=pm
 /matrix=in(*)
 /ac file='presi.acp'
 /mo nx=4 nk=2 ph=st
 /fr lx(1,1) lx(2,1) lx(3,2) lx(4,2)
 /ou

```

### 6.6.3.8 Vereinbaren von Etiketten für die manifesten Variablen (Subkommando LA(bels))

Mit dem LA(bels)-Subkommando können Etiketten für die manifesten Variablen vereinbart werden. Es ist überflüssig (und verboten) beim Einlesen von Variablen aus der aktiven SPSS-Systemdatei oder aus einer SPSS-Matrix-Systemdatei; in diesen Fällen werden die Variablennamen aus dem Deklarationsteil der Systemdatei automatisch als Etiketten übernommen.

|                                               |
|-----------------------------------------------|
| <code>/LA [ FI='dateiname' [FO] [RE] ]</code> |
| <code>[(/fortran-a-format)]</code>            |
| <code>[/ erste zeile mit etiketten]</code>    |
| <code>[/ zweite zeile mit etiketten]</code>   |
| <code>[/ . . .]</code>                        |

FI, FO, RE            Siehe Subkommando RA.

- Regeln:
- Obwohl es sich bei den Format- oder Labelzeilen innerhalb des Programms nicht um Subkommandos handelt, müssen diese eingerückt und mit einem Schrägstrich eingeleitet werden.
  - Die Etiketten müssen in derselben Reihenfolge angegeben werden, in der die Variablen auf dem Speichermedium vorliegen.
  - Die voreingestellten Etiketten lauten: "VAR 1", "VAR 2", ...
  - LISREL kürzt alle Etiketten auf acht Zeichen: Bei freiformatigem Lesen wird rechts, bei festformatigem Lesen wird links abgeschnitten.
  - Leerzeichen sind in Etiketten nicht erlaubt.
  - Bei freiformatigem Lesen bestehen folgende Abkürzungsmöglichkeiten:
    - Sollen für die letzten  $r$  Variablen die voreingestellten Label beibehalten werden, so kann die Liste der Etiketten entsprechend früher durch einen Schrägstrich beendet werden.
    - Man kann eine Variable bei der Vergabe von Etiketten überspringen, indem man an ihrer Position zwei Kommas setzt. Analog kann man mit drei Kommas zwei Variablen überspringen usw.
  - Beim festformatigem Lesen muß ein A-Format verwendet werden (vgl. zweites Beispiel).

Beispiele:

```
- /la
 /Y1 Y2 Y3 Y4 X1 X2 X3 X4 X5 X6 X7
- /la
 /(11a2)
 /Y1Y2Y3Y4X1X2X3X4X5X6X7
- lisrel
 /da ni=8 no=428
 /cm fi='kovmat'
 /la
 /x1 y1, ,x2 y2 /
 /se
 /Y1 Y2 X1 X2 /
 /mo ...
 /ou
```

Im dritten Beispiel werden für eine Analyse vier Variablen ausgewählt, die in einer Datei mit acht Variablen die Positionen 1, 2, 4 und 5 einnehmen (vgl. SE-Subkommando). Mit dem LA-Subkommando werden nur für die benötigten Variablen Etiketten vergeben.

### 6.6.3.9 Auswählen und Umordnen der Variablen für eine Analyse (Subkommando SE(lect))

Durch den Steuerparameter NI des Subkommandos DA erfährt LISREL, wieviele Eingabevariablen auf dem Speichermedium bereitstehen. Über die Steuerparameter NY und NX des Subkommandos MO (s.u.) teilt der Benutzer mit, wieviele Y- bzw. X-Variablen sein Modell enthält. Per Voreinstellung faßt LISREL die ersten NY einzulesenden Variablen als Y-Variablen auf und die nächsten NX einzulesenden Variablen als X-Variablen. Mit dem SE(lect)-Subkommando kann die Rollenverteilung gesteuert werden durch Angabe einer Liste mit (NY + NX) Variablen-Bezeichnungen, womit die NY zuerst genannten als Y-Variablen und die NX zuletzt genannten als X-Variablen deklariert werden.

Die Auswahlliste kann entweder hinter dem SE-Subkommando im LISREL-Aufruf abgelegt sein oder in einer externen Textdatei. Sie wird auf jeden Fall in freiem Format gelesen, weshalb eine Formatspezifikation überflüssig und verboten ist. Die Variablen können entweder über ihre Position in den Daten oder über ihre mit dem LA-Subkommando vereinbarten Etiketten angesprochen werden.

|                      |
|----------------------|
| /SE [FI='dateiname'] |
|----------------------|

|                                      |
|--------------------------------------|
| [/liste mit variablen-bezeichnungen] |
|--------------------------------------|

Regeln:

- Obwohl es sich bei der Liste mit Variablen-Bezeichnungen nicht um ein Subkommando handelt, muß sie eingerückt und mit einem Schrägstrich eingeleitet werden.
- Wenn weniger als NI Variablen selektiert werden, muß die Liste durch "/" abgeschlossen werden (siehe Beispiele).
- Eine abkürzende Spezifikation von Variablenlisten mit dem Schlüsselwort "TO" ist im SE-Subkommando nicht erlaubt.

Beispiel:

Soll bei der Analyse eines LISREL-Modells auf eine Korrelationsmatrix zurückgegriffen werden, in der die Variablen nicht Modell-konform angeordnet sind, kann mit dem SE-Subkommando für die richtige Reihenfolge gesorgt werden. Im Beispiel enthält die Korrelationsmatrix fünf EingabevARIABLEN, von denen im LISREL-Modell die vierte als erste Y-Variable, die fünfte als zweite Y-Variable, die erste als erste X-Variable und die dritte als zweite X-Variable verwendet werden soll. Zunächst wird die Variablenauswahl über Positions-Nummern verwendet, anschließend wird auch die Auswahl über Etiketten demonstriert:

```
/Berufliche Ambitionen von Jugendlichen
/da ni=5 no=767 ma=km
/la
/intellnz bldngvat berufvat schullst ambition
/km
/1
/.277 1
/.250 .611 1
/.572 .294 .248 1
/.335 .303 .331 .478 1
/se
/4 5 1 3 /
/mo ny=2 nx=2 ...
```

Selektion über Etiketten:

```
/se
/schullst ambition intellnz berufvat /
```

#### 6.6.4 Modell-Spezifikation

Im Modell-Spezifikations-Block können folgende Subkommandos zum Beschreiben des Modells verwendet werden:

|    |                                                                                                                                                                                            |
|----|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| MO | Spezifikation zentraler Eigenschaften des Modells<br>Das Subkommando leitet den Modell-Spezifikations-Block ein und ist obligatorisch für die Durchführung einer Kovarianzstrukturanalyse. |
| FR | Festlegung des Typs "frei" für eine Liste von Parametern                                                                                                                                   |
| FI | Festlegung des Typs "fixiert" für eine Liste von Parametern                                                                                                                                |
| NF | Betrifft die automatische Modellsuche: die angegebenen Parameter dürfen nicht frei gesetzt werden.                                                                                         |
| PA | Einlesen einer Matrix mit Parametertypen                                                                                                                                                   |
| EQ | Deklaration von Gleichheits-Restriktionen                                                                                                                                                  |
| VA | Besetzung fixierter Parameter mit Werten ungleich Null                                                                                                                                     |
| ST | Festlegung von Startwerten zu nicht-fixierten Parametern (für iterative Schätzverfahren)                                                                                                   |
| MA | Einlesen einer Matrix mit Parameterwerten (feste Werte für fixierte Parameter und Startwerte für nicht-fixierte)                                                                           |
| PL | Plotten der Fit-Funktion gegen einen Parameter                                                                                                                                             |
| LE | Einlesen von Etiketten für die $\eta$ -Variablen                                                                                                                                           |
| LK | Einlesen von Etiketten für die $\xi$ -Variablen                                                                                                                                            |

Innerhalb des Modell-Spezifikations-Blocks können nach dem einleitenden und obligatorischen MO-Subkommando die zur Modell-Beschreibung erforderlichen sonstigen Subkommandos beliebig angeordnet werden.

### 6.6.4.1 Das MO(dell)-Subkommando

Im MO(dell)-Subkommando werden zentrale Eigenschaften des zu analysierenden Modells deklariert (Dimensionen der Zufallsvektoren  $Y$ ,  $X$ ,  $\eta$  und  $\xi$ ; Formen und Modi der Parametermatrizen). Es ist obligatorisch für die Durchführung einer Kovarianzstrukturanalyse.

```
/MO [NY={0|p}] [NX={0|q}] [NE={0|m}] [NK={0|n}] [{mn=mf,mm | mn=mf | mn=mm}] [{mn= ...}] [FI]
```

- NY,NX,NE,NK**      Dimensionen der im Modell enthaltenen Zufallsvektoren  
 Durch Weglassung bestimmter Dimensionsangaben, d.h. durch Beibehaltung der voreingestellten Nullen, wird ein Submodell gemäß Abschnitt 3.5 gewählt.
- mn**                Name einer LISREL-Parametermatrix (vgl. Tabelle 3.6). Die erlaubten Werte sind also: LY, LX, BE, GA, PH, PS, TE, TD.  
 In Abhängigkeit vom Modell ist es erforderlich bzw. sinnvoll, für eine Teilmenge der acht Parametermatrizen die Matrix-Form und/oder den Matrix-Modus zu spezifizieren.
- mf**                Matrixform  
 Tabelle 3.5 enthält eine Liste der möglichen Formen von Parametermatrizen. Welche Formen für eine spezielle Parametermatrix erlaubt sind, kann man der Tabelle 3.6 entnehmen.  
 Von einer Matrix der Form DI wird nur die Hauptdiagonale im Rechner gespeichert. Matrizen der Formen ID oder ZE werden überhaupt nicht im Speicher gehalten. Ein nicht gespeicherter Parameter kann bei der weiteren Modellspezifikation nicht angesprochen werden und wird auch bei der Berechnung der Modifikationsindikatoren nicht berücksichtigt. Sind für einen per Voreinstellung ausgeblendet (und damit auf Null bzw. Eins fixierten) Parameter Modifikationsindikatoren erwünscht, so muß der betroffenen Matrix eine andere Form zugewiesen werden.  
 Beispiel:      Die Matrix B ist per Voreinstellung von der Form ZE. Um Modifikationsindikatoren für die auf Null fixierten Parameter  $\beta_{ij}$  zu erhalten, kann man der Matrix B die Form FU zuweisen:  
                   /mo ny=2 nx=3 be=fu  
 Für die im LISREL-Modell grundsätzlich nicht existenten Parameter  $\beta_{ii}$  (Effekte endogener Variablen auf sich selbst) werden dabei natürlich keine Modifikationsindikatoren geliefert.
- mm**                Matrixmodus  
 Gemäß Tabelle 3.4 sind hier die Werte FI (fixiert) und FR (frei) erlaubt.  
 Regeln:        - Wird bei  $\Theta_\epsilon$  oder  $\Theta_\delta$  die voreingestellte Form DI auf SY abgeändert und keine Angabe zum Modus gemacht, so sind anschließend die Diagonalelemente frei (gemäß dem voreingestellten Matrix-Modus), die Nicht-Diagonal-Elemente jedoch auf Null fixiert. Mit "TE=SY,FR" bzw."TD=SY,FR" kann jeweils die gesamte Matrix frei gesetzt werden.

- Nach der Spezifikation PH=ST sind die Diagonalelemente von  $\Phi$  auf Eins fixiert und die übrigen Elemente sind frei. Dieser Status kann durch die Subkommandos FR, FI oder PA (s.u.) nicht geändert werden. Die Spezifikationen "PH=ST,FI" und "PH=ST,FR" sind verboten.
- Nach der Spezifikation BE=SD sind die Elemente unterhalb der Hauptdiagonalen von B frei und alle übrigen Elemente sind auf Null fixiert. Durch die Zuweisung eines Matrix-Modus' kann dieser Status nicht geändert werden, jedoch können mit den Subkommandos FI, EQ und PA (s.u.) Subdiagonalelemente fixiert oder restringiert werden.

|            |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
|------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| FI         | FIXED-X<br>Mit diesem in Abschnitt 3.9 erläuterten logischen Steuerparameter werden die $\xi$ -Variablen mit den X-Variablen identifiziert, d.h. $NK=NX$ , $\Lambda_X=I$ , $\Theta_\delta=O$ , $\Phi=S_X$ (fixiert, $S_X$ = Kovarianzmatrix von X). Er ist automatisch aktiv, wenn für NE und NK die Voreinstellung 0 beibehalten wird, d.h. im Submodell 2 (Kausalmodell für direkt beobachtete Variablen).                                                                                                                                                                                                                                                                                                                                                     |
| Regel:     | Das MO-Subkommando ist obligatorisch für die Durchführung einer Kovarianzstrukturanalyse; wenn es fehlt, gibt LISREL lediglich die zu analysierende Matrix aus.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
| Beispiele: | <ul style="list-style-type: none"> <li>- /mo nx=4 nk=2 td=sy<br/>Zur Analyse von Beispiel 1.2 wird das Submodell 1 (konfirmatorische Faktorenanalyse für X-Variablen) spezifiziert. Für alle beteiligten Parametermatrizen kann die Voreinstellung übernommen werden. Für <math>\Theta_\delta</math> wird aber trotzdem die Form SY festgelegt, damit für die Residual-Kovarianzen Modifikationsindikatoren berechnet werden.</li> <li>- /mo ny=4 nx=7 ne=2 nk=3 be=fu<br/>Zur Analyse von Beispiel 1.3 muß für die Matrix B die voreingestellte Form ZE auf FU geändert werden, weil das Modell gerichtete Effekte zwischen den <math>\eta</math>-Variablen enthält. Bei allen anderen Parametermatrizen kann die Voreinstellung beibehalten werden.</li> </ul> |

#### 6.6.4.2 Freisetzen bzw. Fixieren von Paramtern (Subkommandos FR(ee)- bzw. FI(x))

Die meisten Modelle verlangen für einzelne Elemente aus den Parametermatrizen eine vom (voreingestellten oder gewählten) Matrix-Modus abweichende Fixierung bzw. Freisetzung. Dazu dienen die Subkommandos FI(x) bzw. FR(ee). Sie bestehen einfach aus dem Subkommandonamen und einer Liste von Parametern, wobei abkürzend auch Parameterbereiche angegeben werden können.

Weil solche Listen auch in anderen Subkommandos auftreten, soll hier zur Vereinfachung der Schreibweise der Platzhalter "**parameterliste**" eingeführt werden:

{mn(zeile,spalte) | mn(zeile,spalte) - mn(zeile,spalte)} [mn ...]

mn Name einer LISREL-Parametermatrix (vgl. Tabelle 3.6). Die erlaubten Werte sind also: LY, LX, BE, GA, PH, PS, TE, TD.

Regel: Bei der Angabe von *Parameterbereichen* muß man sich an der natürlichen Anordnung der nicht-redundanten, gemäß Matrix-Form im Rechner gespeicherten Elemente einer Parametermatrix orientieren (vgl. Abschnitt 3.7).  
 In Beispiel 1.3 mit drei  $\xi$ -Variablen ist die Matrix  $\Phi$  von der voreingestellten Form SY:

$$\Phi = \begin{bmatrix} \phi_{11} & & \\ \phi_{21} & \phi_{22} & \\ \phi_{31} & \phi_{32} & \phi_{33} \end{bmatrix}$$

Damit ergibt sich für die nicht-redundanten Elemente von  $\Phi$  die folgende natürliche Anordnung:

$$\phi_{11}, \phi_{21}, \phi_{22}, \phi_{31}, \phi_{32}, \phi_{33}$$

Beispiele: - ly(2,1) ly(3,2) be(4,1)  
 - ga(2,1) td(1,1) -td(4,4)  
 Wenn  $\Theta_\delta$  die voreingestellte Form DI hat, ist umfaßt diese Liste neben  $\gamma_{21}$ :  
 $\theta_{11}^{(\delta)}, \theta_{22}^{(\delta)}, \theta_{33}^{(\delta)}, \theta_{44}^{(\delta)}$ .

Syntax-Box für die Subkommandos FR und FI:

```
/{FR | FI} parameterliste
```

FR Mit diesem Subkommado wird für alle aufgelisteten Parameter der Modus "frei" vereinbart, d.h. sie dürfen frei geschätzt werden.

FI Mit diesem Subkommado wird für alle aufgelisteten Parameter der Modus "fixiert" vereinbart, d.h. sie werden beim Schätzen auf dem voreingestellten Wert oder auf einem Benutzer-definierten Wert (s.u.) konstant gehalten. Der voreingestellte Fixierungswert ist bei Parametern, deren Freiheit vom Benutzer aufgehoben werden kann, stets Null (vgl. VA(lue)-Subkommando).

Regeln: - Es können nur solche Parameter angesprochen werden, die aufgrund der Matrix-Form im Rechner gespeichert sind (vgl. Abschnitt 3.7).  
 - Von mehreren Angaben zu einem Parameter zählt die Letzte.  
 - Nach der Spezifikation PH=ST im MO-Subkommando sind die Hauptdiagonalelemente von  $\Phi$  auf 1 fixiert und die übrigen Elemente sind frei. Dieser Status kann mit FI- oder FR-Subkommandos nicht geändert werden.

Beispiele: - /mo ny=4 nx=7 ne=2 nk=3 be=FU  
 /fr ly(2,1) ly(4,2)  
           lx(2,1) lx(3,1) lx(3,2) lx(5,2) lx(7,3)  
           be(2,1) be(1,2)  
 Da  $\Lambda_y, \Lambda_x$  und B den voreingestellte Modus FI haben, müssen die freien Parameter in einem FR-Subkommando spezifiziert werden.  
 - /fi ga(1,3) ga(2,2)  
 $\Gamma$  hat den voreingestellten Modus FR. Daher müssen zu fixierende  $\gamma$ -Parameter explizit in einem FI-Subkommando aufgeführt werden.

### 6.6.4.3 Freisetzung von Parametern bei der automatischen Modellsuche verhindern (Subkommando NF(ree))

LISREL kann eine automatische Modellsuche betreiben, indem es Schritt für Schritt die Fixierung oder Gleichheitsrestriktion zum größten Modifikationsindex aufhebt, solange dieser statistisch signifikant ist (siehe Abschnitt 6.6.5). Mit dem NF-Subkommando (never free) kann für eine Liste von Parametern die Freisetzung verhindert werden. Für diese Parameter werden auch keine Modifikationsindikatoren berechnet.

|                    |
|--------------------|
| /NF parameterliste |
|--------------------|

parameterliste      Siehe Beschreibung der Subkommandos FI und FR (Abschnitt 6.6.4.2)

Hinweis:              In der Parameter-Spezifikation am Anfang der LISREL-Ausgabe (siehe Abschnitt 3.6) werden die angegebenen Parameter durch -1 gekennzeichnet, bei der Ausgabe der Modifikationsindikatoren erscheint eine Null.

Beispiel:             /nf   1x(1,1) 1x(2,1) 1x(3,2) 1x(4,2)

### 6.6.4.4 Deklaration freier und fixierter Parameter durch eine Mustermatrix (Subkommando PA(ttern))

Zur Spezifikation der freien und fixierten Elemente einer Parametermatrix kann alternativ zur Verwendung von FR- und FI-Subkommandos mit Hilfe des PA(ttern)-Subkommandos eine Muster-Matrix von gleicher Form wie die Parametermatrix eingelesen werden, deren Element  $(i,j)$  gleich 0 bzw. 1 ist, wenn der korrespondierende Parameter fixiert bzw. frei sein soll. Für das Einlesen einer Mustermatrix gelten weitgehend dieselben Regeln wie für das Einlesen einer Assoziationsmatrix aus einer Textdatei (siehe Abschnitt 6.6.3.5), sie kann also im LISREL-Aufruf oder in einer externen Datei stehen und optional durch ein FORTRAN-Format eingeleitet werden.

|                                     |
|-------------------------------------|
| /PA mn [ FI='dateiname' [FO] [RE] ] |
|-------------------------------------|

|                      |
|----------------------|
| [(fortran-i-format)] |
|----------------------|

|                                          |
|------------------------------------------|
| [/ erste zeile mit typ-spezifikationen]  |
| [/ zweite zeile mit typ-spezifikationen] |
| [/ . . .]                                |

mn                     Name einer LISREL-Parametermatrix (vgl. Tabelle 3.6). Die erlaubten Werte sind also: LY, LX, BE, GA, PH, PS, TE, TD.

FI, FO, RE            Siehe RA-Subkommando  
Zusätzlich ist zu beachten: Es sind nur Integer-Formate erlaubt.

Regeln:              - Die Mustermatrix muß von derselben Form wie die zugehörige Parametermatrix sein. Hat also z.B.  $\Theta_8$  die Form DI, darf die Mustermatrix nur aus einem Vektor zur Spezifikation der Hauptdiagonalelemente bestehen.  
- Bei freiformatigem Einlesen besteht folgende Abkürzungsmöglichkeit: Sollen die letzten  $r$  Parameter einer Matrix als fixiert spezifiziert werden, kann die entsprechende Folge von Nullen durch einen Schrägstrich ersetzt werden.

Beispiele:

```
- /pa lx
 /(3i1)
 /010
 /100
 /110
 /000
 /010
 /000
 /001
```

Diese Matrix spezifiziert das Muster freier und fixierter Parameter in  $\Lambda_X$  aus Beispiel 1.3.

```
- /mo nx=4 nk=1 td=sy
 /pa td
 /1
 /1 1
 /0 0 1
 /0 0 1 1
```

Hier hat  $\Theta_\delta$  die Form SY(mmetrisch), so daß die Mustermatrix in unterer Dreiecksgestalt (inklusive Hauptdiagonale) präsentiert werden muß.

#### 6.6.4.5 Deklaration von Gleichheitsrestriktionen (Subkommando EQ(ual))

Für jede Gruppe gleichzusetzender Parameter wird ein EQ(ual)-Subkommando benötigt.

|                    |
|--------------------|
| /EQ parameterliste |
|--------------------|

parameterliste      Siehe Beschreibung der Subkommandos FI und FR (Abschnitt 6.6.4.2)

Beispiel:

```
/mo nx=4 nk=2
/eq td(1,1)-td(4,4)
```

Dieses EQ-Subkommando setzt für das testtheoretische Beispiel 1.2 die Fehlervarianzen der manifesten Variablen gleich. Die explizite Auflistung der Hauptdiagonalelemente von  $\Theta_\delta$  kann hier durch eine Bereichsangabe ersetzt werden, weil  $\Theta_\delta$  die (voreingestellte) Matrixform DI(agonal) hat.

#### 6.6.4.6 Spezifikation von Werten für eine Liste von Parametern (Subkommandos VA(lue)- und ST(art))

Ein aufgrund der gewählten Matrixform (vgl. die Abschnitt 3.7 und 6.6.4.1) ansprechbarer und (per Voreinstellung oder Benutzerwahl) fixierter Parameter ist mit Null vorbesetzt. Viele LISREL-Modelle enthalten jedoch Parameter, die auf andere Werte fixiert sein sollen. Solche alternativen Werte können mit dem VA(lue)-Subkommando vereinbart werden.

Die bei iterativen Schätzverfahren benötigten Startwerte für alle schätzbaren Parameter werden normalerweise von LISREL automatisch generiert (vgl. Abschnitt 2.4.4). Allerdings kann der Benutzer mit dem ST(art)-Subkommando auch alternative Startwerte verwenden. Dabei ist zu bedenken, daß die Erfolgswahrscheinlichkeit und der Zeitaufwand einer iterativen Schätzprozedur von der Güte der Startwerte abhängen.

Bei manchen Modellen kann LISREL zu schätzbaren Parametern keine automatischen Startwerte generieren, so daß diese vom Benutzer mit dem ST(art)-Subkommando angegeben werden müssen.

Die Subkommandos VA und ST können gemeinsam in einer Syntax-Box dargestellt werden:

```
/{VA | ST} wert {parameterliste | ALL}
```

|                |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
|----------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| VA             | Mit dem VA-Subkommando kann man Werte für fixierte Parameter festlegen.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
| ST             | Mit dem ST-Subkommando kann man Startwerte für die iterative Schätzung nicht-fixierter Parameter festlegen.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
| parameterliste | Siehe Beschreibung der Subkommandos FI und FR (Abschnitt 6.6.4.2)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
| ALL            | <p>Das Schlüsselwort bedeutet im ...</p> <ul style="list-style-type: none"> <li>- VA-Subkommando: Alle Modellparameter (frei, fixiert oder restringiert) werden auf den angegebenen Wert gesetzt. Dieser wird bei nicht-fixierten Parametern als Startwert für das Schätzverfahren verwendet.</li> <li>- ST-Subkommando: Alle nicht-fixierten Parameter werden mit dem angegebenen Startwert versorgt.</li> </ul> <p>Zusammen mit der Möglichkeit, mehrere VA- bzw. ST-Subkommandos zu verwenden, kann das Schlüsselwort gelegentlich die Wertzuweisung vereinfachen. "ALL" darf nicht abgekürzt werden.</p>                                                                                                                                                                                                                                                   |
| Regel:         | <p>LISREL kennt zwar für das Zuweisen fester Werte und für das Vereinbaren von Startwerten jeweils ein eigenes Subkommando, doch sind diese beiden Subkommandos (VA und ST) weitgehend wirkungsgleich. Mit beiden Subkommandos werden den aufgelisteten Parametern Werte zugewiesen, die in Abhängigkeit von den Parameter-Modi als fixe Werte oder als Startwerte aufgefaßt werden. VA und ST haben jedoch unterschiedliche Effekte bei Verwendung des Schlüsselwortes ALL (s.o.) und (in analoger Weise) bei Angabe von Parameterbereichen (z.B. td(1,1)-td(4,4)):</p> <ul style="list-style-type: none"> <li>- Das VA-Subkommando weist in diesen Fällen <i>allen</i> angesprochenen Parametern den gewünschten Wert zu.</li> <li>- Das ST-Subkommando weist in diesen Fällen <i>nur den nicht-fixierten</i> Parametern den gewünschten Wert zu.</li> </ul> |
| Beispiel:      | <pre>/va 1 ly(1,1) ly(3,2) lx(1,1) lx(4,2) lx(6,3)</pre> <p>Mit diesem Subkommando wird in Beispiel 1.3 für jede latente Variable zur Festlegung der Maßeinheit eine Ladung auf 1 fixiert.</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |

#### 6.6.4.7 Spezifikation von (Start-)Werten für eine Parametermatrix (Subkommando MA)

Das MA-Subkommando dient zum Einlesen einer Matrix mit (Start-)Werten für eine LISREL-Parametermatrix. Es steht damit zum VA-Subkommando bzw. zum ST-Kommando in ähnlicher Relation wie das PA-Subkommando zum FI- und zum FR-Subkommando. Fixierte Parameter werden auf den eingelesenen Wert gesetzt, freie bzw. restringierte Parameter erhalten entsprechende Startwerte für iterative Schätzverfahren.

Für das Einlesen einer Matrix mit (Start-)Werten gelten weitgehend dieselben Regeln wie für das Einlesen einer Assoziationsmatrix aus einer Textdatei (siehe Abschnitt 6.6.3.5), sie kann also im LISREL-Aufruf oder in einer externen Datei stehen und optional durch ein FORTRAN-Format eingeleitet werden.

|                                                                                                               |
|---------------------------------------------------------------------------------------------------------------|
| <code>/MA mn [ FI='dateiname' [FO] [RE] ]</code>                                                              |
| <code>[/ (fortran-format)]</code>                                                                             |
| <code>[/ erste zeile mit werten]</code><br><code>[/ zweite zeile mit werten]</code><br><code>[/ . . .]</code> |

**mn** Name einer LISREL-Parametermatrix (vgl. Tabelle 3.6). Die erlaubten Werte sind also: LY, LX, BE, GA, PH, PS, TE, TD.

**FI, FO, RE** Siehe Subkommando RA.

**Regeln:**

- Die einzulesende Matrix mit (Start-)Werten muß von derselben Form wie die zugehörige Parameter-Matrix sein. Hat also z.B.  $\Theta_{\delta}$  die Form DI, darf nur ein Vektor mit (Start-)Werten für die Hauptdiagonalelemente eingelesen werden.
- Bei freiformatigem Einlesen besteht folgende Abkürzungsmöglichkeit: Sollen die letzten  $r$  Parameter einer Matrix ihren derzeitigen Wert behalten, kann die entsprechende Folge von Zahlen durch einen Schrägstrich ersetzt werden.

**Beispiel:**

```

/ma 1x
1 0 0
0 0 0
0 0 0
0 1 0
0 0 0
0 0 1
0 0 0

```

Diese Sequenz kann an Stelle des VA-Subkommandos aus dem letzten Abschnitt verwendet werden.

#### 6.6.4.8 Vereinbarung von Etiketten für latente Variablen (Subkommandos LE(ta) und LK(si))

Mit dem LE(ta)- bzw. dem LK(si)-Subkommando können Etiketten für die latenten Variablen vereinbart werden. Es gelten dieselben Regeln wie für das LA(bels)-Subkommando (siehe Abschnitt 6.6.3.8).

|                                                                                                                     |
|---------------------------------------------------------------------------------------------------------------------|
| <code>/{LE   LK} [ FI='dateiname' [FO] [RE] ]</code>                                                                |
| <code>[/ (fortran-a-format)]</code>                                                                                 |
| <code>[/ erste zeile mit etiketten]</code><br><code>[/ zweite zeile mit etiketten]</code><br><code>[/ . . .]</code> |

**LE** Das Subkommando LE dient zum Einlesen von Etiketten für die  $\eta$ -Variablen.

**LK** Das Subkommando LK dient zum Einlesen von Etiketten für die  $\xi$ -Variablen.

|            |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
|------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| FI, FO, RE | Siehe Subkommando RA.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
| Regeln:    | <ul style="list-style-type: none"> <li>- Die Etiketten müssen in derselben Reihenfolge angegeben werden, in der die <math>\eta</math>- bzw. <math>\xi</math>-Variablen in den Parametermatrizen stehen.</li> <li>- Die voreingestellten Etiketten lauten: <ul style="list-style-type: none"> <li>für die <math>\eta</math>-Variablen: ETA 1, ETA 2, ...</li> <li>für die <math>\xi</math>-Variablen: KSI 1, KSI 2, ...</li> </ul> </li> <li>- LISREL kürzt alle Etiketten auf acht Zeichen: Bei freiformatigem Lesen wird rechts, bei festformatigem Lesen wird links abgeschnitten.</li> <li>- Bei freiformatigem Lesen bestehen folgende Abkürzungsmöglichkeiten: <ul style="list-style-type: none"> <li>- Sollen für die letzten <math>r</math> <math>\eta</math>- bzw. <math>\xi</math>-Variablen die voreingestellten Label beibehalten werden, so kann die Liste der Etiketten entsprechend früher durch einen Schrägstrich beendet werden.</li> <li>- Man kann eine Variable bei der Vergabe von Etiketten überspringen, indem man an ihrer Position zwei Kommas setzt. Analog kann man mit drei Kommas zwei Variablen überspringen usw.</li> </ul> </li> <li>- Beim festformatigem Lesen muß ein A-Format verwendet werden.</li> </ul> |

#### 6.6.4.9 Plotten der Fit-Funktion in Abhängigkeit von einem Parameter (Subkommando PL(ot))

Mit dem PL(ot)-Subkommando kann für jeden Parameter  $\vartheta$  (frei, fixiert oder restringiert) ein Plot der Fit-Funktion (ULS, GLS, ML, WLS oder DWLS) in Abhängigkeit von seinen Ausprägungen angefordert werden. Bei der Ermittlung des Fit-Wertes zu einer konkreten Ausprägung von  $\vartheta$  geht LISREL folgendermaßen vor: Neben den fixierten Parametern wird auch  $\vartheta$  (auf dem interessierenden Wert) konstant gehalten, bzgl. der restlichen Parameter wird die Fitfunktion unter Berücksichtigung der bestehenden Gleichheits-Restriktionen neu minimiert. Diese Auswertung führt LISREL für 11, in einem relevanten Intervall automatisch gewählte, Parameter-Ausprägungen durch. Ein Plot verursacht also annähernd den Aufwand von 11 LISREL-Analysen.

/PL parameterliste [FROM a TO b]

|                |                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
|----------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| parameterliste | Name einer LISREL-Parametermatrix (vgl. Tabelle 3.6). Die erlaubten Werte sind also: LY, LX, BE, GA, PH, PS, TE, TD.                                                                                                                                                                                                                                                                                                                                                          |
| FROM a TO b    | <p>Der Plot wird für 11 äquidistante Parameter-Ausprägungen im Intervall von a bis b erstellt.</p> <p>Per Voreinstellung werden die 11 Stützstellen in folgendem Intervall gewählt:</p> <ul style="list-style-type: none"> <li>- bei freien Parametern: im 95%-Vertrauensintervall um den geschätzten Wert</li> <li>- bei fixierten Parametern: symmetrisch um den bei Freisetzung angenommenen Wert mit der doppelten geschätzten Veränderung als Intervallbreite</li> </ul> |
| Regel:         | In einem LISREL-Lauf dürfen maximal 10 Plots angefordert werden.                                                                                                                                                                                                                                                                                                                                                                                                              |
| Beispiel:      | /pl td(4,2)                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |

### 6.6.5 Spezifikation der Ausgabe und der Schätzmethode (Subkommando OU(tput))

Das OU(tput)-Subkommando hat mehr Funktionen, als sein Name vermuten läßt. Im einzelnen kann man damit:

- die Schätzmethode wählen,
- optionale Ausgaben anfordern,
- Merkmale des Programmablaufs kontrollieren (z.B. Breite einer Ausgabezeile, maximal erlaubte Rechenzeit, maximale Anzahl von Iterationen, Konvergenzkriterium),
- das Sichern einer LISREL-Matrix in eine externe Textdatei anfordern,
- eine automatische Modellmodifikation anfordern.

Das OU-Subkommando ist **obligatorisch** und muß stets am Ende einer LISREL-Problembeschreibung stehen<sup>1</sup>. Falls alle Voreinstellungen beibehalten werden sollen, besteht es nur aus dem Subkommandonamen:

/OU

Die Syntaxdarstellung des OU-Subkommandos ist aus drucktechnischen Gründen umgebrochen. Für ein konkretes OU-Subkommando gelten die allgemeinen Regeln zum Gebrauch von Fortsetzungszeilen (siehe Abschnitt 6.3).

```
/OU [ME={IV | TS | UL | GL | ML | WL | DW}] [NS] [RO] [RC=c] [AD={10 | c | OFF}] [SO]
[SE] [TV] [PC] [RS] [EF] [MR] [MI] [FS] [SS] [SC] [ALL]
[TM={60 | s}] [IT=i] [EP={0.000001 | e}] [{TO | WP}] [ND={3 | d}]
[mn='dateiname'] [mn=...]
[AM] [SL={1 | w}]
```

#### Kontrolle der Schätzmethode:

|           |                                                                                                                                                                                                                                                                             |
|-----------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| ME        | Schätzmethode<br>Voreinstellung: ML, mit folgenden Ausnahmen: <ul style="list-style-type: none"> <li>- Bei Verwendung des AC-Subkommandos ist die WLS-Methode voreingestellt.</li> <li>- Bei Verwendung des AV-Subkommandos ist die DWLS-Methode voreingestellt.</li> </ul> |
| IV        | Es werden nur die (nicht-iterativen) IV-Schätzer berechnet.                                                                                                                                                                                                                 |
| TS        | Es werden nur die (nicht-iterativen) TSLS-Schätzer berechnet.                                                                                                                                                                                                               |
| UL        | Ausgehend von Startwerten nach der IV-Methode werden iterativ ULS-Schätzer bestimmt.                                                                                                                                                                                        |
| GL        | Ausgehend von Startwerten nach der TSLS-Methode werden iterativ GLS-Schätzer bestimmt.                                                                                                                                                                                      |
| <u>ML</u> | Ausgehend von Startwerten nach der TSLS-Methode werden iterativ ML-Schätzer bestimmt.                                                                                                                                                                                       |
| WL        | Ausgehend von Startwerten nach der TSLS-Methode werden iterativ WLS-Schätzer bestimmt.<br>Diese Methode ist voreingestellt, wenn mit Hilfe des AC-Subkommandos die geschätzten asymptotischen Kovarianzen der zu analysierenden Assoziationsmaße eingelesen worden sind.    |
| DW        | Ausgehend von Startwerten nach der TSLS-Methode werden iterativ DWLS-Schätzer bestimmt.<br>Diese Methode ist voreingestellt, wenn mit Hilfe des AV-Subkommandos die geschätzten asymptotischen Varianzen der zu analysierenden Assoziationsmaße eingelesen worden sind.     |

<sup>1</sup> Wenn im Rahmen *eines* LISREL-Aufrufs mehrere Probleme bearbeitet werden sollen, folgen nach dem abschließenden OU-Subkommando zum ersten Problem noch die Subkommandos zu den restlichen Problemen.

NS Keine automatische Berechnung von Startwerten. Iterative Schätz-Algorithmen beginnen mit den vom Benutzer per ST-, VA- oder MA-Subkommando spezifizierten Startwerten.

RO, RC Wenn die zu analysierende Matrix  $S$  nicht positiv definit ist, wird an ihrer Stelle folgende Matrix analysiert:

$$S + c(\text{diag}[S])$$

Als Wert für  $c$  wird zunächst die sogenannte Ridge-Konstante eingesetzt, die auf 0.001 voreingestellt ist und mit Hilfe des RC-Steuerparameters anders festgelegt werden kann. Nötigenfalls wird der Wert von  $c$  solange mit 10 multipliziert, bis die resultierende Matrix positiv definit ist.

RO Mit RO (Ridge-Option) wird verlangt, daß LISREL auf jeden Fall statt  $S$  die oben angegebene Matrix analysieren soll.

RC Erlaubt die Wahl der Ridge-Konstanten  
Voreinstellung: 0.001

AD Bei der iterativen Suche nach dem globalen Minimum einer Fit-Funktion können prinzipiell auch Punkte außerhalb des zulässigen Parameterraumes erreicht werden (z.B. mit negativen Varianzen für einzelne Variablen). LISREL prüft daher nach jeweils  $c$  Iterationen die Zulässigkeit und bricht eventuell das Schätzverfahren ab.

$c$  Erlaubte Werte: 0,1,2,3,...  
Voreinstellung: 10

OFF Die Zulässigkeitsprüfung wird abgeschaltet.

SO Verifikation der Maßeinheiten für latente Variablen abschalten (scaling check off)  
Per Voreinstellung prüft LISREL, ob in dem zu analysierenden Modell für jede latente Variable mit einer der in Abschnitt 3.8 beschriebenen Methoden eine Maßeinheit definiert worden ist, und bricht bei negativem Prüfergebnis die Bearbeitung ab. Bei einigen speziellen Modellen werden jedoch Maßeinheiten für latente Variablen auf andere Weise identifiziert, so daß in diesen Fällen die voreingestellte Prüfung mit dem logischen Steuerparameter SO abgeschaltet werden muß.

Beispiel: /OU ME=UL  
Hier wird an Stelle der voreingestellten ML-Schätzmethode das voraussetzungsärmere ULS-Verfahren gewählt.

### Optionale Ausgaben:

SE Standardfehler für die Parameterschätzungen

TV T-Werte für die Parameterschätzungen

PC Korrelationen der Parameter-Schätzer

RS Die vom Modell implizierte Kovarianzmatrix  $\hat{\Sigma}$ , die Matrix  $S - \hat{\Sigma}$  der Residuen, die Matrix der normalisierten Residuen und der Q-Plot

EF Totale Effekte und indirekte Effekte

MR Kovarianzen der Y-Variablen mit den  $\eta$ -Variablen, Kovarianzen der Y-Variablen mit den  $\xi$ -Variablen, Kovarianzen der X-Variablen mit den  $\eta$ -Variablen, Kovarianzen der X-Variablen mit den  $\xi$ -Variablen

MI Modifikationsindikatoren und prognostizierte Veränderungen bei Freisetzung der Parameter

FS Koeffizienten zur Schätzung von Faktorwerten

|     |                                                                                                                                     |
|-----|-------------------------------------------------------------------------------------------------------------------------------------|
| SS  | Standardisierte Lösung<br>Latente Variablen sind standardisiert, manifeste Variablen haben noch die ursprüngliche Metrik (Varianz). |
| SC  | Vollständig standardisierte Lösung<br>Latente und manifeste Variablen sind standardisiert.                                          |
| ALL | Alle Ausgaben                                                                                                                       |

Hinweis: Per Voreinstellung produziert LISREL für eine Kovarianzstrukturanalyse folgende Ausgaben:

- die (eingelesene oder berechnete) zu analysierende Matrix,
- Parameter-Spezifikationen (fixiert, restringiert oder frei),
- geschätzte Startwerte für die Parameter
- die Ergebnisse der iterativen Schätzung (inkl. Kovarianzen der  $\eta$ -Variablen mit den  $\xi$ -Variablen)
- quadrierte multiple Korrelationen und Determinationskoeffizienten für alle Strukturgleichungen und faktorenanalytischen Bestimmungsgleichungen,
- globale Maße für die Anpassung des Modells an die Daten,
- Statistiken zur Verteilung der Residuen.

Beispiel: `/ou se tv rs mi`  
Hier werden als zusätzliche Ausgaben angefordert: Standardfehler, T-Werte, Residuen und Modifikationsindikatoren.

### Kontrolle über verschiedene Merkmale des Programmablaufs:

|           |                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
|-----------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| TM        | Maximal erlaubte Rechenzeit in CPU-Sekunden, Voreinstellung: 60                                                                                                                                                                                                                                                                                                                                                                                           |
| IT        | Maximale Anzahl von Iterationen im Schätz-Algorithmus<br>Voreinstellung: Das Dreifache der Anzahl freier Parameter                                                                                                                                                                                                                                                                                                                                        |
| EP        | Konvergenz-Kriterium Epsilon, Voreinstellung: 0.000001                                                                                                                                                                                                                                                                                                                                                                                                    |
| TO,WP     | Umschalten zwischen 80- und 132-spaltiger Ergebnisausgabe<br>Hinweise: - Der SPSS-Systemparameter WIDTH zur Steuerung der SPSS-Ausgabebreite, der mit dem SPSS-Kommando SET festgelegt werden kann, hat <b>keinen</b> Effekt auf die Breite der LISREL-Ausgabe.<br>- Eine Änderung der Voreinstellung (80 Spalten) auf (132 Spalten) gilt nur für das aktuelle LISREL-Problem, hat also (im Unterschied zum SPSS-Kommando SET) keine andauernden Effekte. |
| <u>TO</u> | Schmale Ausgabe (terminal output, maximal 80 Spalten pro Zeile)                                                                                                                                                                                                                                                                                                                                                                                           |
| <u>WP</u> | Breite Ausgabe (wide print, maximal 132 Spalten pro Zeile)                                                                                                                                                                                                                                                                                                                                                                                                |
| ND        | Anzahl der Dezimalstellen in der Ergebnisausgabe, Voreinstellung: 3                                                                                                                                                                                                                                                                                                                                                                                       |
| d         | Erlaubte Werte: 0..8                                                                                                                                                                                                                                                                                                                                                                                                                                      |

Beispiel: `/ou wp`  
Umschalten auf 132-spaltige Ergebnisausgabe

**Ausgabe von LISREL-Matrizen in externe Textdateien:**

|           |                                                                                                                              |                                                                                                                                                                                                                                                                                                                            |
|-----------|------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| mn        | Name der auszugebenden Matrix (matrix name)<br>Mögliche Angaben:<br>- LY, LX, BE, GA, PH, PS, TE, TD<br>- MA<br>- SI<br>- EC | Schätzergebnisse für die Parametermatrizen gemäß Tabelle 3.6<br>Die analysierte Assoziationsmatrix (nach der Selektion und/oder Umordnung von Variablen per SE-Subkommando)<br>Die vom geschätzten Modell implizierte Kovarianzmatrix $\hat{\Sigma}$<br>Die geschätzte asymptotische Kovarianzmatrix der Parameterschätzer |
| dateiname | Name der Textdatei, in die geschrieben werden soll <sup>1</sup>                                                              |                                                                                                                                                                                                                                                                                                                            |

Beispiel: `/ou se tv mi rs ef mr lx='lx.dat'`  
 Hier wird neben optionalen Ergebnissen die Ausgabe der geschätzten Ladungsmatrix  $\Lambda_x$  in die Textdatei LX.DAT angefordert.  
 Beim hypothetischen Modell aus Beispiel 1.3 (vgl. Abschnitt 5.1) erhält man folgende Datei:

```
(6D13.6) LX
0.100000D+01 0.000000D+00 0.000000D+00 0.129138D+01 0.000000D+00 0.000000D+00
0.919803D+00 0.109208D+01 0.000000D+00 0.000000D+00 0.100000D+01 0.000000D+00
0.000000D+00 0.107908D+01 0.000000D+00 0.000000D+00 0.000000D+00 0.100000D+01
0.000000D+00 0.000000D+00 0.143652D+01
```

Die Ausgabematrix ist also im FORTRAN-Format (6D13.6) als "langer Vektor" gespeichert worden.

<sup>1</sup> Bei Fertigstellung dieses Manuskriptes galt für die BS2000-Version von LISREL 7 unter SPSS 4.1: Es müssen permanente Dateien verwendet werden; diese dürfen noch nicht existieren. Um eine eventuell bereits existierende Altversion der Ausgabedatei vor dem LISREL-Aufruf löschen zu lassen, kann man folgendes SPSS-Kommando verwenden:

```
HOST 'DELETE-FILE dateiname'
```

Beispiel: `host 'delete-file lx.dat'`  
`lisrel`  
`/. . .`  
`/ou se tv mi rs ef mr lx='lx.dat'`

**Automatische Modellmodifikation:**

- AM** Einschalten der automatischen Modellmodifikation  
Das Modell wird sukzessive liberalisiert, indem pro Schritt von den bislang fixierten oder restringierten Parametern derjenige mit dem höchsten Modifikationsindikator freigesetzt wird, sofern der Modifikationsindikator signifikant ist bei dem Niveau, das der Steuerparameter SL angibt.  
Mit dem Subkommando NF kann das Freisetzen für bestimmte Parameter verhindert werden.
- SL** Signifikanzniveau für die Modifikationsindikatoren in Prozent, Voreinstellung: 1  
Die automatische Modifikation endet, wenn für keinen Modifikationsindex das Signifikanzniveau kleiner als w% ist.
- Beispiel:** /ou am sl=5  
Es wird eine automatische Modellmodifikation angefordert, bei der das Model so lange liberalisiert werden soll, bis auf dem 5% - Niveau kein Modifikationsindikator mehr signifikant ist.

## Übungsaufgaben

### Aufgabe 6.1 (Simulationen zum polychorischen Korrelationskoeffizienten)

In dieser Aufgabe zu PRELIS soll die Überlegenheit des polychorischen bzw. polyserialen Korrelationskoeffizienten bei ordinalen Daten gegenüber alternativen Koeffizienten demonstriert werden.

Produzieren Sie wie in Abschnitt 4.3 Simulationsdaten zum Latent-State-Modell der Angst, wobei jedoch nur die Variablen  $X_1$  und  $X_2$  per RECODE auf Ordinalniveau heruntertransformiert werden sollen. Damit stehen also vier manifeste Variablen zur Verfügung, von denen zwei metrische Skalenqualität besitzen ( $X_3$  und  $X_4$ ) und zwei ordinalskaliert sind ( $X_1$  und  $X_2$ ). Die theoretischen Korrelationen zwischen den ursprünglichen (nicht-rekodierten) Variablen können mit den Rechenregeln aus Abschnitt 2.1 bestimmt werden:

|       | $X_1$             | $X_2$             | $X_3$             | $X_4$ |
|-------|-------------------|-------------------|-------------------|-------|
| $X_1$ | 1                 |                   |                   |       |
| $X_2$ | $\approx 0.40825$ | 1                 |                   |       |
| $X_3$ | $\approx 0.21333$ | $\approx 0.26128$ | 1                 |       |
| $X_4$ | $\approx 0.26128$ | 0.32              | $\approx 0.40825$ | 1     |

Ziehen Sie eine Zufallsstichprobe der Größe  $N = 500$  und werten Sie die Daten folgendermaßen aus:

- Berechnen Sie mit der SPSS-Prozedur CORRELATIONS die empirische Korrelationsmatrix für  $X_1$  bis  $X_4$  vor und nach der Rekodierung von  $X_1$  und  $X_2$ .
- Analysieren Sie die Variablen  $X_1$  bis  $X_4$  (nach der Rekodierung) mit PRELIS. Deklarieren Sie dabei die Variablentypen korrekt als ordinal bzw. metrisch und akzeptieren Sie als Typ der zu schätzenden Assoziationsmatrix die Voreinstellung CORRELATIONS. Vergleichen Sie die von PRELIS ermittelte Korrelationsmatrix mit der unter a) für die rekodierten Variablen berechneten Matrix. Welche Matrix enthält die besseren Schätzer. Wie sind die Unterschiede zu erklären?
- Wählen Sie in einem weiteren PRELIS-Lauf folgende Spezifikationen:
  - TYPE=POLYCHOR
  - Im PRINT-Subkommando: Ausgabe der geschätzten asymptotischen Kovarianzmatrix der Korrelations-Schätzer in die Ergebnisliste)
  - Im WRITE-Subkommando: Ausgabe der geschätzten asymptotischen Kovarianzmatrix der Korrelations-Schätzer in die Datei POL.ACOV

Vergleichen Sie die geschätzten polychorischen bzw. polyserialen Korrelationen mit den wahren Werten und mit den weiter oben ermittelten Schätzern.

**Aufgabe 6.2** (Latent-State-Modell der Angst; Auswertung mit LISREL)

Untersuchen Sie mit LISREL das Latent-State-Modell der Angst in Beispiel 1.2 anhand der empirischen Kovarianzmatrix in Abschnitt 2.2. Die Identifikation des Modells wurde schon in Abschnitt 2.3 nachgewiesen. Zum Einlesen der Daten können Sie den ersten Teil des Programms aus Aufgabe 2.3 verwenden. Wählen Sie für  $\Theta_{\delta}$  die Matrix-Form SY(mmetrisch), damit für die fixierten Kovarianzen der  $\delta$ -Variablen Modifikationsindikatoren ausgegeben werden. Lassen Sie sich die Standardfehler und T-Werte zu den geschätzten Parametern, die Residuen und die Modifikationsindikatoren ausgeben.

Liefern die Ergebnisse Hinweise auf Schwächen des Modells? Welche Empfehlungen ergeben sich für die Verbesserung des Meßinstrumentes?

Unter Annahme der Modellgültigkeit läßt sich die Reliabilität von  $X_1$  schätzen durch:

$$\hat{Rel}(X_1) = \frac{\hat{\phi}_{11}}{\hat{\phi}_{11} + \hat{\theta}_{11}^{(\delta)}}$$

Analoge Formeln gelten für die übrigen manifesten Variablen. Berechnen Sie alle geschätzten Reliabilitäten.

**Aufgabe 6.3** (Berufliche Ambitionen von Jugendlichen; Auswertung mit LISREL)

Untersuchen Sie mit LISREL das Modell aus Beispiel 1.1 zum Bedingungs hintergrund der beruflichen Ambitionen von Jugendlichen anhand der Korrelationsmatrix in Abschnitt 2.2. Beweisen Sie bitte zunächst die Identifikation des Modells. Präsentieren Sie der Übung halber die Daten als Teil des LISREL-Aufrufs unter Verwendung des voreingestellten Freiformats. Vergeben Sie geeignete Etiketten für die manifesten Variablen.

Wie beurteilt LISREL die bei klassisch-pfadanalytischer Auswertung ungeprüft vorausgesetzte Rekursivitäts-Annahme, die in unserem Beispiel äquivalent ist zur Annahme unkorrelierter Fehler  $\zeta_1$  und  $\zeta_2$  (vgl. Abschnitt 1.1)?

Hinweise: - In der einzulesenden Korrelationsmatrix stehen die Variablen nicht in der richtigen Reihenfolge ( $\rightarrow$  Subkommando SE).  
- Bei der Parametermatrix B müssen Sie die voreingestellte Form ZE abändern auf SD.

## Lösungsvorschläge für die Übungsaufgaben

### Aufgabe 2.1

$$\begin{aligned}
 \gamma &= b_{YX} = \frac{\text{Cov}(X, Y)}{\text{Var}(X)} \\
 &= \frac{\text{Cov}(X_T + \delta, \gamma_T X_T + \zeta_T)}{\text{Var}(X_T + \delta)} \\
 &= \frac{\gamma_T \text{Var}(X_T)}{\text{Var}(X_T) + \text{Var}(\delta)} \\
 &= \gamma_T \text{Rel}(X) < \gamma_T
 \end{aligned}$$

### Aufgabe 2.2

```

title Berufl. Ambit. von Jugendl., Auswertung per Mult. Regr.
matrix data var=intellnz bldngvat berufvat schullst ambition
 /contents=corr /n=767
begin data
1
.277 1
.250 .611 1
.572 .294 .248 1
.335 .303 .331 .478 1
end data
regression matrix=in(*)
 /dep=schullst /enter=intellnz bldngvat
 /dep=ambition /enter=berufvat schullst

```

### Aufgabe 2.3

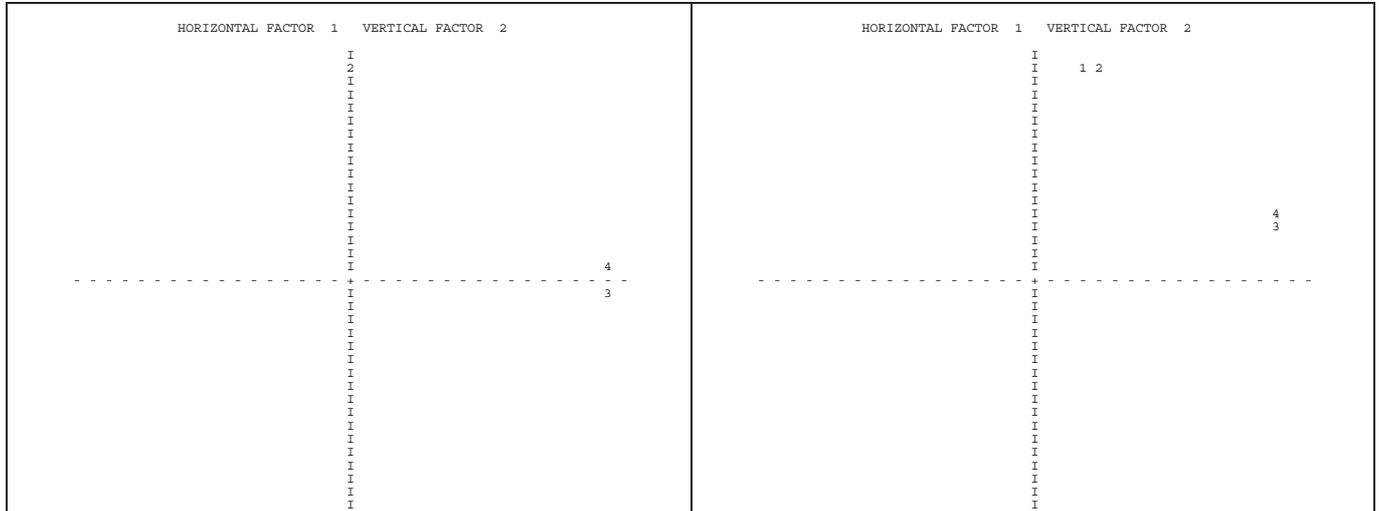
```

title Latent-State-Modell d. Angst, Auswert. per expl. Fakt.an.
matrix data var=x1 x2 x3 x4
 /contents=cov /n=179
begin data
24.670
21.895 25.135
10.353 10.624 27.239
11.665 12.636 25.258 28.683
end data
mconvert
factor matrix=in(cor=*)
 /print=corr def repr
 /plot=eigen rotation(1,2)
 /extraction=paf /rotation=oblimin /rotation=varimax

```

**Anmerkungen:**

**Zu a)** Von den beiden Lösungen ist die schiefwinklige deutlich zu bevorzugen. Dies läßt sich am eindrucksvollsten durch die beiden Plots der Ladungen im Faktorenraum belegen (links: schiefwinklig, rechts: orthogonal):



**Zu b)** Die schiefwinklige Lösung ist mit unseren Modellannahmen "gut vereinbar", obwohl wir bei dem rein exploratorischen Vorgehen gar keine Möglichkeit hatten, das Modell zu formulieren: Die Anzahl der Faktoren wurde nach dem Eigenwertkriterium bestimmt, und das Ladungsmuster resultierte aus dem Extraktionsverfahren und der Rotation nach dem Prinzip der Einfachstruktur. Erfreulicherweise sind im Ladungsmuster (siehe PATTERN MATRIX) die Schätzer zu den im Modell auf Null gesetzten Parametern tatsächlich sehr klein:

| PATTERN MATRIX: |          |          |
|-----------------|----------|----------|
|                 | FACTOR 1 | FACTOR 2 |
| X1              | -.00950  | .93758   |
| X2              | .01110   | .93675   |
| X3              | .96084   | -.03055  |
| X4              | .93862   | .03537   |

| FACTOR CORRELATION MATRIX: |          |          |
|----------------------------|----------|----------|
|                            | FACTOR 1 | FACTOR 2 |
| FACTOR 1                   | 1.00000  |          |
| FACTOR 2                   | .47819   | 1.00000  |

Die Korrelation zwischen den Faktoren (latenten Variablen) wird auf 0.48 geschätzt.

**Zu c)** Zwar bietet die SPSS-Prozedur FACTOR für unser Modell keinen Gültigkeits-Test an, aber immerhin können wir die Residuen inspizieren, d.h. die Unterschiede zwischen den vom Modell reproduzierten und den tatsächlich beobachteten Korrelationen.

| REPRODUCED CORRELATION MATRIX: |         |         |         |         |
|--------------------------------|---------|---------|---------|---------|
|                                | X1      | X2      | X3      | X4      |
| X1                             | .87062* | .00038  | .00623  | -.00639 |
| X2                             | .87888  | .88756* | -.00626 | .00642  |
| X3                             | .39315  | .41228  | .89607* | .00031  |
| X4                             | .44490  | .46419  | .90332  | .91401* |

THE LOWER LEFT TRIANGLE CONTAINS THE REPRODUCED CORRELATION MATRIX; THE DIAGONAL, COMMUNALITIES; AND THE UPPER RIGHT TRIANGLE, RESIDUALS BETWEEN THE OBSERVED CORRELATIONS AND THE REPRODUCED CORRELATIONS.

THERE ARE 0 ( .0%) RESIDUALS (ABOVE DIAGONAL) THAT ARE > 0.05

Die Residuen sind allesamt sehr klein und geben keine Hinweise auf Schwachstellen des Modells.

### Aufgabe 5.1

Die zur Lösung erforderlichen LISREL-Aufrufe können durch leichte Veränderungen des Programms in Abschnitt 5.1 gewonnen werden.

#### Zu a)

Die ML-Schätzer liegen näher an den wahren Werten. Dies überrascht nicht, da gemäß Abschnitt 2.4 bei der hier gegebenen Voraussetzung der multivariaten Normalverteilung der manifesten Variablen die ML-Schätzer asymptotisch effizient sind.

#### Zu b)

Aufgrund der schon in Abschnitt 5.1.1 wiedergegebenen Fitbeurteilung ( $\chi^2 = 29.10$  bei  $df = 33$ ) kann das Modell mit  $\gamma_{11}$  und  $\gamma_{12}$  als freien Parametern beibehalten werden. Aus diesem uneingeschränkten Modell ergibt sich durch die  $H_0: \gamma_{11} = \gamma_{12}$  ein eingeschränktes Modell mit 34 Freiheitsgraden. Die Voraussetzungen für die in Abschnitt 5.2 beschriebene  $D^2$ -Teststrategie liegen vor. Im Modell-Spezifikations-Block des LISREL-Aufrufs zum uneingeschränkten Modell muß folgendes Subkommando eingefügt werden:

$$/eq\ ga(1,1)\ ga(1,2)$$

Für das eingeschränkte Modell liefert LISREL  $\chi^2 = 30.01$ , so daß die Prüfgröße den Wert  $D^2 = 30.01 - 29.10 = 0.91$  annimmt. Da dieser Wert kleiner als 3.84 ist ( $\approx 5\%$ -Fraktile der  $\chi^2$ -Verteilung mit  $df = 1$ ), kann die  $H_0$  beibehalten werden.

## Aufgabe 6.1

Das folgende SPSS-Programm führt alle benötigten Rechnungen aus:

```

set seed=5555
set mxloops=10000
input program
+ loop #i=1 to 500
+ compute #ksi=normal(1)
+ compute #ksi1=0.8*#ksi + normal(0.6)
+ compute #ksi2=0.8*#ksi + normal(0.6)
+ compute X1=0.7071*#ksi1 + normal(1)
+ compute X2=#ksi1 + normal(1)
+ compute X3=0.7071*#ksi2 + normal(1)
+ compute X4=#ksi2 + normal(1)
+ end case
+ end loop
+ end file
end input program

correlations x1 to x4 /stat=descr

recode x1 (lo thru -2 = 1)
 (-2 thru -1.5 = 2)
 (-1.5 thru 0 = 3)
 (0 thru 2 = 4)
 (2 thru hi = 5)/
 x2 (lo thru -2.5 = 1)
 (-2.5 thru -1 = 2)
 (-1 thru 1 = 3)
 (1 thru 1.5 = 4)
 (1.5 thru 2.5 = 5)
 (2.5 thru hi = 6)

correlations x1 to x4 /stat=descr

prelis /var = x1 x2 (ord) x3 x4 (con)

prelis /var = x1 x2 (ord) x3 x4 (con)
/type = polychor
/print = acov
/write = acov 'pol.acov'

```

Das Programm produziert typischerweise Ergebnisse wie in folgender Tabelle:

|               | theoret.<br>Wert | CORRELA-<br>TIONS:<br>PM <sup>1</sup> vor<br>Rekod. | CORRELA-<br>TIONS:<br>PM nach<br>Rekod. | PRELIS:<br>PM mit nor-<br>mal scores | PRELIS:<br>polych. bzw.<br>polyserial |
|---------------|------------------|-----------------------------------------------------|-----------------------------------------|--------------------------------------|---------------------------------------|
| $r(X_1, X_2)$ | .4083            | .3860                                               | .3248                                   | .335                                 | .385                                  |
| $r(X_1, X_3)$ | .2133            | .1884                                               | .1842                                   | .183                                 | .198                                  |
| $r(X_1, X_4)$ | .2613            | .2449                                               | .2368                                   | .240                                 | .258                                  |
| $r(X_2, X_3)$ | .2613            | .2434                                               | .2235                                   | .234                                 | .246                                  |
| $r(X_2, X_4)$ | .3200            | .3199                                               | .2891                                   | .297                                 | .313                                  |
| $r(X_3, X_4)$ | .4083            | .4110                                               | .4110                                   | .411                                 | .411                                  |

<sup>1</sup> PM = Produkt-Moment-Korrelation

Diese beispielhaften Ergebnisse illustrieren folgende Tendenzen bei der Korrelationsschätzung aus ordinalen Daten:

- Die schlechtesten Schätzergebnisse erhält man bei der Berechnung von Produkt-Moment-Korrelationen.
- Leichte Besserungen ergeben sich durch die Verwendung von normal scores.
- Befriedigende Präzision bieten nur die polychorischen bzw. polyserialen Schätzer.

Damit werden die Simulationsergebnisse von Jöreskog & Sörbom (1988), über die in Abschnitt 4.1.2.3 berichtet wurde, voll bestätigt.

## Aufgabe 6.2

```
matrix data var=x1 x2 x3 x4
 /contents=cov /n=179
begin data
24.670
21.895 25.135
10.353 10.624 27.239
11.665 12.636 25.258 28.683
end data

lisrel
/"Latent-State-Modell der Angst"
/da ni=4
/mo nx=4 nk=2 td=sy
/va 1.0 lx(1,1) lx(2,1) lx(3,2) lx(4,2)
/eq td(1,1) td(2,2) td(3,3) td(4,4)
/ou se tv rs mi
```

Zur Statistik: Wie schon in Abschnitt 2.5.5 diskutiert, scheinen die beiden aus der zweiten Testhälfte stammenden Indikatoren ( $X_2$  und  $X_4$ ) über den Einfluß der latenten Variablen hinaus eine gemeinsame Wurzel zu haben. Dies führt trotz akzeptablem Wert der  $\chi^2$ -Statistik (7.53,  $p = .274$ ) zum Modifikationsindikator 4.55 für den auf 0 fixierten Parameter  $\theta_{42}^{(\delta)}$ , der bei Testung zum Niveau  $\alpha = 0.05$  bedeutsam ist (kritische Grenze: 4). Wird der Parameter  $\theta_{42}^{(\delta)}$  freigesetzt (zusätzliches Subkommando im Modell-Spezifikations-Block: /FR TD(4,2)), liefert LISREL dafür den Schätzwert 0.849. Aus dieser Kovarianz läßt sich unter Verwendung der von LISREL auf 2.866 geschätzten gemeinsamen Varianz von  $\delta_2$  bzw.  $\delta_4$  die geschätzte Korrelation der Residuen berechnen (vgl. die Definition einer Korrelation in Abschnitt 2.1):

$$\text{Cor}(\delta_2, \delta_4) \approx \frac{0.849}{2.866} \approx 0.30$$

Bei einer Weiterentwicklung des Meßinstrumentes sollte also angestrebt werden, das konstrukt fremde Spezifikum in der zweiten Testhälfte zu identifizieren und zu eliminieren.

Unter Verwendung des verbesserten Modells (Parameter  $\theta_{42}^{(\delta)}$  frei) erhalten wir für die Reliabilität von  $X_1$  den Schätzwert:

$$\hat{\text{Rel}}(X_1) = \frac{\hat{\phi}_{11}}{\hat{\phi}_{11} + \hat{\theta}_{11}^{(\delta)}} \approx \frac{21.782}{21.782 + 2.866} \approx 0.88$$

### Aufgabe 6.3

Wir sollten uns zunächst der lästigen Aufgabe widmen, die Identifikation des Modells nachzuweisen. Diese steht zwar wegen der Rekursivität des Pfadmodells außer Frage, doch ein wenig Übung mit derlei Rechnungen ist sicherlich nützlich.

i) Nachweis der Identifikation von  $\gamma_{11}$  und  $\gamma_{12}$ :

Wegen der Modellannahmen und der Rechenregeln in Abschnitt 2.1 gelten folgende Gleichungen:

$$\begin{aligned} \text{Cov}(X_1, Y_1) &= \text{Cov}(X_1, \gamma_{11}X_1 + \gamma_{12}X_2 + \zeta_1) \\ &= \gamma_{11}\text{Var}(X_1) + \gamma_{12}\text{Cov}(X_1, X_2) \\ \text{Cov}(X_2, Y_1) &= \text{Cov}(X_2, \gamma_{11}X_1 + \gamma_{12}X_2 + \zeta_1) \\ &= \gamma_{11}\text{Cov}(X_1, X_2) + \gamma_{12}\text{Var}(X_2) \end{aligned}$$

Das resultierende System mit zwei Gleichungen kann nach den beiden Unbekannten  $\gamma_{11}$  und  $\gamma_{12}$  aufgelöst werden, sofern die folgende Koeffizienten- bzw. Kovarianzmatrix regulär ist:

$$\begin{bmatrix} \text{Var}(X_1) & \text{Cov}(X_1, X_2) \\ \text{Cov}(X_1, X_2) & \text{Var}(X_2) \end{bmatrix}$$

Diese Regularität ist garantiert, sofern  $X_1$  und  $X_2$  linear unabhängig sind. Lineare Unabhängigkeit ist aber in einem sinnvollen Modell für jedes Paar manifester Variablen gegeben.

ii) Nachweis der Identifikation von  $\gamma_{23}$  und  $\beta_{21}$ :

$$\begin{aligned} \text{Cov}(X_3, Y_2) &= \text{Cov}(X_3, \gamma_{23}X_3 + \beta_{21}Y_1 + \zeta_2) \\ &= \gamma_{23}\text{Var}(X_3) + \beta_{21}\text{Cov}(X_3, Y_1) \\ \text{Cov}(Y_1, Y_2) &= \text{Cov}(Y_1, \gamma_{23}X_3 + \beta_{21}Y_1 + \zeta_2) \\ &= \gamma_{23}\text{Cov}(X_3, Y_1) + \beta_{21}\text{Var}(Y_1) \end{aligned}$$

Das erhaltene Gleichungssystem läßt sich mit den Argumenten aus i) nach den unbekannt Parametern  $\gamma_{23}$  und  $\beta_{21}$  auflösen.

c) Nachweis der Identifikation von  $\psi_{11}$  und  $\psi_{22}$ :

$$\begin{aligned} \text{Var}(Y_1) &= \text{Var}((\gamma_{11}X_1 + \gamma_{12}X_2) + \zeta_1) \\ &= \text{Var}(\gamma_{11}X_1 + \gamma_{12}X_2) + \psi_{11} \\ &= \gamma_{11}^2\text{Var}(X_1) + 2\gamma_{11}\gamma_{12}\text{Cov}(X_1, X_2) + \gamma_{12}^2\text{Var}(X_2) + \psi_{11} \\ \Rightarrow \psi_{11} &= \text{Var}(Y_1) - \gamma_{11}^2\text{Var}(X_1) - 2\gamma_{11}\gamma_{12}\text{Cov}(X_1, X_2) - \gamma_{12}^2\text{Var}(X_2) \end{aligned}$$

Da  $\gamma_{11}$  und  $\gamma_{12}$  aus i) schon bekannt sind, kann  $\psi_{11}$  berechnet werden. Analog läßt sich auch  $\psi_{22}$  bestimmen.

Mit dem folgenden Programm wird die gewünschte LISREL-Analyse ausgeführt. Instruktionsgemäß liegt die zu analysierende Korrelationsmatrix innerhalb des LISREL-Aufrufs. Folglich werden keine Daten von SPSS übernommen. Beim Aufruf von LISREL muß jedoch trotzdem eine aktive SSPS-Systemdatei existieren. Daher wird zuvor mit einem kleinen SPSS-Eingabeprogramm eine Pseudo-Systemdatei erstellt.

```
input program
numeric dummy
end file
end input program

lisrel
/"Berufliche Ambitionen von Jugendlichen"
/da ni=5 no=767 ma=km
/km
/1
/.277 1
/.250 .611 1
/.572 .294 .248 1
/.335 .303 .331 .478 1
/la
/intellnz bldngvat gerufvat schullst ambition
/se
/4 5 1 2 3
/mo ny=2 nx=3 be=sd
/fi ga(1,3) ga(2,1) ga(2,2) ps(2,1)
/ou se tv ef rs mi
```

Einige Anmerkungen zu den Ergebnissen:

- Wir erhalten eine  $\chi^2$  - Prüfgröße von 5.99, der bei  $df = 3$  eine Überschreitungswahrscheinlichkeit von  $p = 0.112$  entspricht, so daß unser Modell nicht verworfen werden muß.
- Die Modifikationsindizes sind allesamt kleiner als 4, so daß die Fixierungen beibehalten werden können.  
Insbesondere ist der Modifikationsindikator zu  $\psi_{21}$  nicht signifikant, d.h. die Unabhängigkeitsannahme für  $\zeta_1$  und  $\zeta_2$  ist akzeptabel.
- Die T-Werte zu allen freien Parameter sind signifikant, unser Modell enthält also offenbar keine überflüssigen Parameter, d.h. es ist nicht über-parametrisiert.

## Literatur

- Baltes-Götz, B. (1993). *Einführung in SPSS 4.1 unter BS2000* (4 revid. Auflage). Trier: Universitäts-Rechenzentrum.
- Backhaus, K., Erichson, B., Plinke, W., Schuchard-Fischer, Chr. & Weiber, R. (1989). *Multivariate Analysemethoden* (5. revid. Auflage). Berlin: Springer.
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.
- Browne, M.W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, 37, 62-83.
- Hayduk, L.A. (1987). *Structural equation modeling with LISREL*. Baltimore: The Johns Hopkins University Press.
- Hodapp, V. (1984). *Analyse linearer Kausalmodelle*. Bern: Huber.
- Jöreskog K.G. & Sörbom, D. (1986). *LISREL VI. Analysis of linear structural relationships by the method of maximum likelihood*. Uppsala: University of Uppsala.
- Jöreskog, K. G. & Sörbom, D. (1988). *PRELIS: A preprocessor for LISREL* (2nd ed.). Mooresville, IN: Scientific Software
- Jöreskog, K. G. & Sörbom, D. (1989). *LISREL 7: A guide to the program and applications* (2nd ed.). Chicago, IL: SPSS
- Kenny, D.A. (1979). *Correlation and causality*. New York: Wiley.
- Kerchoff, A.C. (1974). *Ambition and attainment*. Rose Monograph Series.
- Kühnel, S. (1993). Lassen sich ordinale Daten mit linearen Strukturgleichungsmodellen analysieren? *ZA-Information*, 33, 29-51.
- Lee, S.Y. (1985). Analysis of covariance and correlation structures. *Computational Statistics and Data Analysis*, 2, 279-295.
- Loehlin, J.C. (1987). *Latent variable models*. Hillsdale, NJ: Lawrence Erlbaum.
- Long, J.S. (1983). *Covariance structure models*. Beverly Hills: SAGE Publications.
- Mardia, K. V. (1970). Measures of multivariate skewness und kurtosis with applications. *Biometrika*, 57, 519-530.
- Pfeifer, A. & Schmidt, P. (1987). *LISREL. Die Analyse komplexer Strukturgleichungsmodelle*. Stuttgart: Fischer.
- SPSS. (1990). *LISREL 7 and PRELIS user's guide and Reference*. Chicago, IL: SPSS.

Steyer, R. (1988). *Experiment, Regression und Kausalität: Die logische Struktur kausaler Regressionsmodelle*. Unveröff. Habilitationsschrift, Universität Trier.

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