

# Faktorenanalyse auf dem Prüfstand

Zum Zusammenhang zwischen der Verteilung  
der Fähigkeitswerte und der Schätzgenauigkeit  
klassischer Faktorenmodelle

Dissertation von Daniel Kasper

Eingereicht an der Technischen Universität Dortmund  
Fakultät 12 Erziehungswissenschaft und Soziologie

März 2012

Erstgutachter: Prof. Dr. Wilfried Bos, Institut für Schulentwicklungsforschung, Technische Universität Dortmund, Vogelpothsweg 78, 44227 Dortmund

Zweitgutachter: Prof. Dr. Ali Ünlü, Lehrstuhl für Methoden der Empirischen Bildungsforschung, TUM School of Education, Lothstraße 17, 80335 München

# Vorwort

In der Erziehungswissenschaft im Allgemeinen und der empirischen Bildungsforschung im Besonderen werden (auch) Erhebungsinstrumente zur Untersuchung von fachspezifischen Sachverhalten eingesetzt. Da in diesen Fällen der Erkenntnisgewinn unter anderem von der Beschaffenheit der Erhebungsinstrumente abhängt, sollten die Instrumente vorher definierte Qualitätsanforderungen erfüllen. Insoweit es sich bei den Erhebungsinstrumenten um quantitative Messverfahren handelt, sind diese Anforderungen in verschiedenen Testtheorien beschrieben.

Nach den Testtheorien stellt die Validität des Messverfahrens ein wesentliches Gütekriterium dar (American Psychological Association, 1954, 1966). Ein Teilaspekt der Validität ist die faktorielle Struktur des Messinstrumentes (Lienert & Raatz, 1998). Die faktorielle Struktur eines Messinstrumentes wird durch die Faktorenanalyse bestimmt. Allerdings gibt es eine Reihe unterschiedlicher Faktorenmodelle, die nicht zu gleichen Ergebnissen führen müssen. Außerdem erfüllen viele Messinstrumente der Erziehungswissenschaft nicht die Voraussetzungen der Faktorenanalyse.

Bei derzeitigem Forschungsstand bleibt unklar, wie sich die Verletzung der Voraussetzungen auf die Schätzgüte der Faktorenmodelle auswirkt. Dieses gilt vor allem, wenn die von vielen Faktorenmodellen vorausgesetzte Normalverteilung der Messwerte nicht angenommen werden kann, was insbesondere für die als kriteriumsorientierte Messung (Glaser, 1963) bezeichneten Verfahren zutrifft. Das primäre Ziel der Dissertation ist es deswegen zu überprüfen, wie sich die Verletzung der Voraussetzung normalverteilter Messwerte auf die Schätzgüte der Faktorenmodelle auswirkt.

Nach einer Darstellung der Ausgangssituation und der Zielstellung der Arbeit in Kapitel 1, werden in Kapitel 2 die theoretischen Grundlagen für die Untersuchung gelegt. Dabei wird zunächst der begriffliche Rahmen erläutert, und dann die untersuchten Faktorenmodelle genauer vorgestellt. Ein besonderes Augenmerk wird dabei auf die durch die Modelle angenommenen Verteilungseigenschaften gelegt, weswegen diesem Aspekt ein Unterkapitel

gewidmet wurde. Abschließend wird der empirische Forschungsstand zum Untersuchungsgegenstand berichtet.

In Kapitel 3 werden die methodischen Grundlagen für den empirischen Teil der Dissertation erläutert. Neben den Forschungsfragen und den Hypothesen wird vor allem auf den Mattson-Ansatz (Mattson, 1997) eingegangen, und es wird gezeigt, dass sich dieser Ansatz hervorragend eignet, um die aufgestellten Hypothesen im Rahmen einer Simulationsstudie zu überprüfen. Abgerundet wird Kapitel 3 durch eine Darstellung des genauen Aufbaus des Simulationsdesigns und die verwendeten Auswertungsmethoden.

Die Ergebnisse der Simulationsstudie sind in Kapitel 4 detailliert aufgeführt. Eine Zusammenfassung der Schrift, Schlussfolgerungen in Bezug auf die Zielstellung der Dissertation sowie Empfehlungen für die wissenschaftliche Forschung und Praxis finden sich in Kapitel 5.

Da in der Literatur für die einzelnen Modelle teilweise sehr unterschiedliche Notationen verwendet werden, wurde versucht diese zu vereinheitlichen. Eine Übersicht über die Bedeutung der verwendeten Notation findet sich im Symbolverzeichnis in Anhang E (Seite 116).

An dieser Stelle endlich darf ich jenen Menschen danken, ohne die diese Arbeit nicht zu Stande gekommen wäre. An erster Stelle ist hier mein Vater Johannes-Peter Kasper zu nennen. Es ist dein steter Glaube an mich, der mir Licht in dunklen Zeiten ist. Philipp's habe ich gleich zwei, bei denen ich mich bedanken kann: Einmal meinem Bruder, der mir die Schwachstellen unseres Systems immer wieder aufzeigt, und einmal meinem geliebten Freund, Partner, Lebensbegleiter, kurz meinem Glückstreffer, der für Wärme, Freude und Geborgenheit in meinem Leben sorgt.

Auf fachlicher Seite sind es meine Betreuer Prof. Dr. Wilfried Bos und Prof. Dr. Ali Ünlü, denen ich für viele nützliche Anregungen danken möchte. Schließlich bedanke ich mich auch bei meinem Projektleiter PD Dr. Tobias C. Stubbe und meinen Kolleginnen Anke Walzebug, Claudia Dohe und Magdalena Buddeberg, die mir gerade in der Schlussphase der Dissertation den Rücken freigehalten haben, sodass ich mich voll und ganz der Fertigstellung dieser Arbeit widmen konnte. Opus finitum (est).

# Inhaltsverzeichnis

<b>1</b>	<b>Ausgangssituation und Zielstellung</b>	<b>1</b>
<b>2</b>	<b>Theoretische Grundlagen</b>	<b>4</b>
2.1	Begrifflicher Rahmen . . . . .	4
2.2	Symbolische Zuordnungen . . . . .	9
2.3	Modelle der klassischen Faktorenanalyse . . . . .	11
2.3.1	Überblick . . . . .	11
2.3.2	Hauptkomponentenmethode . . . . .	11
2.3.3	Explorative Faktorenanalyse . . . . .	32
2.3.4	Hauptfaktorenanalyse . . . . .	50
2.4	Rotationsproblematik . . . . .	52
2.5	Kriterien zur Bestimmung der faktoriellen Struktur . . . . .	56
2.6	Verteilungseigenschaften von $\mathbf{F}$ und $\mathbf{Z}$ . . . . .	59
2.7	Empirischer Forschungsstand . . . . .	60
<b>3</b>	<b>Methodische Grundlagen</b>	<b>68</b>
3.1	Forschungsfragen . . . . .	68
3.2	Hypothesen . . . . .	70
3.3	Mattson-Ansatz . . . . .	72
3.4	Simulationsdesign . . . . .	76
3.5	Auswertungsmethoden . . . . .	82
<b>4</b>	<b>Ergebnisse</b>	<b>85</b>
4.1	Anzahl der extrahierten Dimensionen . . . . .	85
4.2	Geschätzte Schiefe der Fähigkeitswerte . . . . .	89
4.3	Fit der geschätzten Ladungsmatrizen . . . . .	96

<b>5 Zusammenfassung, Schlussfolgerungen und Empfehlungen</b>	<b>99</b>
<b>A Basissyntax zur Datensimulation</b>	<b>104</b>
<b>B Geschätzte Schiefewerte für <math>k = 4</math></b>	<b>107</b>
<b>C Geschätzte Schiefewerte für <math>k = 8</math></b>	<b>110</b>
<b>D Verteilung der absoluten Differenz</b>	<b>113</b>
<b>E Symbolverzeichnis</b>	<b>116</b>
<b>Literaturverzeichnis</b>	<b>117</b>

# Kapitel 1

## Ausgangssituation und Zielstellung

Den Ausgangspunkt für die vorliegende Dissertation bildet die von Glaser (1963) begründete Theorie der kriteriumsorientierten Messung.<sup>1</sup> Nach dieser Theorie ist ein Lehrziel genau dann erreicht, wenn eine Menge von Aufgaben richtig gelöst wird. Die notwendige Anzahl an richtig gelösten Aufgaben stellt dabei eine Idealnorm dar, d.h. sie orientiert sich nicht an empirischen Fakten, sondern wird in der Regel durch Experten festgelegt.

Um überprüfen zu können, ob ein Lehrziel erreicht ist, werden entsprechend Aufgaben zu einem kriteriumsorientierten Test zusammengestellt. Dieser Test hat bestimmte Gütekriterien zu erfüllen (Kapitel 2.1), wobei die Kontenvalidität (Inhaltsvalidität) einer der wichtigsten Eigenschaften lehrzielorientierter Tests darstellt. Darüber hinaus ist zu fordern, dass kriteriumsorientierte Tests konstruktvalide sind (Klauer, 1987).

Cronbach und Meehl (1955) unterscheiden sieben methodische Zugänge, um die Konstruktvalidität eines Tests zu bestimmen. Neben einer inhaltlich-logischen Analyse der Testelemente und einer statistisch geleiteten Interpretation von inter- sowie intraindividuellen Testergebnissen stehen die Bestimmung der konvergenten und diskriminanten Validität (Campbell & Fiske, 1959) sowie die faktorielle Validität im Vordergrund.

---

<sup>1</sup>Klauer (1987) verweist auf die Namensvielfalt dieser Theorie. So wird anstelle von kriteriumsorientierter Messung auch von lehrzielorientierter Messung oder schlichtweg von Kompetenztests gesprochen. Um einerseits die Kohäsion zu wahren und andererseits einer gewissen Situationalität gerecht zu werden, wird in der Dissertation entweder von kriteriumsorientierter Messung bzw. Testung oder von lehrzielorientierter Messung bzw. Testung gesprochen, ohne dass damit irgendeine spezifische Ausrichtung dieser Theorie angedeutet oder ausgegrenzt werden soll.

Um die faktorielle Validität eines Tests zu bestimmen, wird eine Faktorenanalyse durchgeführt, wobei die faktorielle Struktur des Tests der faktoriellen Validität entspricht. Allerdings gibt es eine Reihe unterschiedlicher Faktorenmodelle (s. Kapitel 2.3), die nicht zu gleichen Ergebnissen führen müssen. Darüber hinaus weisen die Ergebnisse von lehrzielorientierten Tests häufig eine rechtssteile Verteilung auf (Micceri, 1989), da die meisten Personen das Lehrziel erreichen und damit die Testaufgaben in der Regel richtig lösen können. Schließlich ist die bei vielen Faktorenmodellen aus der Berechnung heraus resultierende Konsequenz normalverteilter Fähigkeitswerte im Kontext von kriteriumsorientierten Tests nicht unbedingt plausibel, weil die meisten der getesteten Personen vor der Testung intensiv in Bezug auf die getestete Fähigkeit trainiert wurden (siehe beispielsweise die Vorbereitung auf die Theoretische Fahrerlaubnisprüfung; Sturzbecher, Kasper, Bönninger und Rüdell (2008)), d.h. die überwiegende Anzahl der Personen besitzt sehr wahrscheinlich eine hohe Fähigkeitsausprägung. In der Terminologie der Testpsychologie ist also anzunehmen, dass sowohl die Testwerte als auch die Fähigkeitswerte der Personen bei kriteriumsorientierten Tests nicht normalverteilt sondern rechtssteil verteilt sind, während viele Faktorenmodelle normalverteilte Fähigkeitswerte annehmen und normalverteilte Testwerte modulieren. Systematische Untersuchungen dazu, welches Faktorenmodell unter diesen Bedingungen die faktorielle Validität des Tests am genauesten schätzt, liegen kaum vor und berücksichtigen in der Regel nur einen der genannten Aspekte. Darüber hinaus finden sich keine Analysen, bei denen im Rahmen einer Simulationsstudie in einem ersten Schritt unterschiedliche faktorielle Strukturen konstruiert wurden, um in einem zweiten Schritt ausgehend von diesen faktoriellen Strukturen Testdatensätze zu generieren, die dann im dritten Schritt mit unterschiedlichen Faktorenmodellen analysiert wurden, um zu überprüfen, wie genau die Faktorenmodelle die den Testdatensätzen zugrundeliegende faktorielle Struktur schätzen. Eine solche Modellierung ist aber notwendig, um festzustellen, wie hoch die faktorielle Validität ist, wenn sie mit unterschiedlichen Faktorenmodellen geschätzt wird. Aus diesem offenen Forschungsstand wurde die Zielstellung der Dissertation erarbeitet.

Primäres Ziel der Dissertation ist es, sowohl auf analytischer als auch auf empirischer Ebene zu überprüfen, welches Faktorenmodell die faktorielle Struktur eines Tests am genauesten schätzt, wenn sowohl die Fähigkeitswerte als auch die Testwerte rechtssteil verteilt sind. Damit liefert diese Doktorarbeit einen Hinweis darauf, wie gut unterschiedliche Faktorenmodelle zur Beurteilung der faktoriellen Validität von kriteriumsorientierten Tests geeignet sind.

Auch wenn sich das primäre Ziel der Dissertation auf die Analyse von kriteriumsorientierten Testdaten bezieht, ergeben sich aus der Dissertation starke Implikationen für andere, diskret skalierte Messinstrumente, wie etwa jene die bei Large-Scale-Assessment-Studien wie PISA (OECD, 2009b), PIRLS (Mullis, Martin, Kennedy, Trong & Sainsbury, 2009) oder TIMSS (Mullis et al., 2005) verwendet werden, weil diese Messinstrumente zumindest in Bezug auf die Skalierung und Verteilung der Testwerte sehr ähnliche Eigenschaften aufweisen wie kriteriumsorientierte Testdaten. Insbesondere sind die Testdaten sehr häufig rechtssteil verteilt, wie leicht durch eine Analyse der freigegebenen Datensätze zu zeigen ist (s. Kapitel 3).<sup>2</sup>

Auf manifester Datenebene ergeben sich also kaum Unterschiede zwischen kriteriumsorientierten Tests und anderen in der Pädagogik eingesetzten Instrumenten. Differenzen lassen sich damit bestenfalls im Bereich der Konstruktion der Messinstrumente und der den Messinstrumenten zugrundeliegenden Konstrukte finden. Inwiefern etwa die bei kriteriumsorientierten Tests plausible Annahme rechtssteil verteilter Fähigkeitswerte auch auf andere Messinstrumente übertragbar ist, müsste gesondert diskutiert werden. Dessen ungeachtet lässt sich schon aufgrund der Ähnlichkeit von kriteriumsorientierten Testdaten und anderen in der Pädagogik üblichen Messdaten ein sekundäres und allgemeineres Ziel der Dissertation formulieren: Es soll überprüft werden, wie sensibel verschiedene Faktorenmodelle hinsichtlich ihrer Schätzgüte auf rechtssteil verteilte Testdatensätze reagieren.

---

<sup>2</sup>Dieser singuläre Befund konnte von Micceri (1989) allgemein abgesichert werden: Von 231 analysierten klassischen Leistungstests waren gut 68 Prozent der manifesten Daten rechtssteil und 32 Prozent linkssteil verteilt; eine Normalverteilung auf manifester Ebene konnte in keinem Fall nachgewiesen werden.



# Kapitel 2

## Theoretische Grundlagen

### 2.1 Begrifflicher Rahmen

Nach Lienert und Raatz (1998, S. 1) ist „ein Test [...] ein wissenschaftliches Routineverfahren zur Untersuchung eines oder mehrerer empirisch abgrenzbarer Persönlichkeitsmerkmale mit dem Ziel einer möglichst quantitativen Aussage über den relativen Grad der individuellen Merkmalsausprägung“. Bei kriteriumsorientierten Tests wird die individuelle Merkmalsausprägung in Beziehung gesetzt zu Idealwerten, die den Lehrzielen entsprechen (Klauer, 1987). Demnach ist ein lehrzielorientierter Test ein wissenschaftliches Routineverfahren zur Untersuchung von Persönlichkeitsmerkmalen mit dem Ziel, einer möglichst quantitativen Aussage darüber, ob ein Lehrziel erreicht wurde. Die Quantifizierung erfolgt hierbei über die kriteriumsorientierte Messung. Aber was bedeutet kriteriumsorientierte Messung?

Messen ist die Abbildung eines empirischen Relativs in ein numerisches Relativ, d.h. empirischen Sachverhalten werden Zahlen zugeordnet. Dies geschieht in der Pädagogik in der Regel dadurch, dass man den Antworten einer Person auf die Aufgaben eines Tests Zahlen zuordnet. Jeder Person werden zunächst die entsprechenden (Aufgaben-)Werte zugeordnet und dann beispielsweise durch Summierung dieser Werte ein Testwert. Diese Form des Messens ist allerdings nur sinnvoll, wenn bestimmte Regeln eingehalten werden, wobei ein Satz von diesen Regeln als Messmodell bzw. als Skala bezeichnet wird. Man unterscheidet dabei deterministische Messmodelle von stochastischen Messmodellen (Steyer & Eid, 1993).

Deterministische Messmodelle ordnen einer Antwortmöglichkeit auf eine Aufgabe immer genau eine Zahl zu. Bei kriteriumsorientierten Messungen kommt dabei nicht selten das nominale Messmodell bzw. die Ordinalskala zum Ein-

satz. Im nominalen Messmodell von kriteriumsorientierten Tests wird der Person in der Regel eine Eins zugeteilt, wenn sie die Aufgabe richtig bearbeitet hat und eine Null, wenn die Aufgabe falsch gelöst wurde. Eine ordinalskalierte Aufgabe liegt vor, wenn nicht nur unterschieden wird, ob die Personen die Aufgabe richtig oder falsch gelöst haben, sondern auch Teillösungen mit Punkten versehen werden. Personen mit einer genaueren Lösung der Aufgabe wird dabei üblicherweise eine höhere Zahl zugeordnet als Personen mit einer weniger genauen Lösung. Die nominal skalierten Aufgaben von kriteriumsorientierten Tests sind in der Regel also dichotom skaliert, während die ordinalskalierten Aufgaben in polytomer Form vorliegen.

Allerdings stellen die Messwerte bei kriteriumsorientierten Tests lediglich einen Indikator für das Persönlichkeitsmerkmal dar. Hohe Werte deuten in der Regel auf eine starke Ausprägung des Merkmals hin, niedrige Werte stehen für eine schwache Ausprägung. Zur genauen Quantifizierung der Merkmalsausprägung muss ein Messmodell erstellt werden. Da aber parallel zum Persönlichkeitsmerkmal auch andere Faktoren bestimmen können, ob eine Aufgabe richtig oder falsch gelöst wird, und weil das Persönlichkeitsmerkmal häufig als stetig konstruiert wird, bedarf es eines geeigneten stochastischen Messmodells, um eine Beziehung zwischen zugewiesenen Aufgabewerten und stetigen latenten Merkmal herzustellen. Wie noch zu zeigen ist, können die in dieser Dissertation betrachteten Faktorenanalysen als eben solche stochastischen Messmodelle aufgefasst werden.

Kriteriumsorientierte Messungen bestehen also aus zwei Messmodellen: Einem deterministischen und einem stochastischen Messmodell. Zu den Gütekriterien, die eine solche Messung erfüllen muss, zählen die Objektivität, die Reliabilität und die Validität (Lienert & Raatz, 1998). Die Objektivität des Verfahrens ist gegeben, wenn die Testergebnisse unabhängig von der Person sind, welche den Test durchführt. Um die Objektivität zu gewährleisten, werden sämtliche Verfahrensabläufe soweit wie möglich standardisiert.

Die Reliabilität eines Tests bezeichnet den Grad der Genauigkeit der Messung. Ein Verfahren ist also umso reliabler, je weniger das Testergebnis von Faktoren abhängt, die nicht zu dem Persönlichkeitsmerkmal zählen. Zur Bestimmung der Reliabilität existieren verschiedene praktisch-methodische Vorgehensweisen, auf die hier aber nicht weiter eingegangen werden muss (s. dazu Lienert & Raatz, 1998).

Ein Test ist schließlich valide, wenn der Test das, was er zu messen vorgibt, tatsächlich misst. Nach den Empfehlungen der American Psychological Association (1954, 1966) lassen sich die vier Validitätsarten prädiktive Validität, konkurrente Validität, Inhaltsvalidität und Konstruktvalidität unterscheiden. Prädiktive und konkurrente Validität werden häufig zur Kriteriumsvalidität

zusammengefasst. Gefordert wird, dass das Testergebnis mit dem Kriterium, welches vorhergesagt bzw. diagnostiziert werden soll, hoch korrelieren.

Inhaltsvalidität liegt gemäß Klauer (1987) genau dann vor, wenn der Test die Gesamtheit einer definierten Menge von Aufgaben enthält oder repräsentiert. Die Definition der Aufgabenmenge lehnt sich dabei an die Definition des Persönlichkeitsmerkmals an: Es sind jene Aufgaben Bestandteil der Menge, von denen aufgrund der Definition des Persönlichkeitsmerkmals angenommen werden kann, dass ihre Lösungshäufigkeit mit der Ausprägung der latenten Eigenschaft korreliert. Die Inhaltsvalidität eines Tests kann also nur theoretisch (z.B. durch die Beurteilung von Experten) bestimmt werden.

Von Konstruktvalidität kann ausgegangen werden, wenn die mit dem Test erfassten Verhaltensweisen jenen Verhaltensweisen entsprechen, die aufgrund der zu messenden Eigenschaft zu erwarten sind. Cronbach und Meehl (1955) unterscheiden sieben methodische Zugänge, um die Konstruktvalidität eines Tests zu bestimmen. Neben einer inhaltlich-logischen Analyse der Testelemente und einer statistisch geleiteten Interpretation von inter- sowie intraindividuellen Testergebnissen steht die Bestimmung der konvergenten und diskriminanten Validität (Campbell & Fiske, 1959) sowie die faktorielle Validität im Vordergrund.

Gemäß der konvergenten Validität müssen die Testergebnisse hoch positiv mit Messdaten von Tests korrelieren, die ein ähnliches Konstrukt erfassen sollen. Diskriminante Validität kann einem Test bescheinigt werden, wenn die Testergebnisse nur gering mit Messdaten von Tests korrelieren, die ein anderes Konstrukt erfassen sollen. Die faktorielle Validität bezieht sich schließlich zum einen auf die Beziehungen zwischen den beobachteten Testergebnissen und Messdaten aus konstruktnahen und konstruktfernen Tests und zum anderen auf die faktorielle Struktur des Tests (Lienert & Raatz, 1998).

Die faktorielle Struktur eines Tests wird durch die Faktorenanalyse bestimmt. Sie entspricht der Anzahl extrahierter Faktoren und deren korrelative Beziehungen untereinander sowie der Korrelation der einzelnen Testaufgaben mit den Faktoren. Die Anzahl extrahierter Faktoren sollte dabei der Dimensionalität des Konstrukts und die korrelativen Beziehungen zwischen den Faktoren der Komplexität des Konstrukts entsprechen, um von einer hohen faktoriellen Validität ausgehen zu können.

Allerdings gibt es eine Reihe unterschiedlicher Faktorenmodelle, die nicht zu gleichen Ergebnissen führen müssen. Gemäß der Latent Structure-Analyse<sup>1</sup>

---

<sup>1</sup>Nach der Latent Structure-Analyse lassen sich beobachtbare Variablen auf latente Merkmale zurückführen, wobei die beobachtbaren Variablen eine Funktion der latenten Variablen sind. Eine Variable ist eine Größe, die verschiedene Werte annehmen kann.

von Lazarsfeld (1950) lassen sich diese wie in Tabelle 2.1 abgebildet klassifizieren.<sup>2</sup> Wenn also beispielsweise die Aufgabenergebnisse und die latenten Merkmale stetig sind, resultieren verschiedene Varianten der klassischen Faktorenanalyse. Sind die Aufgabenergebnisse indes diskret und die latenten Variablen stetig skaliert, dann ergeben sich die faktorenanalytischen Modelle der Item-Response-Theorie (IRT).

Tabelle 2.1: Klassifikation von Modellen der Latent Structure-Analyse nach Lazarsfeld (1950).

Modell der Latent Structure-Analyse	Aufgabenergebnis	Latentes Merkmal
Latent Class-Analyse	diskret	diskret
Latent Profile-Analyse	stetig	diskret
Item-Response-Theorie	diskret	stetig
Klassische Faktorenanalyse	stetig	stetig

Wie aus den vorangegangenen Überlegungen deutlich geworden ist, wird in dem stochastischen Messmodell der kriteriumsorientierten Messung in der Regel eine Beziehung zwischen latenten stetigen Merkmalen und diskreten Aufgabenergebnissen hergestellt. Insofern handelt es sich dabei um Faktorenmodelle der Item-Response-Theorie. In dieser Dissertation werden dennoch ausschließlich die verschiedenen Modelle der klassischen Faktorenanalyse untersucht. Für die Einschränkung auf klassische Faktorenmodelle sprechen vor allem drei Gründe:

1. **Praktische Relevanz:** Zur Bestimmung der faktoriellen Validität von kriteriumsorientierten Tests bzw. zur Analyse von diskret skalierten Messinstrumenten werden häufig klassische Faktorenanalysen durchgeführt (z. B. Kölbl, Tiedemann & Billmann-Mahecha, 2006; M. Martin, Mullis & Kennedy, 2007; Neuenschwander & Kunz-Makarova, 2000; OECD, 2009a; Olson, Martin & Mullis, 2008; Sparfeldt, Rost & Schilling, 2004), d.h. die Ergebnisse dieser Dissertation enthalten Implikationen für die gängige Forschungspraxis.
2. **Ökonomische Einschränkung:** In gängigen statistischen Programmen wie etwa R (R Development Core Team, 2010), Mplus oder ConQuest

<sup>2</sup>In Abgrenzung zur Latent Structure-Analyse, welche eher die Unterschiede zwischen Modellen mit latenten Variablen betont, sind neuere Konzepte wie beispielsweise der *generalized latent variable* Ansatz zu sehen, welche eher die Gemeinsamkeiten zwischen den Modellen mit latenten Variablen herausarbeiten (Bartholomew, Knott & Moustaki, 2011; Moustaki & Knott, 2000; Skondal & Rabe-Hesketh, 2004).

(Wu, Adams & Wilson, 1997) werden im Rahmen von IRT-Analysen die latenten Fähigkeitswerte so geschätzt, dass sie einer Normalverteilung folgen, d.h. Simulationen, die auf rechtssteil verteilten Fähigkeitswerten aufbauen, sind nicht ohne weiteres mit diesen Programmen umsetzbar. Prinzipiell könnten zwar auch andere Verteilungsformen für die latenten Fähigkeitswerte angenommen werden, die dazugehörigen Algorithmen müssten aber erst programmiert werden, was in dafür angelegten Projekten erfolgen sollte.

3. **Strukturanalytische Herausforderungen:** Bei IRT-Modellen lässt sich aufgrund der im Strukturmodell mit aufgeführten Aufgabenparameter nicht unbedingt ein direkter funktionaler Zusammenhang zwischen manifesten Verteilungen und Verteilung der latenten Fähigkeitswerte bestimmen, d.h. eine eventuell vorhandene Schiefe der Testwerte kann durch entsprechende Kalibrierung der Aufgabenparameter auch aufgefangen werden. Inwiefern ein solches Vorgehen sinnvoll ist, müsste zunächst in der Fachöffentlichkeit diskutiert werden.

Wenngleich die Einschränkung auf klassische Faktorenmodelle begründbar ist, muss bei der Interpretation der Ergebnisse berücksichtigt werden, dass sich diese zunächst einmal (nur) auf die Güte verschiedener klassischer Faktorenmodelle beziehen können. Inwiefern darüber hinaus Aussagen über IRT-Modelle möglich sind, ist im Einzelfall zu überprüfen. Auch ist daran zu denken, dass die Simulationen aufgrund der Einschränkung auf klassische Faktorenmodelle (notwendigerweise) auf stetig skalierte manifeste Daten aufbauen, während empirische Forschungsdaten häufig diskret sind.

Wie die Ausführungen zum empirischen Forschungsstand (s. Kapitel 2.7) zeigen werden, impliziert die Verwendung von stetigem Datenmaterial eine Überschätzung der Schätzgüte klassischer Faktorenmodelle, zumindest dann, wenn die Schätzgüte in Bezug auf die übliche Anwendung in der Forschung abgeschätzt werden soll. Oder anders ausgedrückt: Es ist davon auszugehen, dass die hier getroffenen Aussagen bezüglich der Schätzgüte von klassischen Faktorenmodellen auch dann gelten, wenn anstelle von stetigem Datenmaterial diskrete Daten analysiert werden würden. Genaugenommen ist anzunehmen, dass sich die Schätzgüte bei Verwendung von diskretem Datenmaterial im Allgemeinen eher reduzieren würde.

Da der zentrale Gegenstand der Dissertation die Schätzgüte klassischer Faktorenmodelle in Bezug auf die faktorielle Struktur bei rechtssteil verteilten Daten bzw. rechts steil verteilten Fähigkeitswerten ist, werden im Nachfolgenden die untersuchten klassischen Faktorenmodelle inklusive deren Voraussetzungen und zugrundeliegenden Schätzalgorithmen vorgestellt. Danach

wird auf das Rotationsproblem eingegangen und es werden die Kriterien zur Bestimmung der faktoriellen Struktur erläutert. Schließlich wird in einem gesonderten Kapitel auf die Verteilungseigenschaften der Testwerte bzw. Fähigkeitswerte bei klassischen Faktorenmodellen eingegangen und es wird der bisherige empirische Forschungsstand zur Schätzgüte der klassischen Faktorenmodelle dargestellt.

Es sei allerdings vorweggestellt, dass mit den Ausführungen zu den Schätzeigenschaften der verschiedenen Modelle keine erschöpfende Darstellung vorhandener Schätzmethoden angestrebt wird. Dies würde angesichts der kaum mehr überschaubaren Anzahl von Schätzalgorithmen im Bereich der Eigenwertbestimmung (s. Watkins, 2007) und der numerischen Verfahren zur Nullstellenbestimmung (s. Fletcher, 2006) den Rahmen dieser Dissertation deutlich überschreiten. Dies gilt umso mehr, als dass gerade die Entwicklungen im Bereich der Eigenwertbestimmung aufgrund der stetigen Leistungssteigerung von Computern rasant voranschreiten.

Anstelle einer erschöpfenden Darstellung von Schätzalgorithmen sollen deswegen jene Schätzverfahren vorgestellt werden, die entweder von historisch herausragender Bedeutung sind (sofern sie als Grundlage dienen können, um die derzeit verwendeten Schätzverfahren verständlich darzustellen), oder die in gängigen Statistikprogrammen, wie beispielsweise R (R Development Core Team, 2010) bzw. in allgemein verbreiteten Programmbibliotheken wie etwa LAPACK (Anderson et al., 1999), installiert sind.

## 2.2 Symbolische Zuordnungen

Bevor verschiedene Modelle der Faktorenanalyse vorgestellt werden können, sind einige symbolische Zuordnungen zu treffen. Im Folgenden soll deswegen  $y_{ij}$  für das Ergebnis stehen, dass Person  $i$  ( $i = 1, \dots, n$ ) in Aufgabe  $j$  ( $j = 1, \dots, p$ ) erzielt hat. Bei kriteriumsorientierten Messungen handelt es sich dabei häufig um eine Eins, wenn die Person die Aufgabe richtig bearbeitet hat und eine Null, wenn die Aufgabe falsch gelöst wurde.

Neben den Aufgabenergebnissen  $y_{ij}$  werden auch die standardisierten Werte  $z_{ij}$  benötigt, wobei gilt:

$$z_{ij} = \frac{y_{ij} - \bar{y}_j}{\sqrt{n-1}s_j} \quad \bar{y}_j = \frac{\sum_{i=1}^n y_{ij}}{n} \quad s_j^2 = \frac{\sum_{i=1}^n (y_{ij} - \bar{y}_j)^2}{n-1}$$

Ein positiver Wert von  $z_{ij}$  bedeutet, dass die Person  $i$  ein überdurchschnittliches Ergebnis in Aufgabe  $j$  erzielt hat, und ein negativer Wert von  $z_{ij}$  weist darauf, dass das Ergebnis unter dem Durchschnitt liegt. Wenn darüber

hinaus die Verteilung von  $y_{ij}$  ( $i = 1, \dots, \infty$ ) bekannt ist (z.B. die Normalverteilung), dann kann mit Hilfe der  $z_{ij}$ -Werte bestimmt werden, mit welcher Wahrscheinlichkeit ein Aufgabenergebnis  $y_{ij}$  erzielt wird.

Die standardisierten Aufgabenergebnisse  $z_{ij}$  können zur Matrix  $\mathbf{Z}$  zusammengefasst werden

$$\mathbf{Z} = \begin{pmatrix} z_{11} & z_{12} & \cdots & z_{1p} \\ z_{21} & z_{22} & \cdots & z_{2p} \\ \vdots & \vdots & \cdots & \vdots \\ z_{n1} & z_{n2} & \cdots & z_{np} \end{pmatrix}$$

so dass

$$\begin{aligned} \mathbf{Z}'\mathbf{Z} &= \begin{pmatrix} z_{11} & z_{21} & \cdots & z_{n1} \\ z_{12} & z_{22} & \cdots & z_{n2} \\ \vdots & \vdots & \cdots & \vdots \\ z_{1p} & z_{2p} & \cdots & z_{np} \end{pmatrix} \begin{pmatrix} z_{11} & z_{12} & \cdots & z_{1p} \\ z_{21} & z_{22} & \cdots & z_{2p} \\ \vdots & \vdots & \cdots & \vdots \\ z_{n1} & z_{n2} & \cdots & z_{np} \end{pmatrix} \\ &= \begin{pmatrix} \sum_{i=1}^n z_{i1}^2 & \sum_{i=1}^n z_{i1}z_{i2} & \cdots & \sum_{i=1}^n z_{i1}z_{ip} \\ \sum_{i=1}^n z_{i2}z_{i1} & \sum_{i=1}^n z_{i2}^2 & \cdots & \sum_{i=1}^n z_{i2}z_{ip} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^n z_{ip}z_{i1} & \sum_{i=1}^n z_{ip}z_{i2} & \cdots & \sum_{i=1}^n z_{ip}^2 \end{pmatrix} \end{aligned}$$

gilt. Die einzelnen Elemente der Matrix  $\mathbf{Z}'\mathbf{Z}$  stellen dabei die empirischen Korrelationskoeffizienten  $r_{vj}$  zwischen den Variablen dar, denn

$$\begin{aligned} \sum_{i=1}^n z_{iv}z_{ij} &= z_{1v}z_{1j} + \cdots + z_{nv}z_{nj} \\ &= \frac{(y_{1v} - \bar{y}_v)(y_{1j} - \bar{y}_j)}{\sqrt{n-1}s_v\sqrt{n-1}s_j} + \cdots + \frac{(y_{nv} - \bar{y}_v)(y_{nj} - \bar{y}_j)}{\sqrt{n-1}s_v\sqrt{n-1}s_j} \\ &= \frac{\sum_{i=1}^n (y_{iv} - \bar{y}_v)(y_{ij} - \bar{y}_j)}{\sqrt{n-1}s_v\sqrt{n-1}s_j} \\ &= \frac{\sum_{i=1}^n (y_{iv} - \bar{y}_v)(y_{ij} - \bar{y}_j)}{\sqrt{\sum_{i=1}^n (y_{iv} - \bar{y}_v)^2} \sqrt{\sum_{i=1}^n (y_{ij} - \bar{y}_j)^2}} \\ &= \frac{\sum_{i=1}^n (y_{iv} - \bar{y}_v)(y_{ij} - \bar{y}_j)}{\sqrt{\sum_{i=1}^n (y_{iv} - \bar{y}_v)^2} \sum_{i=1}^n (y_{ij} - \bar{y}_j)^2} \\ &= r_{vj} \end{aligned}$$

für alle  $v, j = 1, \dots, p$ . Es gilt also  $\mathbf{Z}'\mathbf{Z} = \mathbf{R}$  mit

$$\mathbf{R} = \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1p} \\ r_{21} & r_{22} & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \cdots & r_{pp} \end{pmatrix}$$

und wie noch gezeigt wird, bildet die Matrix  $\mathbf{R}$  häufig die Basis für die Durchführung einer Faktorenanalyse.<sup>3</sup>

## 2.3 Modelle der klassischen Faktorenanalyse

### 2.3.1 Überblick

Zu den Verfahren der klassischen Faktorenanalyse gehören die Hauptkomponentenmethode (Pearson, 1901; Hotelling, 1933a, 1933b; Kelley, 1935), die explorative Faktorenanalyse (Spearman, 1904; Burt, 1909; Thurstone, 1931, 1965) und die Hauptfaktorenanalyse (Thurstone, 1931, 1965). Für jedes dieser Verfahren wird deswegen das Strukturmodell, die Voraussetzungen und die Parameterschätzung dargestellt.

### 2.3.2 Hauptkomponentenmethode

#### 2.3.2.1 Strukturmodell

Das Strukturmodell der Hauptkomponentenmethode lautet

$$z_{ij} = f_{i1}l_{j1} + f_{i2}l_{j2} + \cdots + f_{ip}l_{jp}$$

d.h. es wird davon ausgegangen, dass sich das Ergebnis der Person  $i$  bei Aufgabe  $j$  in die Hauptkomponenten  $f_{ik}$  ( $k = 1, \dots, p$ ) zerlegen lässt, wobei diese mit  $l_{jk}$  zu gewichten sind.

In Matrixschreibweise gilt für das Hauptkomponenten-Modell:

$$\begin{pmatrix} z_{11} & z_{12} & \cdots & z_{1p} \\ z_{21} & z_{22} & \cdots & z_{2p} \\ \vdots & \vdots & \cdots & \vdots \\ z_{n1} & z_{n2} & \cdots & z_{np} \end{pmatrix} = \begin{pmatrix} f_{11} & \cdots & f_{1p} \\ \vdots & & \vdots \\ f_{n1} & \cdots & f_{np} \end{pmatrix} \begin{pmatrix} l_{11} & \cdots & l_{p1} \\ \vdots & & \vdots \\ l_{1p} & \cdots & l_{pp} \end{pmatrix}$$

<sup>3</sup>In den nachfolgenden Kapiteln wird bei den algebraischen Herleitungen die Symbolisierung  $\mathbf{Z}'\mathbf{Z}$  verwendet, um die Umformungen transparent zu gestalten. Die konkreten Beispiele hingegen basieren ebenso wie die faktischen Berechnungen auf der Matrix  $\mathbf{R}$ .



$$\mathbf{Z} = \mathbf{F}\mathbf{L}' \quad (2.1)$$

mit  $\mathbf{Z}$   $n \times p$ -Matrix der Aufgabenergebnisse  $z_{ij}$   
 $\mathbf{F}$   $n \times p$ -Matrix der Hauptkomponenten  $f_{ik}$   
 $\mathbf{L}$   $p \times p$ -Matrix der Gewichte  $l_{jk}$

### 2.3.2.2 Voraussetzungen

Als rein datenmanipulierendes Verfahren ist die Hauptkomponentenmethode an keine Verteilungsannahmen gebunden. Es muss jedoch vorausgesetzt werden, dass:

1. Die empirischen Momente von  $\mathbf{Z}$  existieren.
2.  $\forall_j s_j^2 \neq 0$  ( $j = 1, \dots, p$ ).
3.  $\text{rg}(\mathbf{Z}) = \text{rg}(\mathbf{R}) = p$ .

Die letzte Annahme ist wenig restriktiv: Wenn der Rang der Matrix  $\mathbf{Z}$  bzw.  $\mathbf{R}$  nicht gleich  $p$  ist, dann sind einige Variablen als lineare Kombination anderer Variablen darstellbar; diese Variablen sind also redundant und können ohne Beschränkung der Allgemeingültigkeit von den Analysen ausgeschlossen werden.

Die zweite Annahme ist schon kritischer: Wenn Variablen keine Varianz haben, ist die Korrelation nicht definiert, d.h. die Matrix  $\mathbf{R}$  kann nicht bestimmt werden. Bei kriteriumsorientierten Tests gilt häufig  $\forall_j s_j^2 \approx 0$ , weil die meisten Personen das Lehrziel erreichen (Klauer, 1987). Dies führt in der Praxis nicht selten dazu, dass die Korrelationskoeffizienten nahe bei Null liegen, also  $\mathbf{R} \approx \mathbf{I}$  gilt (Sturzbecher et al., 2008).

Im beschriebenen Fall wird nach den gültigen Kriterien für die Extraktion von Hauptkomponenten aber lediglich eine Komponente extrahiert, wie anhand der nachfolgenden Erläuterungen zur Schätzmethode und faktoriellen Struktur deutlich wird. Wenngleich lehrzielorientierte Tests also inhaltlich verschiedene Skalen aufweisen können, läßt sich empirisch meist nur die Bildung eines einzigen Testgesamtwert rechtfertigen, d.h. empirisch lassen sich die Skalen eines kriteriumsorientierten Tests meistens nicht bestätigen.

Die erste Annahme ist essenziell für die Hauptkomponentenmethode: Wenn die empirischen Momente wie etwa  $s_j$  nicht existieren, dann können sie auch

nicht ermittelt werden, was insofern ein Problem darstellt, als dass die Schätzmethoden klassischer Faktorenmodelle die ermittelten empirischen Momente verwenden (s. die nachfolgenden Ausführungen zu den Schätzmethoden klassischer Faktorenmodelle). Im Kontext der Hauptkomponentenmethode wird in der Regel auch davon ausgegangen, dass  $\mathbf{Z}$ ,  $\mathbf{F}$  und  $\mathbf{L}'$  intervallskaliert mit  $\mathbf{Z} \in \mathbb{R}^{n \times p}$ ,  $\mathbf{F} \in \mathbb{R}^{n \times p}$  und  $\mathbf{L}' \in \mathbb{R}^{p \times p}$  sind. Sofern nicht abweichend notiert, wird bei den weiteren Ausführungen davon ausgegangen, dass die Voraussetzungen der Hauptkomponentenmethode erfüllt sind.

### 2.3.2.3 Parameterschätzung

#### Ausgangspunkt

In Gleichung 2.1 sind die Koeffizienten der Matrix  $\mathbf{Z}$  bekannt, die Matrizen  $\mathbf{F}$  und  $\mathbf{L}$  müssen hingegen geschätzt werden. Ausgangspunkt dieser Schätzungen sind die gewichteten Testgesamtwerte  $g_{im}$  ( $m = 1, \dots, k$  und  $k \in j$ ; zur Bestimmung von  $k$  siehe Kapitel 2.5) für die gilt

$$\begin{aligned} g_{im} &= \sum_{j=1}^p c_{jm} z_{ij} \\ &= \begin{pmatrix} z_{i1} & z_{i2} & \cdots & z_{ip} \end{pmatrix} \begin{pmatrix} c_{1m} \\ c_{2m} \\ \vdots \\ c_{pm} \end{pmatrix} \\ &= \mathbf{z}_i \cdot \mathbf{c}_m \end{aligned}$$

wobei die Summe der quadrierten Gewichte eins sein soll, d.h.  $\sum_{j=1}^p c_{jm}^2 = 1$ .<sup>4</sup> Da theoretisch unendlich viele Gewichtsvektoren  $\mathbf{c}_m$  existieren die diese Bedingung erfüllen, müssen weitere Forderungen formuliert werden, um eine Auswahl geeigneter Vektoren zu treffen. Bei der Hauptkomponentenmethode wird gefordert, dass die ersten Testgesamtwerte  $g_{i1} = \sum_{j=1}^p c_{j1} z_{ij}$  maximal

---

<sup>4</sup>Durch die Normierung werden Vielfache von  $\mathbf{c}_m$  nicht als Gewichtsvektoren akzeptiert. Dies scheint sinnvoll, da ein Vielfaches von  $\mathbf{c}_m$  zu einem Vielfachen der Testgesamtwerte und damit zu einer *künstlichen* Erhöhung der Testvarianz führen. Sei beispielsweise  $\mathbf{c}_m$  ein geeigneter Vektor, dann gilt  $g_{im} = \mathbf{z}_i \cdot \mathbf{c}_m$ . Würde auch  $2\mathbf{c}_m = \mathbf{c}_m^*$  als Gewichtsvektor akzeptiert, dann wäre  $g_{im}^* = \mathbf{z}_i \cdot \mathbf{c}_m^* = 2\mathbf{z}_i \cdot \mathbf{c}_m = 2g_{im}$ . Und wenn die Varianz von  $g_{im}$  mit  $\sigma_{g_{im}}^2$  bezeichnet wird, dann gilt für die Varianz von  $g_{im}^*$ :  $\sigma_{g_{im}^*}^2 = 4\sigma_{g_{im}}^2$ , d.h. die Varianz der Testgesamtwerte  $\sigma_{g_{im}^*}^2$  würde bei Verwendung des Gewichtsvektors  $\mathbf{c}_m^*$  um den Faktor 4 überschätzt werden.

variieren, d.h. es soll

$$\begin{aligned} \max (\mathbf{g}'_1 \mathbf{g}_1) &= \max \left( \begin{pmatrix} g_{11} & g_{21} & \cdots & g_{n1} \end{pmatrix} \begin{pmatrix} g_{11} \\ g_{21} \\ \vdots \\ g_{n1} \end{pmatrix} \right) \\ &= \max (\mathbf{c}'_1 \mathbf{Z}' \mathbf{Z} \mathbf{c}_1) \end{aligned}$$

gelten. Der zweite Gewichtsvektor ist so zu wählen, dass die zweiten Testgesamtwerte  $g_{i2} = \sum_{j=1}^p c_{j2} z_{ij}$  sowohl unabhängig sind von den ersten Testgesamtwerten  $g_{i1}$  als auch maximal variieren. Der Prozess wird schließlich beendet, wenn weitere Testgesamtwerte  $g_{il}$  ( $l = (k+1), \dots, p$ ) eine bestimmte minimale Varianz nicht mehr überschreiten (s.u.).

Die Funktionen  $\mathbf{f}(\phi)_j = \mathbf{g}'_j \mathbf{g}_j$  müssen also unter Berücksichtigung der jeweiligen Nebenbedingungen maximiert werden. Wie aus der Analysis bekannt, ist beim Maximum einer Funktion die erste partielle Ableitung Null. Also werden die ersten partiellen Ableitungen der Funktionen  $\mathbf{f}(\phi)_j$  gebildet und nach den gesuchten Vektoren  $\mathbf{c}_j$  aufgelöst. Das Vorgehen soll am Beispiel der Funktionen  $\mathbf{f}(\phi)_1 = \mathbf{g}'_1 \mathbf{g}_1$  und  $\mathbf{f}(\phi)_2 = \mathbf{g}'_2 \mathbf{g}_2$  verdeutlicht werden.<sup>5</sup>

Die Funktion  $\mathbf{f}(\phi)_1 = \mathbf{g}'_1 \mathbf{g}_1$  muss unter der Nebenbedingung  $\mathbf{c}'_1 \mathbf{c}_1 = 1$  maximiert werden. Also wird die erste partielle Ableitung von  $\mathbf{g}'_1 \mathbf{g}_1$  nach  $\mathbf{c}_1$  unter der Nebenbedingung  $\mathbf{c}'_1 \mathbf{c}_1 = 1$  gebildet

$$\begin{aligned} \frac{\partial \mathbf{g}'_1 \mathbf{g}_1 - \lambda_1 (\mathbf{c}'_1 \mathbf{c}_1 - 1)}{\partial \mathbf{c}_1} &= \frac{\partial \mathbf{c}'_1 \mathbf{Z}' \mathbf{Z} \mathbf{c}_1 - \lambda_1 (\mathbf{c}'_1 \mathbf{c}_1 - 1)}{\partial \mathbf{c}_1} \\ &= 2\mathbf{Z}' \mathbf{Z} \mathbf{c}_1 - 2\lambda_1 \mathbf{c}_1 \end{aligned}$$

und Null gesetzt:

$$\begin{aligned} 2\mathbf{Z}' \mathbf{Z} \mathbf{c}_1 - 2\lambda_1 \mathbf{c}_1 &= \mathbf{0} \\ (\mathbf{Z}' \mathbf{Z} - \lambda_1 \mathbf{I}) \mathbf{c}_1 &= \mathbf{0} \end{aligned} \tag{2.2}$$

Gleichung 2.2 wird als Eigenwertproblem bezeichnet (Werner, 1997): Es wird ein von Null verschiedener Vektor für  $\mathbf{c}_1$  gesucht, der nur gefunden werden kann, wenn die Determinante von  $(\mathbf{Z}' \mathbf{Z} - \lambda_1 \mathbf{I})$  Null ist was bedeutet, dass für  $\lambda_1$  einer der  $p$  Eigenwerte der Matrix  $\mathbf{Z}' \mathbf{Z}$  zu wählen ist.<sup>6</sup> Allerdings kann

<sup>5</sup>Im Folgenden wird davon ausgegangen, dass der Leser mit den Grundlagen der mehrdimensionalen Differenzialrechnung vertraut ist. Eine Einführung in das Thema sowie zentrale Theoreme finden sich beispielsweise bei Magnus und Neudecker (1999).

<sup>6</sup>Natürlich würde auch der Vektor  $\mathbf{c}_1 = \mathbf{0}$  die Gleichung 2.2 lösen, aber  $\mathbf{c}_1 = \mathbf{0}$  ist trivial, weil in diesem Fall die Testgesamtwerte null wären.

für  $\lambda_1$  kein beliebiger Eigenwert von  $\mathbf{Z}'\mathbf{Z}$  ausgewählt werden, denn

$$\begin{array}{l} \mathbf{Z}'\mathbf{Z}\mathbf{c}_1 = \lambda_1\mathbf{c}_1 \\ \mathbf{c}'_1\mathbf{Z}'\mathbf{Z}\mathbf{c}_1 = \lambda_1\mathbf{c}'_1\mathbf{c}_1 \\ \mathbf{c}'_1\mathbf{Z}'\mathbf{Z}\mathbf{c}_1 = \lambda_1 \\ \mathbf{g}'_1\mathbf{g}_1 = \lambda_1 \end{array} \quad \left| \begin{array}{l} \mathbf{c}'_1 \\ \mathbf{c}'_1\mathbf{c}_1 = 1 \end{array} \right.$$

und da  $\mathbf{g}'_1\mathbf{g}_1$  maximiert werden soll, kann  $\lambda_1$  nur der größte Eigenwert der Matrix  $\mathbf{Z}'\mathbf{Z}$  und  $\mathbf{c}_1$  der dazugehörige Eigenvektor sein.

Wird als erster Gewichtsvektor der Eigenvektor zum größten Eigenwert der Matrix  $\mathbf{Z}'\mathbf{Z}$  gewählt, dann weisen die ersten Testgesamtwerte  $g_{i1} = \sum_{j=1}^p c_{j1}z_{ij}$  maximale Varianz auf. Der zweite Gewichtsvektor ist so zu wählen, dass die zweiten resultierenden Testgesamtwerte  $g_{i2} = \sum_{j=1}^p c_{j2}z_{ij}$  zum einen unabhängig von den ersten ermittelten Testgesamtwerten  $g_{i1}$  sind und zum anderen wiederum maximal variieren, d.h. es soll  $\mathbf{g}'_2\mathbf{g}_2 = \max$  unter den Nebenbedingungen  $\mathbf{g}'_2\mathbf{g}_1 = 0$  und  $\mathbf{c}'_2\mathbf{c}_2 = 1$  gelten. Die erste partielle Ableitung der Funktion  $f(\phi)_2 = \mathbf{g}'_2\mathbf{g}_2$  nach  $\mathbf{c}_2$  lautet folglich:

$$\begin{aligned} \frac{\partial \mathbf{g}'_2\mathbf{g}_2 - \lambda_2(\mathbf{c}'_2\mathbf{c}_2 - 1) - \mu\mathbf{g}'_2\mathbf{g}_1}{\partial \mathbf{c}_2} &= \frac{\partial \mathbf{c}'_2\mathbf{Z}'\mathbf{Z}\mathbf{c}_2 - \lambda_2(\mathbf{c}'_2\mathbf{c}_2 - 1)}{\partial \mathbf{c}_2} \\ &\quad - \frac{\mu\mathbf{c}'_2\mathbf{Z}'\mathbf{Z}\mathbf{c}_1}{\partial \mathbf{c}_2} \\ &= 2\mathbf{Z}'\mathbf{Z}\mathbf{c}_2 - 2\lambda_2\mathbf{c}_2 - \mu\mathbf{Z}'\mathbf{Z}\mathbf{c}_1 \end{aligned}$$

Mit  $\mu = 0$  liefert jede Lösung des Eigenwertproblems

$$(\mathbf{Z}'\mathbf{Z} - \lambda_2\mathbf{I})\mathbf{c}_2 = \mathbf{0} \tag{2.3}$$

einen Vektor für  $\mathbf{c}_2$ , und da  $\mathbf{g}'_2\mathbf{g}_2 = \lambda_2$  zu maximieren ist, ist für  $\lambda_2$  der zweitgrößte Eigenwert von  $\mathbf{Z}'\mathbf{Z}$  zu wählen und  $\mathbf{c}_2$  ist der dazugehörige Eigenvektor.

Die weiteren Eigenvektoren sind in gleicher Weise zu bestimmen, so daß schließlich gilt

$$\begin{array}{l} (\mathbf{g}_1 \ \mathbf{g}_2 \ \cdots \ \mathbf{g}_p) = \mathbf{Z} (\mathbf{c}_1 \ \mathbf{c}_2 \ \cdots \ \mathbf{c}_p) \\ \mathbf{G} = \mathbf{Z}\mathbf{C} \\ \mathbf{G}(\mathbf{C}'\mathbf{C}) = \mathbf{Z}\mathbf{C} \quad \quad \quad |\mathbf{C}'\mathbf{C} = \mathbf{I} \\ (\mathbf{G}\mathbf{C}')\mathbf{C} = \mathbf{Z}\mathbf{C} \end{array}$$

und damit  $\mathbf{Z} = \mathbf{G}\mathbf{C}'$ .

Im letzten Schritt der Hauptkomponentenmethode werden die Testgesamtwerte noch mit  $\Lambda^{-1/2}$  normiert, wobei  $\Lambda$  eine Diagonalmatrix ist mit den  $p$

Eigenwerten der Matrix  $\mathbf{R}$  in der Diagonalen. Die normierten Testgesamtwerte  $\mathbf{F} = \mathbf{G}\mathbf{\Lambda}^{-1/2}$  heißen schließlich Hauptkomponenten und es gilt

$$\mathbf{F} = \mathbf{G}\mathbf{\Lambda}^{-1/2} \text{ bzw. } \mathbf{L} = \mathbf{C}\mathbf{\Lambda}^{1/2} \text{ mit } \mathbf{F}'\mathbf{F} = \mathbf{I} \quad (2.4)$$

wobei  $\mathbf{L}'$  als Ladungsmatrix bezeichnet wird in der die Korrelationen zwischen den Hauptkomponenten und den Variablen stehen denn  $\mathbf{F}'\mathbf{Z} = \mathbf{F}'\mathbf{F}\mathbf{L}' = \mathbf{L}'$ .

Bisher ist offen geblieben, wie die Eigenwerte  $\lambda_j$  und Eigenvektoren  $\mathbf{c}_j$  von  $\mathbf{Z}'\mathbf{Z} = \mathbf{R}$  bestimmt werden. Deswegen wird hierauf im Nachfolgenden eingegangen. Da sich die Auswahl des Verfahrens zur Bestimmung der Eigenwerte und Eigenvektoren von  $\mathbf{R}$  an der Größe dieser Matrize orientiert, wurden die Ausführungen in drei Unterabschnitte gegliedert: Schätzung von  $\lambda_j$  und  $\mathbf{c}_j$  bei kleinen Matrizen, Schätzung von  $\lambda_j$  und  $\mathbf{c}_j$  bei mittleren Matrizen und Schätzung von  $\lambda_j$  und  $\mathbf{c}_j$  bei großen Matrizen.

### Schätzung von $\lambda$ und $\mathbf{c}$ bei kleinen Matrizen

Bei kleinen Matrizen ( $p \leq 4$ ) können die gesuchten Eigenwerte durch die Lösung des charakteristischen Polynoms

$$a_p\lambda^p + a_{p-1}\lambda^{p-1} + \dots + a_1\lambda^1 + a_0 = 0 \quad (2.5)$$

gefunden werden.<sup>7</sup> Das Polynom entspricht dem Determinanten-Ausdruck  $|\mathbf{Z}'\mathbf{Z} - \lambda_j\mathbf{I}|$  und hat  $p$  Lösungen, nämlich die Eigenwerte  $\lambda_j$ . Zu jedem Eigenwert lassen sich über die Gleichung

$$\mathbf{Z}'\mathbf{Z}\mathbf{c}_j = \lambda_j\mathbf{c}_j \quad (2.6)$$

die dazugehörigen Eigenvektoren  $\mathbf{c}_j$  finden.

**Beispiel 2.3.1:** Die Matrix  $\mathbf{R}$  habe folgende Werte

$$\mathbf{R} = \begin{pmatrix} 1 & 0.4 \\ 0.4 & 1 \end{pmatrix}$$

---

<sup>7</sup>Die Matrix  $\mathbf{R}$  ist quadratisch und, wenn die Voraussetzungen der Hauptkomponentenmethode erfüllt sind, regulär. Bei quadratisch regulären Matrizen entspricht die Anzahl der Eigenwerte der Matrize der Ordnung dieser Matrize (Werner, 1997), d.h.  $p$  steht hier sowohl für die Ordnung der Matrix  $\mathbf{R}$ , also die Spaltenanzahl von  $\mathbf{R}$ , als auch für die Anzahl der Eigenwerte von  $\mathbf{R}$ .

dann lautet die Determinante bzw. das charakteristische Polynom des Ausdrucks  $|\mathbf{R} - \lambda_j \mathbf{I}|$ :

$$\begin{aligned} \left| \begin{pmatrix} 1 - \lambda_j & 0.4 \\ 0.4 & 1 - \lambda_j \end{pmatrix} \right| &= (1 - \lambda_j)^2 - 0.16 = 0 \\ (1 - \lambda_j)^2 &= 0.16 \\ \lambda_{1,2} &= \pm 0.4 + 1 \\ \lambda_1 &= 1.4 \\ \lambda_2 &= 0.6 \end{aligned}$$

Somit sind  $\lambda_1 = 1.4$  und  $\lambda_2 = 0.6$  die Eigenwerte der Matrix  $\mathbf{R}$ . Um die dazugehörigen Eigenvektoren zu finden, wird zunächst die Gleichung 2.6 umformuliert:

$$\begin{aligned} (r_{11} - \lambda_1)c_{11} + r_{12}c_{21} &= 0 \\ r_{21}c_{11} + (r_{22} - \lambda_1)c_{21} &= 0 \end{aligned}$$

Nach einsetzen der Koeffizienten von  $\mathbf{R}$  und  $\lambda_1 = 1.4$  in dieses Gleichungssystem gilt

$$\begin{aligned} 0.4(c_{21} - c_{11}) &= 0 \\ 0.4(c_{11} - c_{21}) &= 0 \end{aligned}$$

so dass unter Berücksichtigung von  $\mathbf{c}'_1 \mathbf{c}_1 = 1$  der ersten Eigenvektor von  $\mathbf{R}$

$$\mathbf{c}_1 = \begin{pmatrix} \sqrt{0.5} \\ \sqrt{0.5} \end{pmatrix}$$

lautet. Für den zweiten Eigenwert  $\lambda_2 = 0.6$  gilt entsprechend

$$\begin{aligned} 0.4(c_{11} + c_{21}) &= 0 \\ 0.4(c_{11} + c_{21}) &= 0 \end{aligned}$$

d.h.

$$\mathbf{c}_2 = \begin{pmatrix} \sqrt{0.5} \\ -\sqrt{0.5} \end{pmatrix}$$

ist der zweite Eigenvektor von  $\mathbf{R}$ .

◁

### Schätzung von $\lambda$ und $\mathbf{c}$ bei mittleren Matrizen

Wie Galois (in Hadlock, 1978) zeigen konnte gibt es für das charakteristische Polynom in Gl. 2.5 nur bis zu einer Ordnung von  $p \leq 4$  explizite Lösungen. Bei größeren Matrizen werden deswegen iterative Algorithmen aus der numerischen linearen Algebra zur Schätzung von  $\lambda$  und  $\mathbf{c}$  angewandt. Für mittlere Matrizen ( $p \leq 25$ ) wird dabei in der Regel auf den *QR*-Algorithmus (Francis, 1961, 1962; Kublanovskaya, 1962) zurückgegriffen, bei großen Matrizen ( $p > 25$ ) kommt hingegen eine Kombination aus dem *divide and conquer* Verfahren (Cuppen, 1981) und dem *QR*-Algorithmus zum Einsatz (einen Überblick über diese und verschiedene andere Verfahren geben Golub & Vorst, 2000).

Ausgangspunkt der *QR*-Iteration ist die Potenzmethode nach von Mises und Pollaczek-Geiring (1929). Demnach konvergieren Folgen der Form

$$\mathbf{x}, \mathbf{A}\mathbf{s}_1, \mathbf{A}^2\mathbf{s}_2, \mathbf{A}^3\mathbf{s}_3, \dots \quad | \quad \mathbf{s}_w = \frac{\mathbf{A}^i \mathbf{s}_{w-1}}{\|\mathbf{A}^i \mathbf{s}_{w-1}\|} \quad (2.7)$$

für nahezu jeden beliebigen Vektor  $\mathbf{x}$  und  $\mathbf{A} \in \mathbb{C}^{n \times n}$  sowie  $\mathbf{s}_0 = \mathbf{x}$ ,  $w = 1 \rightarrow \infty$ ,  $i = 0 \rightarrow \infty$  zum dominanten Eigenvektor  $\mathbf{c}_1$  von  $\mathbf{A}$  falls es einen solchen dominanten Eigenvektor gibt. Als dominanter Eigenvektor wird dabei der Eigenvektor zum betragsmäßig größten Eigenwert  $\lambda_1$  von  $\mathbf{A}$  bezeichnet, wenn für die Eigenwerte von  $\mathbf{A}$  die Ungleichung  $|\lambda_1| > |\lambda_2| \geq \dots \geq |\lambda_n|$  gilt. Mit Hilfe der Potenzmethode kann also der Eigenvektor  $\mathbf{c}_1$  zum betragsmäßig größten Eigenwert von  $\mathbf{A}$  bestimmt werden und der betragsmäßig größte Eigenwert von  $\mathbf{A}$  rekursiv über die Formel  $\mathbf{A}\mathbf{c}_1 = \lambda_1\mathbf{c}_1$ .

In der Praxis hat sich gezeigt, dass die Konvergenzrate (d.h. die Anzahl notwendiger Iterationsschritte bis zur Konvergenz der Potenzmethode bzw. der nachfolgend angeführten *QR*-Iteration) in der Regel kleiner ausfällt, wenn anstelle von  $\mathbf{A}$  die Matrix  $p(\mathbf{A}) = (\mathbf{A} - q_b\mathbf{I})(\mathbf{A} - q_{b-1}\mathbf{I}) \cdots (\mathbf{A} - b_1\mathbf{I})$  verwendet wird (Bai & Demmel, 1989; Braman, Byers & Mathias, 2002). Wie noch bei den Erläuterungen zum Konvergenzkriterium des *QR*-Algorithmus gezeigt wird erweist es sich dabei als nützlich, wenn für die einzelnen  $q$ 's Schätzungen für die  $b$  kleinsten Eigenwerte von  $\mathbf{A}$  gewählt werden wobei  $b$  nicht größer als sechs werden sollte. Die Matrix  $p(\mathbf{A})$  hat dieselben Eigenvektoren wie  $\mathbf{A}$ , aber andere Eigenwerte und zwar  $p(\lambda_1), \dots, p(\lambda_n)$ . Dies stellt insofern kein Problem da, als das die  $q$ 's bekannt sind und somit die Eigenwerte jederzeit rückrechenbar sind.

Wird in Gleichung 2.7 anstelle eines einzelnen Vektors  $\mathbf{x}$  eine nahezu beliebige  $n \times k$  Matrix  $\mathbf{X} = \mathbf{S}_0$  eingesetzt und anstelle von  $\mathbf{A}$  die Matrix  $p(\mathbf{A})$  resultiert die Folge

$$\mathbf{S}_0, p(\mathbf{A})\mathbf{S}_1, p(\mathbf{A})^2\mathbf{S}_2, p(\mathbf{A})^3\mathbf{S}_3, \dots \quad (2.8)$$

mit

$$\begin{aligned} p(\mathbf{A})^i \mathbf{S}_w &= p(\mathbf{A})^i (s_{1w} \quad s_{2w} \quad \cdots \quad s_{kw}) \\ &= (p(\mathbf{A})^i s_{1w} \quad p(\mathbf{A})^i s_{2w} \quad \cdots \quad p(\mathbf{A})^i s_{kw}) \end{aligned}$$

und

$$s_{tw} = \frac{p(\mathbf{A})^i s_{tw-1}}{\|p(\mathbf{A})^i s_{tw-1}\|}$$

für  $t = 1, \dots, k$ , wobei gemäß der Potenzmethode jedes  $p(\mathbf{A})^i s_{tw}$  zum dominanten Eigenvektor  $\mathbf{c}_1$  konvergiert, wenn  $w = 1 \rightarrow \infty$ ,  $i = 0 \rightarrow \infty$  gilt. Verschiedene Eigenvektoren von  $\mathbf{A}$  bzw.  $p(\mathbf{A})$  resultieren, wenn für  $\mathbf{S}_0$  bzw.  $\mathbf{S}_w$  eine orthonormale Basis gewählt wird.

Eine orthonormale Basis zu  $\mathbf{S}_0$  bzw.  $\mathbf{S}_w$  kann über die Lösung der Gleichung

$$\mathbf{Q}_v^{-1} \mathbf{S}_v = \mathbf{R}_v$$

mit  $v = 0 \rightarrow \infty$  und  $\mathbf{S}_0 = \mathbf{X}$  gefunden werden, wobei  $\mathbf{Q}_v$  die gesuchte orthonormale Basis und  $\mathbf{R}_v$  eine obere Dreiecksmatrix ist.<sup>8</sup> Für  $\mathbf{Q}_v^{-1}$  kann auch

$$\mathbf{Q}_v^{-1} = \mathbf{Q}_{vk}^{-1} \mathbf{Q}_{vk-1}^{-1} \cdots \mathbf{Q}_{v1}^{-1}$$

geschrieben werden (s. Watkins, 2007), sodass  $\mathbf{Q}_{v1}^{-1}$  die erste Spalte von  $\mathbf{R}_v$ ,  $\mathbf{Q}_{v2}^{-1}$  die zweite Spalte von  $\mathbf{R}_v$  und allgemein  $\mathbf{Q}_{vt}^{-1}$  die  $t$ -te Spalte von  $\mathbf{R}_v$  erzeugt.

Zur Bestimmung der einzelnen  $\mathbf{Q}_{vt}$  sind verschiedene Verfahren anwendbar, wobei die Householder Transformation von Householder (1958) eine der gängigsten Varianten darstellt (einen ausführlicheren Überblick über verschiedene andere Verfahren gibt Fischer, 2009). Demnach gilt für  $\mathbf{Q}_{vt}$

$$\mathbf{Q}_{vt} = \mathbf{I} - \beta \mathbf{u}_{vt} \mathbf{u}_{vt}'$$

wobei  $\beta = 2/\|\mathbf{u}_{vt}\|^2$ ,  $\mathbf{u}_{vt} = \gamma^{-1}(\mathbf{s}_v \cdot \mathbf{t} - \alpha \mathbf{e}_v \mathbf{t})$  und  $\gamma = \mathbf{d}' \mathbf{s}_v \cdot \mathbf{t} \neq 0$  für jedes  $\mathbf{d}$  dass die vorhergehende Ungleichung erfüllt sowie  $\alpha \neq 0$  ( $\mathbf{s}_v \cdot \mathbf{t}$  ist die  $t$ -te Spalte von  $\mathbf{S}_v$  und  $\mathbf{e}_v \mathbf{t}$  ein Vektor mit einer Eins an der  $t$ -ten Stelle und sonst Nullen).

Die auf diese Weise ermittelten  $\mathbf{Q}_v$  werden nun anstelle der  $\mathbf{S}_v$  in die Potenzfolge 2.8 eingesetzt, d.h. ausgehend von  $\mathbf{Q}_0$  wird zunächst  $p(\mathbf{A})\mathbf{Q} = \mathbf{S}_1$

<sup>8</sup>Die hier verwendete Matrix  $\mathbf{R}$  ist nicht mit der  $\mathbf{R}$ -Matrize der vorherigen Abschnitte zu verwechseln. Während es sich in diesem Abschnitt um eine Dreiecksmatrix handelt, symbolisierte  $\mathbf{R}$  in den letzten Abschnitten die Korrelationsmatrix.



gebildet, danach wird  $\mathbf{S}_1$  über die Gleichung  $\mathbf{S}_1 = \mathbf{Q}_1 \mathbf{R}_1$  orthonomalisiert um  $p(\mathbf{A})\mathbf{Q}_1 = \mathbf{S}_2$  zu berechnen. Dieser Prozess wird weitergeführt, bis ein bestimmtes Konvergenzkriterium erreicht ist, und er lässt sich allgemeiner als QR-Iteration formulieren:

$$p(\mathbf{A})\mathbf{Q}_{w-1} = \mathbf{S}_w \qquad \mathbf{S}_w = \mathbf{Q}_w \mathbf{R}_w$$

d.h.

$$p(\mathbf{A})\mathbf{Q}_{w-1} = \mathbf{Q}_w \mathbf{R}_w$$

Aber woran ist zu erkennen, dass die QR-Iteration konvergiert ist?

Nach Watkins (2007) ist dies anhand der Ähnlichkeitstransformation von  $\mathbf{A}$  erkennbar

$$\mathbf{A}_w = \mathbf{Q}_w^{-1} \mathbf{A} \mathbf{Q}_w = \mathbf{Q}_w^{-1} \mathbf{Q}_{w-1} \mathbf{A}_{w-1} \mathbf{Q}_{w-1}^{-1} \mathbf{Q}_w = \mathbf{Q}_w^{-1} \mathbf{A}_{w-1} \mathbf{Q}_w$$

wobei sich  $\mathbf{A}_w$  auch in der Gestalt

$$\mathbf{A}_w = \begin{pmatrix} \mathbf{A}_{11}^w & \mathbf{A}_{12}^w \\ \mathbf{A}_{21}^w & \mathbf{A}_{22}^w \end{pmatrix}$$

schreiben lässt. In dieser Gleichung ist  $\mathbf{A}_{21}^w$  von der Ordnung  $(n-k) \times k$  und bei Konvergenz der QR-Iteration gilt  $\mathbf{A}_{21}^w = \mathbf{0}$ , d.h. wenn im Falle von  $w \rightarrow \infty$  die QR-Iteration gegen die Lösung konvergiert dann konvergiert  $\mathbf{A}_{21}^w$  im selben Fall gegen Null. Die Ähnlichkeitstransformation  $\mathbf{A}_w$  bietet einen weiteren Vorteil: Die Eigenwerte von  $\mathbf{A}_w$  sind mit den Eigenwerten von  $\mathbf{A}$  identisch, und da  $\mathbf{A}_{21}^w$  im Falle der Konvergenz gegen Null strebt, strebt  $\mathbf{A}_w$  im Falle der Konvergenz gegen die Blockdiagonalmatrix

$$\mathbf{A}_w = \begin{pmatrix} \mathbf{A}_{11}^w & \mathbf{A}_{12}^w \\ \mathbf{0} & \mathbf{A}_{22}^w \end{pmatrix}$$

deren Eigenwerte mit den Eigenwerten der Diagonalblöcke  $\mathbf{A}_{11}^w$  und  $\mathbf{A}_{22}^w$  identisch sind.

Die Eigenwerte der Diagonalblöcke  $\mathbf{A}_{11}^w$  und  $\mathbf{A}_{22}^w$  lassen sich leicht finden, wenn mit jedem einzelnen Diagonalblock zunächst eine QR-Zerlegung und dann eine Ähnlichkeitstransformation durchgeführt wird (und mit den resultierenden Diagonalblöcken wiederum), bis die Blockdiagonalmatrizen entweder obere Dreiecksmatrizen sind (in diesem Fall stehen die Eigenwerte der Matrix in der Diagonalen) oder die einzelnen Blöcke die Ordnung vier nicht überschreiten, sodass die Eigenwerte über die Lösung des dazugehörigen charakteristischen Polynoms gewonnen werden können. Die so ermittelten Eigenwerte sind schließlich die Eigenwerte von  $\mathbf{A}$ .

Das beschriebene Vorgehen zur Ermittlung der Eigenwerte von  $\mathbf{A}$  würde zwar prinzipiell funktionieren, es erscheint allerdings sehr aufwendig: Nach der QR-Iteration müsste  $\mathbf{A}_w$  gebildet werden um zu überprüfen, ob  $\mathbf{A}_{21}^w$  gegen Null konvergiert ist. Wenn das nicht der Fall ist, müsste ein weiterer Iterationsschritt durchgeführt werden um  $\mathbf{A}_{w+1}$  zu ermitteln. Dieser Prozess wäre solange fortzuführen, bis  $\mathbf{A}_{21}^w$  gegen Null konvergiert ist. Es erhebt sich also die Frage, ob die Matrix  $\mathbf{A}_w$  nicht direkt aus der Matrix  $\mathbf{A}_{w-1}$  gewonnen werden kann ohne den Umweg über die QR-Zerlegung von  $\mathbf{S}_v$ ?

Eine direkte Bestimmung von  $\mathbf{A}_w$  ist möglich, wenn die QR-Iteration an  $p(\mathbf{A}_{w-1})$  durchgeführt wird (Watkins, 2007):

$$p(\mathbf{A}_{w-1}) = \mathbf{Q}_w \mathbf{R}_w \quad \mathbf{A}_w = \mathbf{Q}_w^{-1} \mathbf{A}_{w-1} \mathbf{Q}_w$$

Um die Eigenwerte von  $\mathbf{A}$  zu bestimmen ist also zunächst  $p(\mathbf{A}) = \mathbf{Q}_1 \mathbf{R}_1$  zu bilden und  $\mathbf{A}_1 = \mathbf{Q}_1^{-1} \mathbf{A} \mathbf{Q}_1$  auf Konvergenz zu überprüfen. Wenn  $\mathbf{A}_1$  nicht konvergiert ist, dann ist  $p(\mathbf{A}_1) = \mathbf{Q}_2 \mathbf{R}_2$  und  $\mathbf{A}_2 = \mathbf{Q}_2^{-1} \mathbf{A}_1 \mathbf{Q}_2$  zu ermitteln und die Konvergenz anhand von  $\mathbf{A}_2$  zu überprüfen. Dieser Prozess wird solange weitergeführt, bis in  $\mathbf{A}_w$  die entsprechende Submatrix  $\mathbf{A}_{21}^w$  gegen Null gestrebt ist. Mit den Blockdiagonalmatrizen von  $\mathbf{A}_w$  ist dann in ähnlicher Weise zu verfahren, d.h. für die Matrizen  $\mathbf{A}_{11}^w$  und  $\mathbf{A}_{22}^w$  (und gegebenenfalls auch für weitere Untermatrizen in  $\mathbf{A}_{11}^w$  und  $\mathbf{A}_{22}^w$ ) ist eine QR-Iteration mit Ähnlichkeitstransformation durchzuführen, bis schließlich obere Dreiecksmatrizen oder einzelnen Blöcke mit einem kleineren Rang als fünf resultieren.

Auch wenn über die direkte Bestimmung von  $\mathbf{A}_w$  die Eigenwerte von  $\mathbf{A}$  wesentlich ökonomischer berechnet werden können als bei indirekter Ermittlung von  $\mathbf{A}_w$  sind immer noch mindestens  $n^4$  Berechnungsschritte notwendig um die gesuchten Eigenwerte zu schätzen (Watkins, 2007). Die Anzahl notwendiger Berechnungsschritte kann allerdings durch die Implementierung des sogenannten Bulge-Chasing Algorithmus für die QR-Iteration weiter reduziert werden. Grundlage für den Bulge-Chasing Algorithmus ist dabei die Hessenbergmatrix von  $\mathbf{A}$ .

Die Hessenberg-Form  $\mathbf{H}$  der Matrix  $\mathbf{A}$  sieht folgendermaßen aus

$$\mathbf{H} = \begin{pmatrix} h_{11} & h_{12} & h_{13} & \cdots & h_{1n} \\ h_{21} & h_{22} & h_{23} & \cdots & h_{2n} \\ 0 & h_{32} & h_{33} & \cdots & h_{3n} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & h_{n,n-1} & h_{nn} \end{pmatrix}$$

d.h. für alle  $i > j + 1$  gilt  $h_{ij} = 0$ . Die Hessenbergmatrix von  $\mathbf{A}$  muss nur einmal vor Beginn der QR-Iteration gebildet werden, da sie bei der QR-Iteration

erhalten bleibt. Eine Methode zur Bildung der Hessenbergmatrix wird ausführlich bei Strang (2009) beschrieben; an dieser Stelle soll das Vorgehen nur skizzenhaft dargestellt werden.

Ausgehend von der Partitionierung

$$\begin{aligned} \mathbf{A} &= \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \\ &= \begin{pmatrix} a_{11} & \mathbf{c}' \\ \mathbf{y} & \mathbf{A}_1 \end{pmatrix} \end{aligned}$$

wird ein  $\hat{\mathbf{G}}_1$  gesucht für das  $\hat{\mathbf{G}}_1^{-1}\mathbf{y} = \alpha_1\mathbf{e}_1$  gilt, wobei  $\mathbf{e}_1$  ein Vektor mit einer Eins an der ersten Stelle und Nullen an den anderen Stellen und  $\alpha_1$  ein passender Skalierungsfaktor ist. Für

$$\mathbf{G}_1 = \begin{pmatrix} 1 & \mathbf{0}' \\ \mathbf{0} & \hat{\mathbf{G}}_1 \end{pmatrix}$$

gilt dann

$$\mathbf{G}_1^{-1}\mathbf{A} = \begin{pmatrix} a_{11} & \mathbf{c}' \\ \alpha_1\mathbf{e}_1 & \hat{\mathbf{G}}_1^{-1}\mathbf{A}_1 \end{pmatrix}$$

und

$$\mathbf{H}_{.1} = \mathbf{G}_1^{-1}\mathbf{A}\mathbf{G}_1 = \begin{pmatrix} a_{11} & \mathbf{c}'\hat{\mathbf{G}}_1 \\ \alpha_1\mathbf{e}_1 & \hat{\mathbf{G}}_1^{-1}\mathbf{A}_1\hat{\mathbf{G}}_1 \end{pmatrix}$$

d.h. die erste Spalte von  $\mathbf{A}$  ist in die erste Spalte von  $\mathbf{H}$  transformiert worden, und da es sich um eine Ähnlichkeitstransformation handelt, besitzt  $\mathbf{H}_{.1}$  dieselben Eigenwerte wie  $\mathbf{A}$ . Das skizzierte Vorgehen wird nun mit  $\mathbf{H}_{.u}$  ( $u = 1, \dots, n-1$ ) weitergeführt, bis schließlich  $\mathbf{A}$  vollständig in die Hessenbergmatrix  $\mathbf{H}$  umgewandelt wurde.

Auf  $\mathbf{H}$  ist der Bulge-Chasing Algorithmus (Francis, 1961, 1962; Bai & Demmel, 1989) anwendbar. Mit dem Bulge-Chasing Algorithmus wird die Ähnlichkeitstransformation  $\mathbf{H}_w$  gebildet, ohne  $\mathbf{p}(\mathbf{H}_{w-1})$  explizit bestimmen zu müssen. Dadurch werden Rechenoperationen gespart, sodass in der Regel nicht mehr als  $n^3$  Berechnungsschritte notwendig sind, um die gesuchten Eigenwerte von  $\mathbf{A}$  zu schätzen.

Im ersten Schritt des Algorithmus wird ein Vektor  $\mathbf{x}$  bestimmt, der proportional zur ersten Spalte von  $\mathbf{p}(\mathbf{H}_{w-1})\mathbf{e}_1$  ist ( $\mathbf{e}_1$  ist ein Vektor mit einer

Eins an der ersten Stelle und sonst Nullen). Da  $\mathbf{p}(\mathbf{H}_{w-1})$   $b$ -Hessenberg ist, d.h. für alle  $i > j + b$  gilt  $h_{ij} = 0$ , sind lediglich die ersten  $b + 1$  Elemente von  $\mathbf{p}(\mathbf{H}_{w-1})\mathbf{e}_1$  von Null verschieden. Daher lässt sich  $\mathbf{x}$  über die folgende Iteration bestimmen:

1. Setze  $\mathbf{x} \leftarrow \mathbf{e}_1$ .
2. Für  $k = 1, \dots, b$  setze  $\mathbf{x} \leftarrow \alpha_k(\mathbf{H}_{w-1} - q_k\mathbf{I})\mathbf{x}$ .

Wurde  $\mathbf{x}$  bestimmt, dann wird im zweiten Schritt eine Matrix  $\mathbf{Q}_1$  gebildet, deren erste Spalte proportional zu  $\mathbf{x}$  ist. Dies kann beispielsweise über die bereits oben vorgestellte Householder Transformation erfolgen, denn wenn  $\mathbf{x}' = (\mathbf{y}' \ \mathbf{0}')$  geschrieben wird (wobei  $\mathbf{y}$  von der Ordnung  $b + 1$  ist) und  $\mathbf{Q}_1 = \text{diag}(\hat{\mathbf{Q}}_1 \ \mathbf{I})$  gilt (wobei  $\hat{\mathbf{Q}}_1$  von der Ordnung  $(b+1) \times (b+1)$  ist), dann muss die erste Spalte von  $\hat{\mathbf{Q}}_1$  proportional zu  $\mathbf{y}$  sein, d.h. es muss  $\hat{\mathbf{Q}}_1\mathbf{e}_1 = \beta\mathbf{y}$  bzw.  $\hat{\mathbf{Q}}_1^{-1}\mathbf{y} = \beta^{-1}\mathbf{e}_1$  gelten, was erfüllt ist, wenn  $\hat{\mathbf{Q}}_1^{-1}$  beispielsweise über die Householder Transformation gewonnen wird.

Anhand der ermittelten  $\mathbf{Q}_1$  Matrix wird im dritten Schritt die Ähnlichkeitstransformation

$$\mathbf{H}_w^* = \mathbf{Q}_1^{-1}\mathbf{H}_{w-1}\mathbf{Q}_1$$

gebildet, wobei  $\mathbf{H}_w^*$  prinzipiell nicht mehr in Hessenberg-Form vorliegt sondern einen *Bulg* aufweist, d.h.  $\mathbf{H}_w^*$  ist im Fall von beispielsweise  $b = 2$  von der Form

$$\mathbf{H}_w^* = \begin{pmatrix} \times & \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times & \times \\ & & \times & \times & \times & \times \\ & & & \times & \times & \times \\ & & & & \times & \times \end{pmatrix}$$

und wenn beispielsweise  $b = 3$  gilt, dann gilt für  $\mathbf{H}_w^*$ :

$$\mathbf{H}_w^* = \begin{pmatrix} \times & \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times & \times \\ & & & \times & \times & \times \\ & & & & \times & \times \end{pmatrix}$$

Der vierte Schritt des Bulge-Chasing Algorithmus besteht deswegen darin, die Matrix  $\mathbf{H}_w^*$  durch entsprechende Eliminationsmatrizen in die Hessenbergform zu transformieren, d.h. für die Ähnlichkeitstransformation von  $\mathbf{H}_w$  nach dem Bulge-Chasing Verfahren gilt

$$\begin{aligned} \mathbf{H}_w &= \mathbf{Q}_{n-1}^{-1} \cdots \mathbf{Q}_2^{-1} \mathbf{H}_w^* \mathbf{Q}_2 \cdots \mathbf{Q}_{n-1} \\ &= \underbrace{\mathbf{Q}_{n-1}^{-1} \cdots \mathbf{Q}_2^{-1} \mathbf{Q}_1^{-1}}_{\mathbf{Q}_w^{-1}} \mathbf{H}_{w-1} \underbrace{\mathbf{Q}_1 \mathbf{Q}_2 \cdots \mathbf{Q}_{n-1}}_{\mathbf{Q}_w} \\ &= \begin{pmatrix} \mathbf{H}_{11}^w & \mathbf{H}_{12}^w \\ \mathbf{H}_{21}^w & \mathbf{H}_{22}^w \end{pmatrix} \end{aligned}$$

wobei in  $\mathbf{H}_{21}^w$  nur noch  $h_{b+1,b}$  von Null verschieden ist. Die Überprüfung der Konvergenz des Bulge-Chasing Verfahren wird durch die Hessenbergtransformation der Matrix  $\mathbf{A}$  also relativ einfach: Im Falle der Konvergenz des Bulge-Chasing Algorithmus gilt  $h_{b+1,b} = 0$  und die Eigenwerte der Matrix  $\mathbf{A}$  entsprechen den Eigenwerten der Matrizen  $\mathbf{H}_{11}$  und  $\mathbf{H}_{22}$ .

An dieser Stelle muss noch erwähnt werden, dass in der Praxis schon aufgrund von Rundungsfehlern selten  $h_{b+1,b}$  exakt Null wird, weswegen ein Konvergenzkriterium festgelegt werden muss bei dem wenigstens  $h_{b+1,b} \approx 0$  gilt. Ein allgemein als Gültig akzeptiertes Konvergenzkriterium ist dabei  $|h_{b+1,b}| < u (|h_{b,b}| + |h_{b+1,b+1}|)$  mit  $u \approx 10^{-16}$ .

Offen ist noch die Frage geblieben, welche Werte für die  $b$ -Shifts einzusetzen sind. Wünschenswert sind sicher Shifts, durch die die Konvergenzrate minimiert wird. Da die Konvergenzrate vom Konvergenzquotienten<sup>9</sup>  $|p(\lambda_{k+1})/p(\lambda_k)|$  abhängt (je kleiner dieser Koeffizient wird, desto weniger Iterationsschritte sind erforderlich) und gleichzeitig für die Eigenwerte von  $\mathbf{H}$  ohne Beschränkung der Allgemeingültigkeit  $|p(\lambda_1)| \geq \cdots \geq |p(\lambda_k)| \geq |p(\lambda_{k+1})| \geq \cdots \geq |p(\lambda_n)|$  gelten soll, kommen als  $m$ -Shifts damit (nur) die  $m$  Eigenwerte der  $m \times m$  Matrix  $\mathbf{H}_{22}$  in Frage, da dann (wie leicht zu überprüfen ist) die Ausdrücke  $|p(\lambda_t)|$  ( $t = 1, \dots, k$ ) größer als Null werden, während die Ausdrücke  $|p(\lambda_j)|$  ( $j = k + 1, \dots, n$ ) und der Konvergenzquotient in der Praxis gegen Null konvergieren, also sehr klein werden.

Durch den QR-Algorithmus bzw. das Bulge-Chasing Verfahrens können also die Eigenwerte der Korrelationsmatrix  $\mathbf{R}$  ermittelt werden. Dazu ist lediglich  $\mathbf{A} = \mathbf{R}$  zu setzen, und eines der beiden beschriebenen Verfahren anzuwenden.<sup>10</sup> Um auch die Eigenvektoren von  $\mathbf{R}$  zu erhalten, sind zunächst die

<sup>9</sup>Der Konvergenzquotient kann als quantitatives Maß für die Rate der Konvergenz verstanden werden (Strang, 2009).

<sup>10</sup>Mit  $\mathbf{R}$  ist die Korrelationsmatrix und nicht die Dreiecksmatrix des QR-Verfahrens gemeint.





mit:

$$\begin{aligned}\dot{Q}^* &= \begin{pmatrix} Q_1 & 0 \\ 0 & Q_2 \end{pmatrix} \\ D &= \begin{pmatrix} \Lambda_1 & 0 \\ 0 & \Lambda_2 \end{pmatrix} \\ l &= \begin{pmatrix} Q'_1 & 0 \\ 0 & Q'_2 \end{pmatrix} m \\ \kappa &= h_{m+1,m}\end{aligned}$$

Die Eigenwerte von  $\dot{H}^*$  entsprechen also den Eigenwerten der Matrix  $D + \kappa ll'$ . Bevor auf die Schätzung der Eigenwerte dieser Matrix eingegangen wird soll gezeigt werden, dass bei den folgenden Ausführungen nur Probleme betrachtet werden müssen, bei denen  $l \neq 0$  und  $d_1 < d_2 < \dots < d_p$  bzw.  $d_1 > d_2 > \dots > d_p$  gilt. Diese Annahmen sind nicht sehr einschränkend, da sie durch Deflation des Ausdrucks  $D + \kappa ll'$  problemlos erfüllt werden können.

Mit Deflation ist ein Prozesse gemeint, bei dem der Ausdruck  $\dot{D} + \kappa \dot{l} \dot{l}'$  mit der Dimension  $\dot{p} \times \dot{p}$  auf den  $p \times p$  Ausdruck  $D + \kappa ll'$  reduziert wird, wobei  $p < \dot{p}$  gilt. Dies geschieht im Prinzip durch streichen von Zeilen bzw. Spalten in  $\dot{D}$  bzw.  $\dot{l} \dot{l}'$ . Zur Veranschaulichung des Konzepts sei angenommen, dass  $l_i = 0$  ( $i = 1, \dots, p$ ) für ein  $i$  gilt. Wie leicht ersichtlich, gilt in diesem Fall

$$\begin{aligned}(D + \kappa ll')e_i &= d_i e_i + \kappa l l_i \\ &= d_i e_i\end{aligned}$$

d.h.  $d_i$  ist ein Eigenwert von  $\dot{H}^*$  und  $e_i$  der zugehörige Eigenvektor. Zur Bestimmung der übrigen Eigenwerte und Eigenvektoren muss also nur das um die  $i$ -te Zeile bzw. Spalte reduzierte Problem  $D + \kappa ll'$  betrachtet werden (ausführlicher zur Deflation berichten Bunch, Nielsen & Sorensen, 1978).

Im Folgenden wird also davon ausgegangen, dass  $l \neq 0$  und  $d_1 < d_2 < \dots < d_p$  bzw.  $d_1 > d_2 > \dots > d_p$  gilt. Zum Auffinden der Eigenwerte der bereits durch Deflation reduzierten Matrix  $D + \kappa ll'$  wird das charakteristische Polynom gebildet und Null gesetzt:

$$\begin{aligned}\det(D + \kappa ll' - \lambda I) &= \det((D - \lambda I)(I + \kappa(D - \lambda I)^{-1} ll')) \\ &= \det(D - \lambda I) \det(I + \kappa(D - \lambda I)^{-1} ll') \\ &= 0\end{aligned}\tag{2.9}$$

Für den Ausdruck  $D - \lambda I$  kann ohne Beschränkung der Allgemeinheit angenommen werden dass er regulär ist, da andernfalls mindestens ein Element



von  $\mathbf{l}$  null werden würde (s. Cuppen, 1981) was den Annahmen widerspricht. Es gelte also  $\det(\mathbf{D} - \lambda\mathbf{I}) = \prod_{j=1}^p (d_j - \lambda) \neq 0$ ; wie anhand von Gleichung 2.9 ersichtlich, muss dann der Ausdruck  $\det(\mathbf{I} + \kappa(\mathbf{D} - \lambda\mathbf{I})^{-1}\mathbf{l}\mathbf{l}')$  null werden.

Zur Lösung der Gleichung  $\det(\mathbf{I} + \kappa(\mathbf{D} - \lambda\mathbf{I})^{-1}\mathbf{l}\mathbf{l}') = 0$  wird auf das Determinantenlemma zurückgegriffen (da der Beweis des Lemmas sehr einfach ist und bei Ding und Zhou (2007) nachgelesen werden kann, soll auf die Beweisführung verzichtet werden).

**Lemma 2.3.1:** *Wenn  $\mathbf{A}$  eine invertierbare  $p \times p$  Matrix ist und  $\mathbf{u}$  und  $\mathbf{v}$  zwei  $p$ -dimensionale Vektoren, dann gilt:*

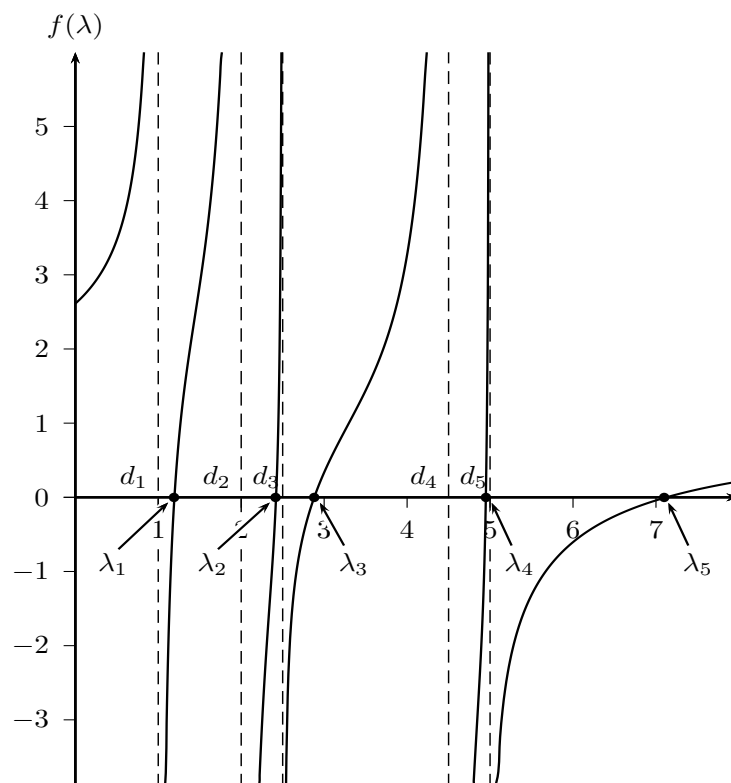
$$\det(\mathbf{A} + \mathbf{u}\mathbf{v}') = (1 + \mathbf{v}'\mathbf{A}^{-1}\mathbf{u})\det(\mathbf{A}) \quad \square$$

Wird nun für  $\mathbf{A} = \mathbf{I}$  angenommen, reduziert sich das Lemma zur spezifischen Form  $\det(\mathbf{I} + \mathbf{u}\mathbf{v}') = 1 + \mathbf{v}'\mathbf{u}$  so dass:

$$\begin{aligned} f(\lambda) &= \det(\mathbf{I} + \underbrace{\kappa(\mathbf{D} - \lambda\mathbf{I})^{-1}\mathbf{l}}_{=: \mathbf{u}} \underbrace{(\mathbf{l}')}_{=: \mathbf{v}'}) = 1 + \mathbf{v}'\mathbf{u} \\ &= 1 + \kappa\mathbf{l}'(\mathbf{D} - \lambda\mathbf{I})^{-1}\mathbf{l} \\ &= 1 + \kappa \sum_{j=1}^p \frac{l_j^2}{d_j - \lambda} \end{aligned}$$

Die gesuchten Eigenwerte von  $\mathbf{D} + \kappa\mathbf{l}\mathbf{l}'$  entsprechen somit den Nullstellen der Funktion  $f(\lambda)$ . Zur Veranschaulichung der Funktion  $f(\lambda)$  ist der zugehörige Graph für  $p = 5$ ,  $\mathbf{d} = (d_1, d_2, d_3, d_4, d_5) = (1, 2, 2.5, 4.5, 5)$ ,  $\mathbf{l}^2 = (l_1, l_2, l_3, l_4, l_5) = (0.2, 0.4, 0.04, 0.5, 0.05)$  und  $\kappa = 3$  in Abbildung 2.1 auf der folgenden Seite dargestellt. Wie zu erkennen, bilden die Nullstellen des Graph  $f(\lambda)$  die gesuchten Eigenwerte  $\lambda_j$  der Matrix  $\mathbf{D} + \kappa\mathbf{l}\mathbf{l}'$  die, sofern  $\mathbf{D} = \text{diag}(d_1, \dots, d_p)$  mit  $d_1 < d_2 < \dots < d_p$  gilt (was leicht durch entsprechende Permutationsmatrizen erreicht werden kann), im Falle von  $\kappa > 0$  zwischen den  $d_i$ -Werten liegen d.h. es gilt  $d_1 < \lambda_1 < \dots < d_p < \lambda_p < d_p + \kappa\mathbf{l}'\mathbf{l}$  (wenn hingegen  $\kappa < 0$  ist, dann sind die Verhältniszeichen des letzten Ausdrucks umzudrehen).

Zur numerischen Bestimmung der gesuchten Eigenwerte von  $\mathbf{H}^*$  sind die Nullstellen der Funktion  $f(\lambda)$  analytisch zu gewinnen. Zur Nullstellenbestimmung wird (wie allgemein üblich) eine Interpolationsfunktion gebildet. Allerdings kann nicht auf die gängige lineare Interpolationstechnik nach Newton zurückgegriffen werden (Kasper, 2006), da  $f(\lambda)$  eine rationale Funktion mit Polen ist und lineare Funktionen nicht zur Approximation von Stützstellen an den Polen geeignet sind (Schaback & Wendland, 2004).


 Abbildung 2.1: Charakteristischer Graph der Funktion  $f(\lambda)$  mit  $\kappa = 3$ .

Als alternatives Verfahren zum Auffinden der gesuchten Nullstellen bietet sich die rationale Interpolation an, weil die rationale Interpolation auf einer rationalen Interpolationsfunktion basiert die ebenso wie  $f(\lambda)$  Pole aufweist und diese deswegen auch approximieren kann. Bunch et al. (1978) haben einen auf rationale Interpolation basierenden Algorithmus für das hier vorliegende Problem entwickelt, der in R bzw. LAPACK implementiert ist. Deswegen sollen die Grundzüge dieses Verfahrens an dieser Stelle kurz vorgestellt werden. Auf die Wiedergabe von Beweisen (s. dazu Bunch et al., 1978) wird dabei ebenso verzichtet wie auf die Besprechung von alternativen Interpolationsalgorithmen (s. dazu Gu & Eisenstat, 1992).

Ausgangspunkt der rationalen Interpolation bildet die Gleichung

$$\begin{aligned} f(\lambda_i) &= 1 + \kappa \sum_{j=1}^p \frac{l_j^2}{d_j - \lambda_i} \\ &= 1 + \kappa \sum_{j=1}^p \frac{l_j^2}{d_j - (d_i + \kappa\mu_i)} \end{aligned}$$

$$\begin{aligned}
&= 1 + \sum_{j=1}^p \frac{l_j^2}{(d_j - d_i)/\kappa - \mu_i} \\
&= 1 + \sum_{j=1}^p \frac{l_j^2}{\gamma_j - \mu_i}
\end{aligned}$$

mit  $\gamma_j = (d_j - d_i)/\kappa$ , und gesucht werden die Nullstellen  $\mu_i$  der Funktion  $f(\lambda_i)$ . Im Weiteren sei (wie bisher auch) angenommen, dass  $\kappa > 0$  und  $1 \leq i < p$  gilt (der Fall  $i = p$  wird später behandelt). Da ferner

$$\phi_1(u_i) \equiv \sum_{j=1}^i \frac{l_j^2}{\gamma_j - u_i} \quad \text{und} \quad \phi_2(u_i) \equiv \sum_{j=i+1}^p \frac{l_j^2}{\gamma_j - u_i}$$

geschrieben werden kann (mit  $u_i$  als Schätzung für  $\mu_i$ ), muss  $u_i$  so bestimmt werden, dass die Gleichung

$$\phi_2(u_i) + 1 = -\phi_1(u_i)$$

erfüllt ist.

Bunch et al. (1978) können zeigen, dass ausgehend von einem beliebigen Startwert  $u_0$  die Sequenz  $u_0, u_1, \dots, u_n$  gegen die gesuchte Nullstelle konvergiert, sofern für  $u_l$  ( $l = 1, \dots, n$ )

$$u_l = u_{l-1} + 2a/(b + \sqrt{b^2 - 4a}) \quad (2.10)$$

mit

$$\begin{aligned}
a &= (d\eta\phi_1)/(\phi_1'c) \\
b &= (d(1 + \phi_2) + \phi_1^2/\phi_1')/c + \phi_1/\phi_1' \\
c &= 1 + \phi_2 - d\phi_2' \\
d &= \gamma - u_{l-1} \\
\eta &= 1 + \phi_1 + \phi_2
\end{aligned}$$

gewählt wird. Für  $i = p$  ist anstelle der Formel 2.10 die Gleichung

$$u_l = u_{l-1} + \left( \frac{1 + \phi_{1l-1}}{\phi_{1l-1}'} \right) \phi_{1l-1}$$

zu verwenden. Als Abbruchkriterium der Iteration empfehlen Bunch et al. (1978)

$$|u_l - u_{l-1}| < \varepsilon |u_{l-1}|$$

wobei  $\varepsilon$  die relative Maschinengenauigkeit ist, die bei heutigen Computern in der Regel mit  $6 \cdot 10^{-8}$  (einfache Genauigkeit) oder  $1,1 \cdot 10^{-16}$  (doppelte Genauigkeit) beziffert werden kann.

Ist die gesuchte Nullstelle gefunden, dann ist über die Gleichung  $\lambda_i = d_i + \kappa \mu_i$  durch einsetzen von  $\mu_i \approx u_i$  der entsprechende Eigenwert zu ermitteln. Der dazugehörige Eigenvektor ergibt sich dann aus der Formel (s. Gu & Eisenstat, 1995)

$$c_i = \left( \frac{\hat{l}_1}{d_1 - \lambda_i}, \dots, \frac{\hat{l}_p}{d_p - \lambda_i} \right)' / \sqrt{\sum_{j=1}^p \frac{\hat{l}_j^2}{(d_j - \lambda_i)^2}}$$

wobei für  $\hat{l}_j$  gilt:

$$\hat{l}_j = \sqrt{\prod_{i=1}^{j-1} \frac{(\lambda_i - d_j)}{(d_i - d_j)} \cdot \prod_{i=j}^{p-1} \frac{(\lambda_i - d_j)}{(d_{i+1} - d_j)} \cdot (\lambda_p - d_j)}$$

### Bestimmung von $\lambda$ und $c$ mit R

Die vorangehenden Erläuterungen dienten dazu, einen Einblick in die Schätzalgorithmen zur Bestimmung von  $\lambda$  und  $c$  zu geben. Die faktischen Berechnungen sind selbstverständlich dem Computer zu überlassen. So können die Eigenwerte und Eigenvektoren bzw. die Hauptkomponenten der Matrix  $\mathbf{R}$  bequem über das Statistikprogramm R ermittelt werden.

**Beispiel 2.3.2:** Für

$$\mathbf{R} = \begin{pmatrix} 1 & 0.4 \\ 0.4 & 1 \end{pmatrix}$$

liefert die Syntax

```
R <- matrix(c(1,0.4,0.4,1),nrow=2)
eigen(R)
```

das Ergebnis

```
$values
[1]      1.4      0.6

$vectors
      [,1]      [,2]
[1,] 0.7071068 -0.7071068
[2,] 0.7071068  0.7071068
```

wobei (wie schon durch Beispiel 2.3.1 gezeigt)  $\lambda_1 = 1.4$  der erste Eigenwert von  $\mathbf{R}$  ist und  $\lambda_2 = 0.6$  der zweite Eigenwert. Die zugehörigen Eigenvektoren lauten  $\mathbf{c}'_1 = (0.7071068 \ 0.7071068)$  bzw.  $\mathbf{c}'_2 = (-0.7071068 \ 0.7071068)$  und entsprechen (bis auf Vorzeichenvertauschungen) ebenfalls den in Beispiel 2.3.1 ermittelten Ergebnissen.  $\triangleleft$

## 2.3.3 Explorative Faktorenanalyse

### 2.3.3.1 Strukturmodell

Die grundlegenden Annahme der explorativen Faktorenanalyse ist, dass das Ergebnis bei einer Aufgabe zum einen von Einflüssen abhängt, die nur bei der Bearbeitung dieser Aufgabe wirksam sind, und zum anderen von Merkmalen, die bei allen Aufgaben eines Tests vorhanden sind. Die Einflüsse, die bei allen Aufgaben wirksam sind, werden gemeinsame Faktoren genannt; zu den Einflüssen, die nur bei der Bearbeitung einer Aufgabe wirken, zählen die spezifischen Faktoren sowie die Messfehler.

Die Faktoren werden in der explorativen Faktorenanalyse als *latente Merkmale* gedacht, d.h. nicht direkt beobachtbare Eigenschaften der Personen. Nach dem explorativen Faktorenmodell begünstigen demnach bestimmte Eigenschaften von Personen das Lösen mehrerer verschiedener Aufgaben, während andere Eigenschaften das Lösen nur einer sehr konkreten Aufgabe erleichtern. Formal lässt sich der geschilderte Sachverhalt durch das Strukturmodell

$$\mathbf{Y} = \mathbf{M} + \mathbf{FL}' + \mathbf{E} \quad (2.11)$$

darstellen, wobei

$$\mathbf{Y} = \begin{pmatrix} y_{11} & \cdots & y_{1p} \\ \vdots & & \vdots \\ y_{n1} & \cdots & y_{np} \end{pmatrix} \quad \mathbf{M} = \begin{pmatrix} \mu_{.1} & \cdots & \mu_{.p} \\ \vdots & & \vdots \\ \mu_{.1} & \cdots & \mu_{.p} \end{pmatrix}$$

$$\mathbf{F} = \begin{pmatrix} f_{11} & \cdots & f_{1k} \\ \vdots & & \vdots \\ f_{n1} & \cdots & f_{nk} \end{pmatrix} \quad \mathbf{L}' = \begin{pmatrix} l_{11} & \cdots & l_{p1} \\ \vdots & & \vdots \\ l_{1k} & \cdots & l_{pk} \end{pmatrix}$$

$$\mathbf{E} = \begin{pmatrix} e_{11} & \cdots & e_{1p} \\ \vdots & & \vdots \\ e_{n1} & \cdots & e_{np} \end{pmatrix}$$

gilt.  $\mathbf{M}$  enthält die Mittelwerte  $\mu_j$  der Variablen  $y_j$ . In  $\mathbf{F}$  stehen die gemeinsamen Faktoren, wobei beispielsweise  $f_{i1}$  das erste latente Merkmal symbolisiert, durch das alle Aufgabenergebnisse beeinflusst werden. Die Stärke des Einflusses wird durch  $\mathbf{L}$  dargestellt. Schließlich enthalten die Komponenten von  $\mathbf{E}$  sowohl die spezifischen Faktoren als auch die Messfehler.

Wird in dem vollständigen explorativen Faktorenmodell der Gleichung 2.11 der Personenindex weggelassen, resultiert ein Strukturmodell auf Ebene der Aufgaben

$$\mathbf{y} = \boldsymbol{\mu} + \mathbf{L}\mathbf{f} + \mathbf{e} \quad (2.12)$$

mit

$$\mathbf{y} = \begin{pmatrix} y_{.1} \\ \vdots \\ y_{.p} \end{pmatrix} \quad \boldsymbol{\mu} = \begin{pmatrix} \mu_{.1} \\ \vdots \\ \mu_{.p} \end{pmatrix}$$

$$\mathbf{L} = \begin{pmatrix} l_{11} & \cdots & l_{1k} \\ \vdots & & \vdots \\ l_{p1} & \cdots & l_{pk} \end{pmatrix} \quad \mathbf{f} = \begin{pmatrix} f_{.1} \\ \vdots \\ f_{.k} \end{pmatrix}$$

$$\mathbf{e} = \begin{pmatrix} e_{.1} \\ \vdots \\ e_{.p} \end{pmatrix}$$

In dieser reduzierten Strukturgleichung des explorativen Faktorenmodells wird dargestellt, aus welchen Komponenten sich ein Aufgabenergebnis zusammensetzt, unabhängig von konkreten Werten. Es steht also  $y_j$  für die Variable  $j$ ,  $\mu_j$  für den Mittelwert der Variable  $j$ ,  $f_m$  für den  $m$ -ten gemeinsamen Faktor ( $m = 1, \dots, k$  und  $k \in j$ ; zur Bestimmung von  $k$  siehe Kapitel 2.5) und  $e_j$  für den  $j$ -ten spezifischen Faktor. Die nachfolgenden Ausführungen beziehen sich, sofern nicht anders erläutert, auf diese reduzierte Strukturgleichung des explorativen Faktorenmodells.

### 2.3.3.2 Voraussetzungen

Ähnlich wie bei der Hauptkomponentenmethode wird angenommen, dass  $\mathbf{y}$ ,  $\boldsymbol{\mu}$ ,  $\mathbf{L}$ ,  $\mathbf{f}$  und  $\mathbf{e}$  intervallskaliert mit  $\mathbf{y} \in \mathbb{R}^{p \times 1}$ ,  $\boldsymbol{\mu} \in \mathbb{R}^{p \times 1}$ ,  $\mathbf{L} \in \mathbb{R}^{p \times k}$ ,  $\mathbf{f} \in \mathbb{R}^{k \times 1}$  und  $\mathbf{e} \in \mathbb{R}^{p \times 1}$  sind. Darüber hinaus wird angenommen, dass  $\mathbf{e}$  und  $\mathbf{f}$  einen Erwartungswert von Null haben und dass die Fehlerterme  $\mathbf{e}$  weder untereinander noch mit den gemeinsamen Faktoren korrelieren. Die gemeinsamen

Faktoren sollen außerdem standardisiert sein. Da die Standardisierung von  $\mathbf{f}$  ohne Probleme zu erzielen ist, ergeben sich insgesamt sechs Voraussetzungen:

1.  $\mathbf{y} \in \mathbb{R}^{p \times 1}, \boldsymbol{\mu} \in \mathbb{R}^{p \times 1}, \mathbf{L} \in \mathbb{R}^{p \times k}, \mathbf{f} \in \mathbb{R}^{k \times 1}$  und  $\mathbf{e} \in \mathbb{R}^{p \times 1}$ .
2.  $\mathbf{y}, \boldsymbol{\mu}, \mathbf{L}, \mathbf{f}$  und  $\mathbf{e}$  intervallskaliert.
3.  $\mathbf{E}(\mathbf{f}) = \mathbf{0}$ .
4.  $\mathbf{E}(\mathbf{e}) = \mathbf{0}$ .
5.  $\text{cov}(\mathbf{e}) = \mathbf{E}(\mathbf{e}\mathbf{e}') = \mathbf{D} = \text{diag}\{e_{.1}^2, \dots, e_{.p}^2\}$ .
6.  $\text{cov}(\mathbf{f}\mathbf{e}') = \mathbf{E}(\mathbf{f}\mathbf{e}') = \mathbf{0}$ .

Sind die gemeinsamen Faktoren überdies nicht korreliert, dann wird vom orthogonalen Faktorenmodell gesprochen, andernfalls vom obliquen Faktorenmodell.<sup>11</sup> Im orthogonalen Faktorenmodell gilt demnach:

$$7. \quad \text{cov}(\mathbf{f}) = \mathbf{E}(\mathbf{f}\mathbf{f}') = \mathbf{I}.$$

Unter diesen Annahmen lässt sich für die Kovarianzmatrix  $\boldsymbol{\Sigma}$  von  $\mathbf{y}$  schreiben

$$\begin{aligned} \boldsymbol{\Sigma} &= \mathbf{E}[(\mathbf{y} - \boldsymbol{\mu})(\mathbf{y} - \boldsymbol{\mu})'] \\ &= \mathbf{E}[(\mathbf{L}\mathbf{f} + \mathbf{e})(\mathbf{L}\mathbf{f} + \mathbf{e})'] \\ &= \mathbf{E}(\mathbf{L}\mathbf{f}\mathbf{f}'\mathbf{L}') + \mathbf{E}(\mathbf{L}\mathbf{f}\mathbf{e}') + \mathbf{E}(\mathbf{e}\mathbf{f}'\mathbf{L}') + \mathbf{E}(\mathbf{e}\mathbf{e}') \\ &= \mathbf{L}\mathbf{L}' + \mathbf{D} \end{aligned} \quad (2.13)$$

d.h. die Varianzen  $\sigma_{jj}^2$  der Aufgabenergebnisse hängen vom Einfluss der gemeinsamen Faktoren und der Fehlerterme  $\mathbf{e}$  ab, und die Kovarianzen  $\sigma_{jt}$  ( $j, t = 1, \dots, p$  und  $j \neq t$ ) zwischen den Aufgabenergebnissen können vollständig durch den Einfluss der gemeinsamen Faktoren erklärt werden:

$$\begin{aligned} \sigma_{jj}^2 &= \sum_{q=1}^k l_{jq}^2 + d_j^2 \\ \sigma_{jt} &= \sum_{q=1}^k l_{jq}l_{tq} \end{aligned} \quad (2.14)$$

Die Ausdrücke  $h_j^2 = \sum_{q=1}^k l_{jq}^2$  werden Kommunalität genannt und bezeichnen den Anteil aufgeklärter Varianz der Aufgabenergebnisse durch die gemeinsamen Faktoren. Bei standardisierten Werten darf dieser Anteil nicht über Eins liegen; unter spezifischen Bedingungen können Schätzungen dieser Anteile allerdings größer als Eins werden, was als Heywood Cases bezeichnet wird. Wie in einem solchen Fall vorzugehen ist, wird später erläutert (S. 43).

<sup>11</sup>Zugunsten der Übersichtlichkeit wird auf die algebraische Darstellung des obliquen Faktorenmodells verzichtet. Der interessierte Leser sei auf Kasper (2006) verwiesen.

Zu erwähnen ist noch, das  $\mathbf{L}$  wie schon bei der Hauptkomponentenmethode als Ladungsmatrix bezeichnet wird und die Kovarianzen zwischen den gemeinsamen Faktoren und den Aufgabenergebnissen enthält:

$$\begin{aligned}\text{cov}(\mathbf{y}, \mathbf{f}) &= \text{E}[(\mathbf{y} - \boldsymbol{\mu})\mathbf{f}'] \\ &= \text{E}[(\mathbf{L}\mathbf{f} + \mathbf{e})\mathbf{f}'] \\ &= \text{E}[\mathbf{L}\mathbf{f}\mathbf{f}' + \mathbf{e}\mathbf{f}'] \\ &= \mathbf{L}\text{E}(\mathbf{f}\mathbf{f}') + \text{E}(\mathbf{e}\mathbf{f}') \\ &= \mathbf{L}\end{aligned}$$

### 2.3.3.3 Parameterschätzung

#### Ausgangspunkt

Für die Schätzung der Parameter in  $\mathbf{L}$  und  $\mathbf{D}$  stehen verschiedene Verfahren zur Verfügung (Harman, 1967); auf die beiden häufigsten Vorgehensweisen (die Minimierung der allgemeinen Diskrepanzfunktion und die Maximum-Likelihood-Methode) soll im nachfolgenden Eingegangen werden. Danach wird erläutert, wie  $\mathbf{F}$  zu schätzen ist.

#### Schätzung von $\mathbf{L}$ und $\mathbf{D}$ über die allgemeine Diskrepanzfunktion

Die allgemeine Diskrepanzfunktion wird wie folgt geschrieben (Browne, 1974)

$$F(\mathbf{S}, \boldsymbol{\Sigma}|\mathbf{V}) = \frac{1}{2} \text{Tr}[\mathbf{V}^{-1}(\mathbf{S} - \boldsymbol{\Sigma})]^2 \quad (2.15)$$

wobei  $\mathbf{S}$  eine erwartungstreue Schätzung der Kovarianzmatrix von  $\mathbf{y}$  aufgrund der Daten ist,  $\boldsymbol{\Sigma}$  die modellbasierte Kovarianzmatrix von  $\mathbf{y}$  darstellt, welche sich auf der Basis des explorativen Faktorenmodells bzw. der Gleichung 2.13 ergibt, und  $\mathbf{V}$  eine Koeffizientenmatrix symbolisiert.

Werden die Parameter in  $\mathbf{L}$  und  $\mathbf{D}$  über die Minimierung der allgemeinen Diskrepanzfunktion geschätzt, dann wird die Abweichung zwischen  $\boldsymbol{\Sigma}$  und  $\mathbf{S}$  minimiert, d.h. die Parameter des explorativen Faktorenmodells werden so geschätzt, dass die empirischen Varianzen und Kovarianzen von  $\mathbf{y}$  möglichst genau erklärt werden können. Allerdings ist die Verwendung der allgemeinen Diskrepanzfunktion an drei Voraussetzungen gebunden: Erstens müssen die vierten Kumulanten von  $\mathbf{y}$  Null sein (was beispielsweise der Fall ist, wenn  $\mathbf{y}$  multivariat Normalverteilt ist), zweitens muss die Matrix  $\mathbf{V}$  positiv definit sein, d.h.  $\mathbf{V}$  muss symmetrisch und invertierbar sein, und drittens ist zu



fordern, dass  $\text{plim}_{n \rightarrow \infty} \mathbf{V} = \boldsymbol{\Sigma}_{\diamond}$  gilt, wobei  $\boldsymbol{\Sigma}_{\diamond}$  die Kovarianzmatrix von  $\mathbf{y}$  in der Population ist (also  $E(\mathbf{S}) = \boldsymbol{\Sigma}_{\diamond}$ ).

Sind alle drei Bedingungen erfüllt, dann werden die über die Minimierung der allgemeinen Diskrepanzfunktion geschätzten Parameter als **best generalised least square estimators** (BGLS) bezeichnet, d.h. von allen generalisierten Kleinstquadrateschätzern weisen sie die geringste Varianz auf. Ist nur die erste Bedingung nicht erfüllt, dann sind die Parameterschätzungen zwar weiterhin konsistent, aber die Standardfehler der Schätzer und damit auch darauf basierende Signifikanztests sind nicht mehr erwartungstreu, weswegen die Standardfehler und Signifikanztests in diesem Fall einer Korrektur bedürfen (s. dazu Browne, 1984). Verletzungen der letzten beiden Bedingungen führen dazu, dass die Gleichung 2.15 in dieser Form nicht mehr gültig ist (Browne, 1982), d.h. eine Schätzung von  $\mathbf{L}$  und  $\mathbf{D}$  über die allgemeine Diskrepanzfunktion ist nicht mehr möglich.

Um  $\mathbf{L}$  und  $\mathbf{D}$  über die allgemeine Diskrepanzfunktion schätzen zu können, soll im Folgenden  $\mathbf{V} = \mathbf{S}$  und  $\text{rg}(\mathbf{S}) = p$  angenommen werden, da dann die letzten beiden Bedingungen der allgemeinen Diskrepanzfunktion erfüllt sind (Browne, 1974). Damit weiterhin auch erwartungstreue Standardschätzfehler resultieren, sollen die vierten Kumulanten von  $\mathbf{y}$  Null sein.

Unter diesen Bedingungen gilt für die Diskrepanzfunktion (Browne, 1982)

1.  $F(\mathbf{S}, \boldsymbol{\Sigma} | \mathbf{V} = \mathbf{S}) \geq 0$ ,
2.  $F(\mathbf{S}, \boldsymbol{\Sigma} | \mathbf{V} = \mathbf{S}) = 0$  dann und nur dann wenn  $\boldsymbol{\Sigma} = \mathbf{S}$ , und
3.  $F(\mathbf{S}, \boldsymbol{\Sigma} | \mathbf{V} = \mathbf{S})$  ist kontinuierlich in  $\mathbf{S}$  und  $\boldsymbol{\Sigma}$ ,

d.h.  $F(\mathbf{S}, \boldsymbol{\Sigma} | \mathbf{V} = \mathbf{S})$  ist differenzierbar in  $\mathbf{S}$  und  $\boldsymbol{\Sigma}$  und hat sein Minimum an der Stelle  $\boldsymbol{\Sigma} = \mathbf{S}$ . Zur Schätzung der Parameter  $\mathbf{L}$  und  $\mathbf{D}$  sind also die partiellen Ableitungen der generalized least square Funktion

$$\begin{aligned} F(\mathbf{S}, \boldsymbol{\Sigma} | \mathbf{S}) &= \frac{1}{2} \text{Tr}[\mathbf{S}^{-1}(\mathbf{S} - \boldsymbol{\Sigma})]^2 \\ &= \frac{1}{2} \text{Tr}[\mathbf{I} - \mathbf{S}^{-1}\boldsymbol{\Sigma}]^2 \end{aligned} \quad (2.16)$$

nach  $\mathbf{L}$  und  $\mathbf{D}$  zu bilden und die Nullstellen dieser Ableitungsfunktionen zu bestimmen.

Die erste partielle Ableitung von  $F(\mathbf{S}, \boldsymbol{\Sigma} | \mathbf{S})$  nach  $\mathbf{L}$  lautet (s. Jöreskog, 1972)

$$\frac{\partial F(\mathbf{S}, \boldsymbol{\Sigma} | \mathbf{S})}{\partial \mathbf{L}} = 2\mathbf{S}^{-1}(\boldsymbol{\Sigma} - \mathbf{S})\mathbf{S}^{-1}\mathbf{L}$$

und nach Nullsetzen dieser partiellen Ableitung gilt<sup>12</sup>:

$$\begin{array}{rcl} \hat{\Sigma}S^{-1}\mathbf{L} - \mathbf{L} & = & \mathbf{0} \quad | + \mathbf{L} \\ \hat{\Sigma}S^{-1}\mathbf{L} & = & \mathbf{L} \quad | \cdot \hat{\Sigma}^{-1} \\ S^{-1}\mathbf{L} & = & \hat{\Sigma}^{-1}\mathbf{L} \end{array} \quad (2.17)$$

Nach der Sherman-Morrison-Woodbury-Formel (auch Woodbury-Formel oder Woodbury-Matrix-Identität genannt) kann

$$\begin{aligned} \hat{\Sigma}^{-1} &= (\mathbf{L}\mathbf{L}' + \mathbf{D})^{-1} \\ &= (\mathbf{D} + \mathbf{L}\mathbf{L}')^{-1} \\ &= \mathbf{D}^{-1} - \mathbf{D}^{-1}\mathbf{L}(\mathbf{I} + \mathbf{L}'\mathbf{D}^{-1}\mathbf{L})^{-1}\mathbf{L}'\mathbf{D}^{-1} \end{aligned} \quad (2.18)$$

geschrieben werden (Duncan, 1944; Guttman, 1946; Sherman & Morrison, 1950; Henderson & Searle, 1981; Hager, 1989). Einsetzen dieses Ausdrucks in Gleichung 2.17 führt zu

$$\begin{aligned} S^{-1}\mathbf{L} &= [\mathbf{D}^{-1} - \mathbf{D}^{-1}\mathbf{L}(\mathbf{I} + \mathbf{L}'\mathbf{D}^{-1}\mathbf{L})^{-1}\mathbf{L}'\mathbf{D}^{-1}]\mathbf{L} \\ &= \mathbf{D}^{-1}\mathbf{L}[\mathbf{I} - (\mathbf{I} + \mathbf{L}'\mathbf{D}^{-1}\mathbf{L})^{-1}\mathbf{L}'\mathbf{D}^{-1}\mathbf{L}] \\ &= \mathbf{D}^{-1}\mathbf{L}[\mathbf{I} - \mathbf{L}\mathbf{L}'(\mathbf{I} + \mathbf{L}'\mathbf{D}^{-1}\mathbf{L})^{-1}\mathbf{L}'\mathbf{D}^{-1}\mathbf{L}] \\ &= \mathbf{D}^{-1}\mathbf{L}(\mathbf{I} + \mathbf{L}'\mathbf{D}^{-1}\mathbf{L})^{-1} \end{aligned}$$

und nach Vormultiplikation mit  $\mathbf{D}^{1/2}$  gilt:

$$(\mathbf{D}^{1/2}\mathbf{S}^{-1}\mathbf{D}^{1/2})\mathbf{D}^{-1/2}\mathbf{L} = \mathbf{D}^{-1/2}\mathbf{L}(\mathbf{I} + \mathbf{L}'\mathbf{D}^{-1}\mathbf{L})^{-1}$$

Eine eindeutige Lösung für dieses Gleichungssystem ergibt sich, wenn zum einen angenommen wird dass  $\mathbf{L}'\mathbf{D}^{-1}\mathbf{L}$  diagonal ist, was durch eine entsprechende Transformationsmatrix erreicht werden kann (Bellman, 1960), und wenn zum anderen  $\mathbf{L}$  so gewählt wird, dass in  $\mathbf{C}_k = \mathbf{D}^{-1/2}\mathbf{L}$  die  $k$  größten

<sup>12</sup>In der Statistik ist es seit Fisher (1922) üblich, zwischen Parametern (den Populationswerten) und deren Schätzungen zu unterscheiden. Dabei werden zur Symbolisierung von Parametern üblicherweise griechische Buchstaben verwendet; in Ausnahmefällen (z.B. bei Großbuchstaben) aber auch lateinische Buchstaben. Schätzungen werden wahlweise durch lateinische Buchstaben mit oder ohne den Akzent ^ oder durch griechische Buchstaben mit dem ^ als Akzent dargestellt. Dabei ist es keineswegs unumstritten, wann bei der Ableitung von Funktionen zur Bestimmung der Schätzwerte die Notation zu wechseln ist. Darüber hinaus ist es gerade in der Literatur zur Faktorenanalyse durchaus üblich, gänzlich auf einen Wechsel der Notation zu verzichten (z.B. Basilevsky, 1994). Im folgenden wird der Empfehlung von Werner (1997) gefolgt, die Notation dann zu wechseln, wenn die Ableitung gleich Null gesetzt wird, weil dann eine spezifische Lösung für die Gleichung angestrebt wird.

Eigenvektoren und in  $\Lambda_k = (\mathbf{I} + \mathbf{L}'\mathbf{D}^{-1}\mathbf{L})^{-1}$  die  $k$  größten Eigenwerte von  $\mathbf{S}_{GLS}^+ = \mathbf{D}^{1/2}\mathbf{S}^{-1}\mathbf{D}^{1/2}$  stehen (Jöreskog, 1972). Ein solches  $\mathbf{L}$  ist leicht zu gewinnen, denn

$$\begin{aligned} (\mathbf{I} + \hat{\mathbf{L}}'\hat{\mathbf{D}}^{-1}\hat{\mathbf{L}}) &= \Lambda_k^{-1} & | & -\mathbf{I} \\ &= \Lambda_k^{-1} - \mathbf{I} & | & \mathbf{C}'_k\mathbf{C}_k = \mathbf{I} \\ &= (\Lambda_k^{-1} - \mathbf{I})^{1/2}\mathbf{C}'_k\mathbf{C}_k(\Lambda_k^{-1} - \mathbf{I})^{1/2} \end{aligned}$$

also

$$\hat{\mathbf{D}}^{-1/2}\hat{\mathbf{L}} = \mathbf{C}_k(\Lambda_k^{-1} - \mathbf{I})^{1/2}$$

und damit auch:

$$\hat{\mathbf{L}} = \hat{\mathbf{D}}^{1/2}\mathbf{C}_k(\Lambda_k^{-1} - \mathbf{I})^{1/2} \quad (2.19)$$

$\mathbf{L}$  lässt sich also vollständig aus den  $k$  größten Eigenwerten und Eigenvektoren von  $\mathbf{S}_{GLS}^+$  gewinnen, wobei zur Schätzung dieser Eigenwerte und Eigenvektoren auf einen der in Kapitel 2.3.2.3 (Seite 13 ff.) vorgestellten Algorithmen zurückgegriffen werden kann. Vor der Bestimmung von  $\mathbf{L}$  gemäß Gleichung 2.19 ist jedoch  $\mathbf{D}$  so zu ermitteln, dass  $F(\mathbf{S}, \boldsymbol{\Sigma}|\mathbf{S})$  ein Minimum erreicht.

Da die Minimierung von  $F(\mathbf{S}, \boldsymbol{\Sigma}|\mathbf{S})$  in Hinblick auf  $\mathbf{D}$  über die Eigenwerte und Eigenvektoren von  $\mathbf{S}_{GLS}^+$  erfolgt, soll zunächst gezeigt werden, dass sich  $F(\mathbf{S}, \boldsymbol{\Sigma}|\mathbf{S})$  als Funktion der  $m = p - k$  kleinsten Eigenwerte von  $\mathbf{S}_{GLS}^+$  schreiben lässt. Wenn nämlich in  $\Lambda_k$  bzw.  $\mathbf{C}_k$  die  $k$  größten und in  $\Lambda_m$  bzw.  $\mathbf{C}_m$  die  $m$  kleinsten Eigenwerte und Eigenvektoren von  $\mathbf{S}_{GLS}^+$  stehen, dann gilt auch

$$\begin{aligned} \mathbf{D}^{1/2}\mathbf{S}^{-1}\mathbf{D}^{1/2} &= \mathbf{C}\boldsymbol{\Lambda}\mathbf{C}' \\ &= (\mathbf{C}_k \quad \mathbf{C}_m) \begin{pmatrix} \Lambda_k & \mathbf{0} \\ \mathbf{0} & \Lambda_m \end{pmatrix} \begin{pmatrix} \mathbf{C}'_k \\ \mathbf{C}'_m \end{pmatrix} \\ &= \mathbf{C}_k\Lambda_k\mathbf{C}'_k + \mathbf{C}_m\Lambda_m\mathbf{C}'_m \end{aligned}$$

bzw.

$$\mathbf{S}^{-1} = \mathbf{D}^{-1/2} [\mathbf{C}_k\Lambda_k\mathbf{C}'_k + \mathbf{C}_m\Lambda_m\mathbf{C}'_m] \mathbf{D}^{-1/2}$$

und

$$\begin{aligned} \hat{\boldsymbol{\Sigma}} &= \mathbf{L}\mathbf{L}' + \mathbf{D} \\ &= \mathbf{D}^{1/2}\mathbf{C}_k(\Lambda_k^{-1} - \mathbf{I})^{1/2}(\Lambda_k^{-1} - \mathbf{I})^{1/2}\mathbf{C}'_k\mathbf{D}^{1/2} + \mathbf{D} \end{aligned}$$

$$\begin{aligned}
&= D^{1/2} C_k (\Lambda_k^{-1} - I) C'_k D^{1/2} + D \\
&= D^{1/2} [C_k \Lambda_k^{-1} C'_k + I - C_k C'_k] D^{1/2} \\
&= D^{1/2} [C_k \Lambda_k^{-1} C'_k + C_m C'_m] D^{1/2}
\end{aligned}$$

also

$$\begin{aligned}
I - S^{-1} \hat{\Sigma} &= I - D^{-1/2} C_k C'_k D^{1/2} - D^{-1/2} C_m \Lambda_m C'_m D^{1/2} \\
&= D^{-1/2} C_m C'_m D^{1/2} - D^{-1/2} C_m \Lambda_m C'_m D^{1/2} \\
&= D^{-1/2} [C_m (I - \Lambda_m) C'_m] D^{1/2}
\end{aligned}$$

so dass

$$\begin{aligned}
\psi(D^{1/2}) &= F(S, \Sigma | S) \\
&= \frac{1}{2} \text{Tr}[I - S^{-1} \Sigma]^2 \\
&= \frac{1}{2} \text{Tr}[D^{-1/2} [C_m (I - \Lambda_m) C'_m] D^{1/2}]^2 \\
&= \frac{1}{2} \text{Tr}[[C_m (I - \Lambda_m) C'_m] D^{1/2} D^{-1/2}]^2 \\
&= \frac{1}{2} \text{Tr}[(I - \Lambda_m) C'_m C_m]^2 \\
&= \frac{1}{2} \text{Tr}(I - \Lambda_m)^2 \\
&= \frac{1}{2} \sum_m^p (\lambda_m - 1)^2
\end{aligned}$$

gilt.

Für das Differential erster Ordnung von  $\psi(D^{1/2})$  lässt sich

$$d\psi(D^{1/2}) = \sum_m^p (\lambda_m - 1) (d\lambda_m) \quad (2.20)$$

schreiben, wobei für den Ausdruck  $(d\lambda_m)$  nach Magnus und Neudecker (1999, S. 158-159)

$$d\lambda_m = c'_j (dS_{GLS}^+) c_j$$

gilt. Umformen von  $d\lambda_m$  führt zu

$$d\lambda_m = c'_j (dS_{GLS}^+) c_j$$

$$\begin{aligned}
&= \mathbf{c}'_j d(\mathbf{D}^{1/2} \mathbf{S}^{-1} \mathbf{D}^{1/2}) \mathbf{c}_j \\
&= \mathbf{c}'_j [(\mathrm{d}\mathbf{D}^{1/2}) \mathbf{S}^{-1} \mathbf{D}^{1/2} + \mathbf{D}^{1/2} \mathbf{S}^{-1} (\mathrm{d}\mathbf{D}^{1/2})] \mathbf{c}_j \\
&= \mathbf{c}'_j [(\mathrm{d}\mathbf{D}^{1/2}) \mathbf{D}^{-1/2} \mathbf{S}_{GLS}^+ + \mathbf{S}_{GLS}^+ \mathbf{D}^{-1/2} (\mathrm{d}\mathbf{D}^{1/2})] \mathbf{c}_j \\
&= \mathbf{c}'_j (\mathrm{d}\mathbf{D}^{1/2}) \mathbf{D}^{-1/2} \mathbf{S}_{GLS}^+ \mathbf{c}_j + \mathbf{c}'_j \mathbf{S}_{GLS}^+ \mathbf{D}^{-1/2} (\mathrm{d}\mathbf{D}^{1/2}) \mathbf{c}_j \\
&= \lambda_m \mathbf{c}'_j (\mathrm{d}\mathbf{D}^{1/2}) \mathbf{D}^{-1/2} \mathbf{c}_j + \lambda_m \mathbf{c}'_j \mathbf{D}^{-1/2} (\mathrm{d}\mathbf{D}^{1/2}) \mathbf{c}_j \\
&= (2/d_j^{1/2}) \lambda_m \mathbf{c}_{jm}^2
\end{aligned} \tag{2.21}$$

sodass nach Einsetzen in Gleichung 2.20

$$d\psi(\mathbf{D}^{1/2}) = (2/d_j^{1/2}) \sum_m^p (\lambda_m^2 - \lambda_m) \mathbf{c}_{jm}^2$$

resultiert. Für den Gradienten bzw. die erste partielle Ableitung von  $\psi(\mathbf{D}^{1/2})$  nach  $d_j^{1/2}$  gilt also:

$$\begin{aligned}
\mathbf{g} &= \begin{pmatrix} \frac{\partial \psi(\mathbf{D}^{1/2})}{\partial d_1^{1/2}} \\ \vdots \\ \frac{\partial \psi(\mathbf{D}^{1/2})}{\partial d_p^{1/2}} \end{pmatrix} \\
&= 2\mathbf{D}^{-1/2} \begin{pmatrix} \sum_m^p (\lambda_m^2 - \lambda_m) \mathbf{c}_{1m}^2 \\ \vdots \\ \sum_m^p (\lambda_m^2 - \lambda_m) \mathbf{c}_{pm}^2 \end{pmatrix} \\
&= 2\mathbf{D}^{-1/2} (\sum_m^p (\lambda_m^2 - \lambda_m) \mathbf{c}_m \odot \mathbf{c}_m)
\end{aligned}$$

wobei  $\mathbf{c}_m \odot \mathbf{c}_m$  mit  $(\mathbf{c}_m \odot \mathbf{c}_m)_j = (\mathbf{c}_m)_j \cdot (\mathbf{c}_m)_j$  das Hadamard Produkt von  $\mathbf{c}_m$  ist (zur Herkunft der Bezeichnung und Anwendungsfelder des Hadamard Produkts siehe Styan, 1973).

Das Differential zweiter Ordnung von  $\psi(\mathbf{D}^{1/2})$  lautet

$$d^2\psi(\mathbf{D}^{1/2}) = \sum_m^p [(\mathrm{d}\lambda_m)^2 + (\lambda_m - 1)\mathrm{d}^2\lambda_m] \tag{2.22}$$

und für den Ausdruck  $(\mathrm{d}^2\lambda_m)$  kann in Anlehnung an Magnus und Neudecker (1999, S. 371)

$$\mathrm{d}^2\lambda_m = 2\mathbf{c}'_m (\mathrm{d}\mathbf{S}_{GLS}^+) [\lambda_m \mathbf{I} - \mathbf{S}_{GLS}^+]^- (\mathrm{d}\mathbf{S}_{GLS}^+) \mathbf{c}_m$$

$$+ \mathbf{c}'_m (d^2 \mathbf{S}_{GLS}^+) \mathbf{c}_m \quad (2.23)$$

geschrieben werden, mit  $\mathbf{S}_{GLS}^+ = \mathbf{D}^{1/2} \mathbf{S}^{-1} \mathbf{D}^{1/2} = \sum_{j=1}^p \lambda_j \mathbf{c}_j \mathbf{c}'_j$  für  $\mathbf{D} \in \mathbb{R}^{p \times p}$  sowie der nach Moore (1920) und Penrose (1955) benannten Moore-Penrose-Inverse  $\mathbf{P}_m^- = [\lambda_m \mathbf{I} - \mathbf{S}_{GLS}^+]^-$  von  $\mathbf{P}_m = [\lambda_m \mathbf{I} - \mathbf{S}_{GLS}^+] = \mathbf{C} \mathbf{\Lambda}^* \mathbf{C}'$ , wobei in der Diagonalen von  $\mathbf{\Lambda}^*$  die Eigenwerte  $\lambda_j^* = \lambda_m - \lambda_j$  von  $\mathbf{P}_m$  stehen.

Einsetzen von

$$d\mathbf{S}_{GLS}^+ = (d\mathbf{D}^{1/2}) \mathbf{D}^{-1/2} \mathbf{S}_{GLS}^+ + \mathbf{S}_{GLS}^+ \mathbf{D}^{-1/2} (d\mathbf{D}^{1/2})$$

bzw.

$$\begin{aligned} d^2 \mathbf{S}_{GLS}^+ &= (d\mathbf{D}^{1/2}) \mathbf{D}^{-1/2} (d\mathbf{S}_{GLS}^+) + (d\mathbf{S}_{GLS}^+) \mathbf{D}^{-1/2} (d\mathbf{D}^{1/2}) \\ &= (d\mathbf{D}^{1/2}) \mathbf{D}^{-1/2} (d\mathbf{D}^{1/2}) \mathbf{D}^{-1/2} \mathbf{S}_{GLS}^+ \\ &\quad + (d\mathbf{D}^{1/2}) \mathbf{D}^{-1/2} \mathbf{S}_{GLS}^+ \mathbf{D}^{-1/2} (d\mathbf{D}^{1/2}) \\ &\quad + (d\mathbf{D}^{1/2}) \mathbf{D}^{-1/2} \mathbf{S}_{GLS}^+ \mathbf{D}^{-1/2} (d\mathbf{D}^{1/2}) \\ &\quad + \mathbf{S}_{GLS}^+ \mathbf{D}^{-1/2} (d\mathbf{D}^{1/2}) \mathbf{D}^{-1/2} (d\mathbf{D}^{1/2}) \\ &= (d\mathbf{D}^{1/2}) \mathbf{D}^{-1/2} (d\mathbf{D}^{1/2}) \mathbf{D}^{-1/2} \mathbf{S}_{GLS}^+ \\ &\quad + \mathbf{S}_{GLS}^+ \mathbf{D}^{-1/2} (d\mathbf{D}^{1/2}) \mathbf{D}^{-1/2} (d\mathbf{D}^{1/2}) \\ &\quad + 2(d\mathbf{D}^{1/2}) \mathbf{D}^{-1/2} \mathbf{S}_{GLS}^+ \mathbf{D}^{-1/2} (d\mathbf{D}^{1/2}) \end{aligned}$$

in Gleichung 2.23 führt nach Umformung zu

$$\begin{aligned} d^2 \lambda_m &= 2\lambda_m^2 \mathbf{c}'_m (d\mathbf{D}^{1/2}) \mathbf{D}^{-1/2} \mathbf{P}_m^- (d\mathbf{D}^{1/2}) \mathbf{D}^{-1/2} \mathbf{c}_m \\ &\quad + 4\lambda_m \mathbf{c}'_m (d\mathbf{D}^{1/2}) \mathbf{D}^{-1/2} \mathbf{S}_{GLS}^+ \mathbf{P}_m^- (d\mathbf{D}^{1/2}) \mathbf{D}^{-1/2} \mathbf{c}_m \\ &\quad + 2\mathbf{c}'_m (d\mathbf{D}^{1/2}) \mathbf{D}^{-1/2} \mathbf{S}_{GLS}^+ \mathbf{P}_m^- \mathbf{S}_{GLS}^+ \mathbf{D}^{-1/2} (d\mathbf{D}^{1/2}) \mathbf{c}_m \\ &\quad + 2\lambda_m \mathbf{c}'_m (d\mathbf{D}^{1/2}) \mathbf{D}^{-1/2} (d\mathbf{D}^{1/2}) \mathbf{D}^{-1/2} \mathbf{c}_m \\ &\quad + 2\mathbf{c}'_m (d\mathbf{D}^{1/2}) \mathbf{D}^{-1/2} \mathbf{S}_{GLS}^+ \mathbf{D}^{-1/2} (d\mathbf{D}^{1/2}) \mathbf{c}_m \\ &= 2\mathbf{c}'_m (d\mathbf{D}^{1/2}) \mathbf{D}^{-1/2} \mathbf{K}_m \mathbf{D}^{-1/2} (d\mathbf{D}^{1/2}) \mathbf{c}_m \quad (2.24) \end{aligned}$$

mit:

$$\mathbf{K}_m = \lambda_m^2 \mathbf{P}_m^- + 2\lambda_m \mathbf{S}_{GLS}^+ \mathbf{P}_m^- + \mathbf{S}_{GLS}^+ \mathbf{P}_m^- \mathbf{S}_{GLS}^+ + \lambda_m \mathbf{I} + \mathbf{S}_{GLS}^+ \quad (2.25)$$

Da für  $\mathbf{P}_m^-$  auch

$$\begin{aligned} \mathbf{P}_m^- &= \mathbf{C} \mathbf{\Lambda}^{*-1} \mathbf{C} \\ &= \sum_{j=1}^p (\lambda_m - \lambda_j)^{-1} \mathbf{c}_j \mathbf{c}'_j \end{aligned}$$

geschrieben werden kann (siehe Searle, 1982, S. 215), gilt

$$S_{GLS}^+ P_m^- = \sum_{j=1}^p \frac{\lambda_j \mathbf{c}_j \mathbf{c}'_j}{(\lambda_m - \lambda_j)} \quad \text{und} \quad S_{GLS}^+ P_m^- S_{GLS}^+ = \sum_{j=1}^p \frac{\lambda_j^2 \mathbf{c}_j \mathbf{c}'_j}{(\lambda_m - \lambda_j)}$$

was nach Einsetzen in Gleichung 2.25 und Umformung bzw. zusammenfassen entsprechender Terme zu

$$\mathbf{K}_m = 2\lambda_m(\mathbf{c}_m \mathbf{c}'_m + \lambda_m \mathbf{P}_m^-) \quad (2.26)$$

führt (siehe Magnus & Neudecker, 1999, S. 372).

Einsetzen von Gleichung 2.21 und 2.24 in Gleichung 2.22 unter Berücksichtigung von 2.26 ergibt

$$\begin{aligned} d^2\psi(\mathbf{D}^{1/2}) &= 4 \sum_m^p \lambda_m \mathbf{T}_m [\lambda_m \mathbf{c}_m \mathbf{c}'_m + (\lambda_m - 1)(\mathbf{c}_m \mathbf{c}'_m + \lambda_m \mathbf{P}_m^-)] \mathbf{T}'_m \\ &= 4 \sum_m^p \lambda_m \mathbf{T} \mathbf{V}_m \mathbf{T}' \end{aligned}$$

mit  $\mathbf{T}_m = \mathbf{c}'_m (\mathbf{d} \mathbf{D}^{1/2}) \mathbf{D}^{-1/2}$ , sodass für die zweiten partiellen Ableitungen von  $\psi(\mathbf{D}^{1/2})$  nach  $d_j^{1/2}$

$$\frac{\partial^2 \psi(\mathbf{D}^{1/2})}{\partial d_i^{1/2} \partial d_j^{1/2}} = \frac{4}{d_i^{1/2} d_j^{1/2}} \sum_m^p \lambda_m c_{mi} c_{mj} V_{mij}$$

gilt ( $i, j = 1, \dots, p$ ) und die Hesse-Matrix

$$\begin{aligned} \mathbf{H} &= \begin{pmatrix} \frac{\partial^2 \psi(\mathbf{D}^{1/2})}{\partial d_1^{1/2} \partial d_1^{1/2}} & \frac{\partial^2 \psi(\mathbf{D}^{1/2})}{\partial d_1^{1/2} \partial d_2^{1/2}} & \cdots & \frac{\partial^2 \psi(\mathbf{D}^{1/2})}{\partial d_1^{1/2} \partial d_p^{1/2}} \\ \frac{\partial^2 \psi(\mathbf{D}^{1/2})}{\partial d_2^{1/2} \partial d_1^{1/2}} & \frac{\partial^2 \psi(\mathbf{D}^{1/2})}{\partial d_2^{1/2} \partial d_2^{1/2}} & \cdots & \frac{\partial^2 \psi(\mathbf{D}^{1/2})}{\partial d_2^{1/2} \partial d_p^{1/2}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 \psi(\mathbf{D}^{1/2})}{\partial d_p^{1/2} \partial d_1^{1/2}} & \frac{\partial^2 \psi(\mathbf{D}^{1/2})}{\partial d_p^{1/2} \partial d_2^{1/2}} & \cdots & \frac{\partial^2 \psi(\mathbf{D}^{1/2})}{\partial d_p^{1/2} \partial d_p^{1/2}} \end{pmatrix} \\ &= 4\mathbf{D}^{-1/2} \left( \sum_m^p \lambda_m \mathbf{c}_m \mathbf{c}'_m \odot \mathbf{V}_m \right) \mathbf{D}^{-1/2} \end{aligned}$$

geschrieben werden kann.

Jöreskog (1972) empfiehlt, die  $\mathbf{D}$ -Matrix so zu transformieren, dass auch Heywood Cases, d.h. Werte von  $h_j^2 > 1$  und damit auch Werte von  $d_j < 0$  schätzbar sind. Dazu ist

$$\theta_j = \log d_j$$

für  $j = 1, \dots, p$  zu bilden und der Gradient  $\mathbf{g}$  bzw. die Hesse-Matrix  $\mathbf{H}$  wie folgt anzupassen

$$\mathbf{g}^+ = \sum_m^p (\lambda_m^2 - \lambda_m) \mathbf{c}_m \odot \mathbf{c}_m$$

und:

$$\begin{aligned} \mathbf{H}^+ &= \sum_m^p \lambda_m \mathbf{c}_m \mathbf{c}_m' \odot \mathbf{V}_m + \frac{(\lambda_m^2 - \lambda_m) \mathbf{c}_m \odot \mathbf{c}_m}{2} \quad |\lambda_m, \dots, \lambda_p \approx 1 \\ &\approx \left( \sum_m^p \mathbf{c}_m \mathbf{c}_m' \right)^2 \end{aligned}$$

Mit Hilfe der ersten und zweiten Ableitung kann für den gesuchten Vektor  $\boldsymbol{\delta} = (d_1, \dots, d_p)'$  (und für  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)'$  entsprechend) die Taylorsche Formel

$$\psi(\boldsymbol{\delta}_n + \mathbf{p}_v) \approx \psi(\boldsymbol{\delta}_n) + \mathbf{g}_n' \mathbf{p}_v + \frac{1}{2} \mathbf{p}_v' \mathbf{H}_n \mathbf{p}_v$$

gebildet werden (Forster, 1999; Gill, Murray & Wright, 1981), wobei  $\psi(\boldsymbol{\delta}_n) = \psi(\mathbf{D}^{1/2} |\boldsymbol{\delta}_n)$ ,  $n = 0, \dots, w$  mit  $w$  als Anzahl der Iterationsschritte,  $v = n + 1$ ,  $\boldsymbol{\delta}_n$  der Fehlervektor nach der  $n$ -ten Iteration,  $\mathbf{g}_n$  der dazugehörige Gradient und  $\mathbf{H}_n$  die entsprechende Hesse-Matrix ist. Am gesuchten Punkt, also am Vektor  $\boldsymbol{\delta}_0$  der Gleichung 2.16 minimiert, fallen  $\boldsymbol{\delta}_{n+1} = \boldsymbol{\delta}_n + \mathbf{p}_v$  und  $\boldsymbol{\delta}_n$  zusammen, d.h.  $\mathbf{p}_v$  ist so zu bestimmen, dass der Ausdruck

$$\Phi(\mathbf{p}_v) = \mathbf{g}_n' \mathbf{p}_v + \frac{1}{2} \mathbf{p}_v' \mathbf{H}_n \mathbf{p}_v$$

minimiert wird. Dies ist etwa dann der Fall, wenn

$$\begin{aligned} \mathbf{H}_n \mathbf{p}_v &= -\mathbf{g}_n \\ \mathbf{p}_v &= -\mathbf{H}_n^{-1} \mathbf{g}_n \end{aligned}$$

gilt, was als Newton-Richtung bezeichnet wird.

Da die exakte Bestimmung der  $\mathbf{H}$ -Matrix häufig schwer ist, kann diese auch approximiert werden, was zur Quasi-Newton Methode führt. Dabei werden drei Approximationsvarianten unterschieden:



1. der Powell-Symmetrisch-Broyden Algorithmus (PSB; Broyden, 1967),
2. der Davidon-Fletcher-Powell Algorithmus (DFP; Davidon, 1959; Fletcher & Powell, 1963), und
3. der Broyden-Fletcher-Goldfarb-Shanno Algorithmus (BFGS; Broyden, 1970a, 1970b; Fletcher, 1970; Goldfarb, 1970; Shanno, 1970).

In allen drei Verfahren wird die Matrix  $\mathbf{H}_n$  durch eine Matrix  $\mathbf{B}_n$  ersetzt, zu deren Bestimmung nur noch die ersten partiellen Ableitungen benötigt werden. So gilt im PSB Algorithmus

$$\mathbf{B}_{n+1} = \mathbf{B}_n + \frac{1}{\mathbf{p}'_v \mathbf{p}_v} \left( (\mathbf{y}_n - \mathbf{B}_n \mathbf{p}_v) \mathbf{p}'_v + \mathbf{p}_v (\mathbf{y}_n - \mathbf{B}_n \mathbf{p}_v)' \right) - \frac{(\mathbf{y}_n - \mathbf{B}_n \mathbf{p}_v)' \mathbf{p}_v}{(\mathbf{p}'_v \mathbf{p}_v)^2} \mathbf{p}_v \mathbf{p}'_v$$

mit

$$\mathbf{y}_n = \left( \frac{\partial \psi(\boldsymbol{\delta}_{n+1} + \mathbf{p}_v)}{\partial (\boldsymbol{\delta}_{n+1} + \mathbf{p}_v)} - \frac{\partial \psi(\boldsymbol{\delta}_{n+1})}{\partial \boldsymbol{\delta}_{n+1}} \right)$$

beim DFP Verfahren

$$\mathbf{B}_{n+1} = \mathbf{B}_n + \frac{1}{\mathbf{y}'_n \mathbf{p}_v} \mathbf{y}_n \mathbf{y}'_n - \frac{1}{\mathbf{p}'_v \mathbf{B}_n \mathbf{p}_v} (\mathbf{B}_n \mathbf{p}_v \mathbf{p}'_v \mathbf{B}_n) + (\mathbf{p}'_v \mathbf{B}_n \mathbf{p}_v) \mathbf{w} \mathbf{w}'$$

mit

$$\mathbf{w} = \frac{1}{\mathbf{y}'_n \mathbf{p}_v} \mathbf{y}_n - \frac{1}{\mathbf{p}'_v \mathbf{B}_n \mathbf{p}_v} \mathbf{B}_n \mathbf{p}_v$$

und im BFGS ist

$$\mathbf{B}_{n+1} = \mathbf{B}_n + \frac{1}{\mathbf{y}'_n \mathbf{p}_v} \mathbf{y}_n \mathbf{y}'_n - \frac{1}{\mathbf{p}'_v \mathbf{B}_n \mathbf{p}_v} (\mathbf{B}_n \mathbf{p}_v \mathbf{p}'_v \mathbf{B}_n)$$

Auf die statistischen Eigenschaften dieser Matrizen und die Verbindungen zwischen ihnen geht Kasper (2006) ein. Ist  $\mathbf{B}_n$  nach einem der Algorithmen bestimmt, gilt für  $\mathbf{p}_v$ :

$$\begin{aligned} \mathbf{B}_n \mathbf{p}_v &\approx -\mathbf{g}_n \\ \mathbf{p}_v &\approx -\mathbf{B}_n^{-1} \mathbf{g}_n \end{aligned}$$

Bei großen Datensätzen kann  $\mathbf{p}_v$  auch über die Conjugate-Gradient Methode bestimmt werden

$$\mathbf{p}_{v+1} \approx -\mathbf{g}_{n+1} + \beta_n \mathbf{p}_v$$

wobei nach Fletcher und Reeves (1964) für  $\beta_n$

$$\beta_n = \frac{\mathbf{g}'_{n+1} \mathbf{g}_{n+1}}{\mathbf{g}'_n \mathbf{g}_n}$$

gilt. Eine andere Spezifikation von  $\beta_n$  wurde von Polak (1971) vorgeschlagen

$$\beta_n^* = \frac{(\mathbf{g}_{n+1} - \mathbf{g}_n)' \mathbf{g}_{n+1}}{\mathbf{g}'_n \mathbf{g}_n}$$

wobei  $\beta_n^*$  vor allem in den ersten Iterationsschritten effizientere Lösungen als  $\beta_n$  für  $\mathbf{p}_v$  liefert.

Unabhängig davon, wie  $\mathbf{p}_v$  geschätzt wird, folgt die Iteration zum Auffinden von  $\delta_o$  nach Festlegung von Startwerten für  $\delta_o$  und Ermittlung von  $\psi(\delta_o)$ ,  $\mathbf{g}_o$  und  $\mathbf{H}_o$  bzw.  $\mathbf{B}_o$  in vier Schritten:

1. Bestimmung von  $\mathbf{p}_v$  auf der Basis von  $\delta_n$ ,  $\mathbf{g}_n$  und  $\mathbf{H}_n$  bzw.  $\mathbf{B}_n$ .
2. Ermittlung von  $\delta_{n+1} = \delta_n + \mathbf{p}_v$ .
3. Bestimmung von  $\psi(\delta_{n+1})$  und Überprüfung ob  $\psi(\delta_n) - \psi(\delta_{n+1}) < \epsilon$  gilt.
4. Wenn  $\psi(\delta_n) - \psi(\delta_{n+1}) < \epsilon$  gilt, dann ist  $\delta_o = \delta_{n+1}$  andernfalls wird die Zuweisung  $\delta_n \leftarrow \delta_{n+1}$  gesetzt und die Iteration vom ersten Schritt erneut durchlaufen.

Als Startvektor für  $\delta_o$  empfiehlt Jöreskog (1967)

$$\delta_o = \left(1 - \frac{k}{2p}\right) \begin{pmatrix} 1/s^{11} \\ \vdots \\ 1/s^{pp} \end{pmatrix}$$

und als Startvektor für  $\theta_o$  (s. Jöreskog, 1972)

$$\theta_o = \log \left[ \left(1 - \frac{k}{2p}\right) \begin{pmatrix} 1/s^{11} \\ \vdots \\ 1/s^{pp} \end{pmatrix} \right]$$

wobei  $s^{jj}$  in beiden Fällen das  $j$ -te Diagonalelement von  $\mathbf{S}^{-1}$  ist. Für  $\epsilon$  kann theoretisch eine beliebig kleine positive Zahl gewählt werden, wobei diese in R (R Development Core Team, 2010) beispielsweise bei  $\epsilon \approx 0.00000001$  liegt. Ein leicht abgewandeltes Vorgehen zum Auffinden der gesuchten Lösung  $\boldsymbol{\delta}_\circ$  erfolgt über die Levenberg-Marquadt Methode (Levenberg, 1944). Demnach ist  $\mathbf{p}_v$  eine Lösung des Eigenwertproblems

$$(\mathbf{H}_n + \lambda \mathbf{I}) \mathbf{p}_v = -\mathbf{g}_n \quad (2.27)$$

und der Levenberg-Marquadt Algorithmus ist nach Festlegung von Startwerten für  $\boldsymbol{\delta}_0$  (z.B. nach der obigen Formel) bzw.  $\lambda_0$  und Ermittlung von  $\mathbf{g}_0$ ,  $\mathbf{H}_0$  sowie  $\mathbf{H}_0 + \lambda_0 \mathbf{I}$  wie folgt definiert:

1. Überprüfen ob  $(\mathbf{H}_n + \lambda_n \mathbf{I})$  positiv definit ist, wenn nicht, mit  $\lambda_n \leftarrow 4\lambda_n$  erneut beginnen sonst weiter mit Schritt drei.
2. Auflösen der Gleichung 2.27 nach  $\mathbf{p}_v$ .
3. Überprüfen der Minimierungsbedingung  $r_n$ :

Wenn  $r_n < 0.25$ , dann gilt  $\lambda_{n+1} \leftarrow 4\lambda_n$ ,  $\boldsymbol{\delta}_{n+1} \leftarrow \boldsymbol{\delta}_n + \mathbf{p}_v$  und weiter mit Schritt eins.

Wenn  $r_n > 0.75$ , dann gilt  $\lambda_{n+1} \leftarrow \lambda_n/2$ ,  $\boldsymbol{\delta}_{n+1} \leftarrow \boldsymbol{\delta}_n + \mathbf{p}_v$  und weiter mit Schritt eins.

Wenn  $0.25 < r_n < 0.75$ , dann gilt  $\lambda_{n+1} \leftarrow \lambda_n$ ,  $\boldsymbol{\delta}_{n+1} \leftarrow \boldsymbol{\delta}_n + \mathbf{p}_v$  und weiter mit Schritt eins.

Wenn  $r_n \leq 0$  dann gilt  $\boldsymbol{\delta}_{n+1} = \boldsymbol{\delta}_n = \boldsymbol{\delta}_\circ$  und die gesuchte Lösung ist gefunden.

Zu Beginn des Levenberg-Marquadt Algorithmus wird ein beliebiges  $\lambda_1 > 0$  gewählt, und für  $r_n$  gilt:

$$r_n = \frac{\psi(\boldsymbol{\delta}_n) - \psi(\boldsymbol{\delta}_{n+1})}{\psi(\boldsymbol{\delta}_n) - \Phi(\mathbf{p}_v)}$$

### Schätzung von $\mathbf{L}$ und $\mathbf{D}$ über die Maximum-Likelihood-Methode

Voraussetzung für die Schätzung von  $\mathbf{L}$  und  $\mathbf{D}$  über die Maximum-Likelihood-Methode ist, dass  $\mathbf{y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  und  $\text{rg}(\boldsymbol{\Sigma}) = p$ . Unter diesen Voraussetzungen ist die empirische Kovarianzmatrix  $\mathbf{S}$

$$\mathbf{S} = \frac{\sum_{i=1}^n (\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{y}_i - \bar{\mathbf{y}})'}{n-1} \quad \text{mit} \quad \bar{\mathbf{y}} = \frac{1}{n} \sum_{i=1}^n \mathbf{y}_i \quad | \quad \mathbf{y}'_i = (y_{i1} \quad \cdots \quad y_{ip})$$

Wishart-Verteilt mit  $(n-1)\mathbf{S} \sim W(\boldsymbol{\Sigma}, n-1)$  und der Dichtefunktion

$$f(\mathbf{S}|\mathbf{L}, \mathbf{D}, \mathbf{y}) = \frac{|\mathbf{S}|^{(n-p-2)/2} \exp[-((n-1)/2)\text{tr}(\mathbf{S}\boldsymbol{\Sigma}^{-1})]}{2^{(n-1)p/2} |\boldsymbol{\Sigma}|^{(n-1)/2} \Gamma_p((n-1)/2)} \quad (2.28)$$

wobei  $\boldsymbol{\Sigma}$  bzw.  $\mathbf{L}$  und  $\mathbf{D}$  so zu bestimmen sind, dass die Funktion  $f(\mathbf{S}|\mathbf{L}, \mathbf{D}, \mathbf{y})$  maximiert wird.

Zum auffinden des Maximums der Funktion  $f(\mathbf{S}|\mathbf{L}, \mathbf{D}, \mathbf{y})$  sind die ersten partiellen Ableitungen der Funktion  $f(\mathbf{S}|\mathbf{L}, \mathbf{D}, \mathbf{y})$  nach  $\mathbf{L}$  bzw.  $\mathbf{D}$  zu bilden. Da Summen allerdings leichter differenzierbar sind als Produkte, wird Gleichung 2.28 vor der partiellen Ableitung logarithmiert

$$\begin{aligned} \ln f(\mathbf{S}|\mathbf{L}, \mathbf{D}, \mathbf{y}) &= (n-p-2)/2 \ln|\mathbf{S}| - ((n-1)/2)\text{tr}(\mathbf{S}\boldsymbol{\Sigma}^{-1}) \\ &\quad - ((n-1)/2) \ln|\boldsymbol{\Sigma}| - a \end{aligned}$$

und weil  $a = -((n-1)p/2) \ln 2 - \ln \Gamma_p((n-1)/2)$  nicht von  $\mathbf{L}$  oder  $\mathbf{D}$  abhängt, wird es nicht weiter betrachtet. Des Weiteren ist ersichtlich, dass das Maximum von  $\ln f(\mathbf{S}|\mathbf{L}, \mathbf{D}, \mathbf{y})$  dem Minimum von

$$\gamma(\mathbf{L}, \mathbf{D}) = \ln|\boldsymbol{\Sigma}| + \text{tr}(\mathbf{S}\boldsymbol{\Sigma}^{-1}) \quad (2.29)$$

entspricht. Im Folgenden wird also das Minimum der Funktion  $\gamma(\mathbf{L}, \mathbf{D})$  gesucht.

Die erste partielle Ableitungen von  $\gamma(\mathbf{L}, \mathbf{D})$  nach  $\mathbf{L}$  lautet

$$\frac{\partial \gamma(\mathbf{L}, \mathbf{D})}{\partial \mathbf{L}} = \frac{\partial \ln|\boldsymbol{\Sigma}|}{\partial \mathbf{L}} + \frac{\partial \text{tr}(\mathbf{S}\boldsymbol{\Sigma}^{-1})}{\partial \mathbf{L}} \quad (2.30)$$

wobei für den ersten Ausdruck nach Lawley und Maxwell (1963)

$$\begin{aligned} \frac{\partial \ln|\boldsymbol{\Sigma}|}{\partial \mathbf{L}} &= \begin{pmatrix} \frac{\partial \ln|\boldsymbol{\Sigma}|}{\partial l_{11}} & \cdots & \frac{\partial \ln|\boldsymbol{\Sigma}|}{\partial l_{1k}} \\ \vdots & & \vdots \\ \frac{\partial \ln|\boldsymbol{\Sigma}|}{\partial l_{p1}} & \cdots & \frac{\partial \ln|\boldsymbol{\Sigma}|}{\partial l_{pk}} \end{pmatrix} \\ &= \begin{pmatrix} \sum_{i,j} \frac{\partial \ln|\boldsymbol{\Sigma}|}{\partial \sigma_{ij}} \frac{\partial \sigma_{ij}}{\partial l_{11}} & \cdots & \sum_{i,j} \frac{\partial \ln|\boldsymbol{\Sigma}|}{\partial \sigma_{ij}} \frac{\partial \sigma_{ij}}{\partial l_{1k}} \\ \vdots & & \vdots \\ \sum_{i,j} \frac{\partial \ln|\boldsymbol{\Sigma}|}{\partial \sigma_{ij}} \frac{\partial \sigma_{ij}}{\partial l_{p1}} & \cdots & \sum_{i,j} \frac{\partial \ln|\boldsymbol{\Sigma}|}{\partial \sigma_{ij}} \frac{\partial \sigma_{ij}}{\partial l_{pk}} \end{pmatrix} \\ &= \frac{1}{|\boldsymbol{\Sigma}|} \begin{pmatrix} 2\text{adj}(\boldsymbol{\Sigma})_{1,l_1} & \cdots & 2\text{adj}(\boldsymbol{\Sigma})_{1,l_k} \\ \vdots & & \vdots \\ 2\text{adj}(\boldsymbol{\Sigma})_{p,l_1} & \cdots & 2\text{adj}(\boldsymbol{\Sigma})_{p,l_k} \end{pmatrix} \end{aligned}$$

$$\begin{aligned}
 &= 2 (\text{adj}(\mathbf{\Sigma}) \setminus |\mathbf{\Sigma}|) \mathbf{L} \\
 &= 2\mathbf{\Sigma}^{-1}\mathbf{L}
 \end{aligned}$$

geschrieben werden kann, und für den zweiten Ausdruck:

$$\begin{aligned}
 \frac{\partial \text{tr}(\mathbf{S}\mathbf{\Sigma}^{-1})}{\partial \mathbf{L}} &= \begin{pmatrix} \frac{\partial \text{tr}(\mathbf{S}\mathbf{\Sigma}^{-1})}{\partial l_{11}} & \cdots & \frac{\partial \text{tr}(\mathbf{S}\mathbf{\Sigma}^{-1})}{\partial l_{1k}} \\ \vdots & \vdots & \vdots \\ \frac{\partial \text{tr}(\mathbf{S}\mathbf{\Sigma}^{-1})}{\partial l_{p1}} & \cdots & \frac{\partial \text{tr}(\mathbf{S}\mathbf{\Sigma}^{-1})}{\partial l_{pk}} \end{pmatrix} \\
 &= \begin{pmatrix} \text{tr} \left( \mathbf{S} \frac{\partial \mathbf{\Sigma}^{-1}}{\partial l_{11}} \right) & \cdots & \text{tr} \left( \mathbf{S} \frac{\partial \mathbf{\Sigma}^{-1}}{\partial l_{1k}} \right) \\ \vdots & \vdots & \vdots \\ \text{tr} \left( \mathbf{S} \frac{\partial \mathbf{\Sigma}^{-1}}{\partial l_{p1}} \right) & \cdots & \text{tr} \left( \mathbf{S} \frac{\partial \mathbf{\Sigma}^{-1}}{\partial l_{pk}} \right) \end{pmatrix} \\
 &= \begin{pmatrix} \text{tr} \left( -\mathbf{S}\mathbf{\Sigma}^{-1} \frac{\partial \mathbf{\Sigma}}{\partial l_{11}} \mathbf{\Sigma}^{-1} \right) & \cdots & \text{tr} \left( -\mathbf{S}\mathbf{\Sigma}^{-1} \frac{\partial \mathbf{\Sigma}}{\partial l_{1k}} \mathbf{\Sigma}^{-1} \right) \\ \vdots & \vdots & \vdots \\ \text{tr} \left( -\mathbf{S}\mathbf{\Sigma}^{-1} \frac{\partial \mathbf{\Sigma}}{\partial l_{p1}} \mathbf{\Sigma}^{-1} \right) & \cdots & \text{tr} \left( -\mathbf{S}\mathbf{\Sigma}^{-1} \frac{\partial \mathbf{\Sigma}}{\partial l_{pk}} \mathbf{\Sigma}^{-1} \right) \end{pmatrix} \\
 &= -2\mathbf{\Sigma}^{-1}\mathbf{S}\mathbf{\Sigma}^{-1}\mathbf{L}
 \end{aligned}$$

Einsetzen der beiden letzten Ausdrücke in Gleichung 2.30 führt zu:

$$\frac{\partial \gamma(\mathbf{L}, \mathbf{D})}{\partial \mathbf{L}} = \mathbf{\Sigma}^{-1}\mathbf{L} - \mathbf{\Sigma}^{-1}\mathbf{S}\mathbf{\Sigma}^{-1}\mathbf{L}$$

Diese Gleichung ist Null zu setzen

$$(\hat{\mathbf{\Sigma}} - \mathbf{S})\hat{\mathbf{\Sigma}}^{-1}\hat{\mathbf{L}} = \mathbf{0} \tag{2.31}$$

und  $\mathbf{L}$  unter der Berücksichtigung von  $\hat{\mathbf{M}} = \hat{\mathbf{L}}'\hat{\mathbf{D}}^{-1}\hat{\mathbf{L}}$  zu schätzen.

Da nach Basilevsky (1994) für  $\mathbf{\Sigma}^{-1}\mathbf{L}$  auch  $\mathbf{D}^{-1}\mathbf{L}(\mathbf{I} + \mathbf{M})^{-1}$  geschrieben werden kann, ändert sich Gleichung 2.31 in die Form

$$(\hat{\mathbf{\Sigma}} - \mathbf{S})\hat{\mathbf{D}}^{-1}\hat{\mathbf{L}} = \mathbf{0}$$

und nach einsetzen von  $\hat{\mathbf{\Sigma}} = \hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\mathbf{D}}$  in diese Gleichung gilt

$$\hat{\mathbf{L}}\hat{\mathbf{L}}'\hat{\mathbf{D}}^{-1}\hat{\mathbf{L}} + \hat{\mathbf{L}} - \mathbf{S}\hat{\mathbf{D}}^{-1}\hat{\mathbf{L}} = \mathbf{0}$$

bzw.

$$[\hat{\mathbf{D}}^{-1/2}\mathbf{S}\hat{\mathbf{D}}^{-1/2} - (\hat{\mathbf{M}} + \mathbf{I})]\hat{\mathbf{D}}^{-1/2}\hat{\mathbf{L}} = \mathbf{0}$$

$$[S_{ML}^+ - (\hat{M} + I)]\hat{D}^{-1/2}\hat{L} = \mathbf{0}$$

mit  $S_{ML}^+ = \hat{D}^{-1/2}S\hat{D}^{-1/2}$ . Wie zu erkennen, stehen in  $\hat{M} + I$  die Eigenwerte von  $S_{ML}^+$  und in  $\hat{D}^{-1/2}\hat{L}$  die dazugehörigen Eigenvektoren.

Wird angenommen, dass in  $\Lambda_k = (I + \hat{M})$  die  $k$  größten Eigenwerte von  $S_{ML}^+$  stehen, dann gilt

$$\begin{aligned} (I + \hat{L}'\hat{D}^{-1}\hat{L}) &= \Lambda_k & | & -I \\ &= \Lambda_k - I & | & C_k' C_k = I \\ &= (\Lambda_k - I)^{1/2} C_k' C_k (\Lambda_k - I)^{1/2} \end{aligned}$$

bzw.

$$\hat{D}^{-1/2}\hat{L} = C_k(\Lambda_k - I)^{1/2}$$

und damit auch:

$$\hat{L} = \hat{D}^{1/2} C_k (\Lambda_k - I)^{1/2} \quad (2.32)$$

$L$  lässt sich also vollständig aus den Eigenwerten und Eigenvektoren von  $S_{ML}^+$  gewinnen. Vor der Schätzung von  $L$  gemäß Gleichung 2.32 ist jedoch  $D$  so zu bestimmen, dass  $\gamma(L, D)$  ein Minimum erreicht.

Für die erste partielle Ableitung von  $\gamma(L, D)$  nach  $D$  gilt gemäß Lawley und Maxwell (1963, S. 13)

$$\frac{\partial \gamma(L, D)}{\partial D} = \text{diag} [\Sigma^{-1} - \Sigma^{-1} S \Sigma^{-1}]$$

was nach Ersetzen von  $\Sigma^{-1}$  durch Gleichung 2.18 (S. 37) und Berücksichtigung von Gleichung 2.31 auch wie folgt geschrieben werden kann:

$$\begin{aligned} \frac{\partial \gamma(L, D)}{\partial D} &= \text{diag} [D^{-1} (\Sigma - S)] \\ &= \text{diag} [D^{-1} (LL' + D - S)] \\ &= \text{diag} [D^{-1/2} C_k (\Lambda_k - I) C_k' D^{-1/2} + D^{-1} - D^{-1} S D^{-1}] \\ &= \text{diag} [D^{-1/2} [C_k (\Lambda_k - I) C_k' + I - D^{-1/2} S D^{-1/2}] D^{-1/2}] \\ &= T \end{aligned}$$

Demnach gilt für den Gradienten von  $\gamma(L, D)$  nach  $D$

$$g = \begin{pmatrix} \frac{\partial \gamma(L, D)}{\partial d_1} \\ \vdots \\ \frac{\partial \gamma(L, D)}{\partial d_p} \end{pmatrix} = \begin{pmatrix} t_{11} \\ \vdots \\ t_{pp} \end{pmatrix}$$

mit (Jöreskog, 1967, S. 450):

$$t_{jj} = 1/d_{jj} \left[ \sum_{z=1}^k (\lambda_z - 1) c_{jz}^2 + 1 - s_{jj}/d_{jj} \right]$$

Mit Hilfe des Gradienten  $\mathbf{g}$  kann über die Quasi-Newton Methode oder über die Conjugate-Gradient Methode der Vektor  $\boldsymbol{\delta}_0$  gefunden werden, welcher die Funktion  $\gamma(\mathbf{L}, \mathbf{D})$  minimiert. Als Startvektor für  $\boldsymbol{\delta}_0$  kann dabei auf den bereits oben aufgeführten und von Jöreskog (1967) empfohlenen Vektor zurückgegriffen werden. Ansonsten erfolgt die Iteration wie beim GLS-Schätzer mit der Ausnahme, dass nicht  $\psi(\boldsymbol{\delta}_n)$  sondern  $\gamma(\boldsymbol{\delta}_n) = \gamma(\mathbf{L}, \mathbf{D}|\boldsymbol{\delta}_n)$  minimiert wird. Wenn Heywood Cases auftauchen, d.h. wenn während der Iteration einer bzw. einige  $\hat{d}_j = 0$  werden, dann ist  $\hat{d}_j$  auf einen beliebig kleinen Wert  $\epsilon > 0$  zu fixieren und die entsprechenden Komponenten in  $\mathbf{g}_n$  sowie die Werte außerhalb der Diagonalen der  $j$ -ten Zeile und der  $j$ -ten Spalte der Matrix  $\mathbf{B}^{-1}$  Null zu setzen.

### Schätzung von $\mathbf{F}$

Nach der Schätzung von  $\mathbf{L}$  und  $\mathbf{D}$  gilt für  $\mathbf{F}$  nach Fahrmeir, Hamerle und Tutz (1996):

$$\hat{\mathbf{F}} = (\mathbf{Y} - \mathbf{M})\hat{\mathbf{D}}^{-1}\hat{\mathbf{L}}(\hat{\mathbf{L}}'\hat{\mathbf{D}}^{-1}\hat{\mathbf{L}})^{-1} \quad (2.33)$$

## 2.3.4 Hauptfaktorenanalyse

### 2.3.4.1 Strukturmodell

Die Hauptfaktorenanalyse basiert auf demselben Strukturmodell wie die explorative Faktorenanalyse, d.h. es gilt

$$\mathbf{Z} = \mathbf{F}\mathbf{L}' + \mathbf{E} \quad (2.34)$$

mit  $\mathbf{Z}$   $n \times p$ -Matrix der Aufgabenergebnisse  $z_{ij}$   
 $\mathbf{F}$   $n \times k$ -Matrix der gemeinsamen Faktoren  $f_{ik}$   
 $\mathbf{L}$   $p \times p$ -Matrix der Gewichte  $l_{jk}$   
 $\mathbf{E}$   $n \times p$ -Matrix der spezifischen Faktoren  $e_{ik}$ .

Im Unterschied zur explorativen Faktorenanalyse wird allerdings nicht die Kovarianzmatrix  $\boldsymbol{\Sigma}$  analysiert sondern die Korrelationsmatrix:

$$\begin{aligned} \mathbf{Z}'\mathbf{Z} &= \mathbf{R} \\ &= \mathbf{L}\mathbf{L}' + \mathbf{D} \end{aligned}$$

### 2.3.4.2 Voraussetzungen

Die notwendigen Voraussetzungen der Hauptfaktorenanalyse lauten:

1.  $E(\mathbf{e}) = \mathbf{0}$ .
2.  $\text{cov}(\mathbf{e}) = E(\mathbf{e}\mathbf{e}') = \mathbf{D} = \text{diag}\{e_1^2, \dots, e_p^2\}$ .
3.  $\text{cov}(\mathbf{f}) = E(\mathbf{f}\mathbf{f}') = \mathbf{I}$ .
4.  $\text{cov}(\mathbf{f}\mathbf{e}') = E(\mathbf{f}\mathbf{e}') = \mathbf{0}$ .
5.  $\text{rg}(\mathbf{Z}) = \text{rg}(\mathbf{R}) = p$ .
6.  $\forall_j s_j^2 \neq 0$ .

Der Stellenwert dieser Voraussetzungen wurde bereits bei der Besprechung der beiden vorherigen Modelle erläutert. Da  $\mathbf{R}$  in der Regel über die Produkt-Moment-Korrelation bestimmt wird (Bravais, 1846), muss darüber hinaus angenommen werden, dass

7.  $\mathbf{Z}$ ,  $\mathbf{F}$ ,  $\mathbf{L}$  und  $\mathbf{E}$  intervallskaliert

sind.

### 2.3.4.3 Parameterschätzung

Zur Schätzung von  $\mathbf{L}$  und  $\mathbf{F}$  wird auf den Algorithmus der Hauptkomponentenmethode zurückgegriffen. Allerdings wird die Eigenwertzerlegung nicht auf  $\mathbf{R}$  sondern auf

$$\mathbf{R}_a = \mathbf{R} - \hat{\mathbf{D}}$$

angewendet, wobei  $\hat{\mathbf{D}}$  eine Schätzung für  $\mathbf{D}$  ist. Zur Schätzung von  $\mathbf{D}$  wird auf die Beziehung  $h_j^2 = 1 - e_j^2$  zurückgegriffen (s. Gleichung 2.14), d.h. es werden zunächst die Kommunalitäten  $h_j^2$  geschätzt um dann  $e_j^2 = 1 - h_j^2$  zu gewinnen (Methoden zur Schätzung der Kommunalitäten werden bei Harman (1967) vorgestellt). Mit  $\hat{\mathbf{D}}$  als Schätzung gilt für  $\mathbf{L}$  und  $\mathbf{F}$

$$\begin{aligned} \hat{\mathbf{L}} &= \mathbf{C}\mathbf{\Lambda}^{1/2} \\ \hat{\mathbf{F}} &= \mathbf{Z}\mathbf{T}\mathbf{\Lambda}^{-1/2} \end{aligned} \tag{2.35}$$

wenn in  $\mathbf{\Lambda}$  die  $k$  größten Eigenwerte von  $\mathbf{R}_a$  stehen und in  $\mathbf{C}$  die dazugehörigen Eigenvektoren (s. Fahrmeir et al., 1996,  $k$  selber kann nach einer der beschriebenen Methoden gewonnen werden).



## 2.4 Rotationsproblematik

Bei den Ausführungen zur Hauptkomponentenmethode, zur explorativen Faktorenanalyse und zur Hauptfaktorenanalyse wurde davon ausgegangen, dass die Matrix  $\mathbf{L}$  so verwendet wird, wie sie geschätzt wurde. Wie aber leicht einsichtig ist, ist die Matrix  $\mathbf{L}$  nur eindeutig bis auf orthogonale Transformationen. Wenn nämlich  $\mathbf{Q}$  eine orthogonale Matrix ist und  $\mathbf{B} = \mathbf{LQ}$  gilt, dann folgt für das Hauptkomponentenmodell

$$\begin{aligned}\mathbf{Z} &= \mathbf{FL}' \\ &= \mathbf{FQ}'^{-1}\mathbf{Q}'\mathbf{L}' \\ &= \mathbf{F}^\perp\mathbf{B}'\end{aligned}$$

und für das explorative Faktorenmodell

$$\begin{aligned}\mathbf{Y} &= \mathbf{M} + \mathbf{FL}' + \mathbf{E} \\ &= \mathbf{M} + \mathbf{FQ}'^{-1}\mathbf{Q}'\mathbf{L}' + \mathbf{E} \\ &= \mathbf{M} + \mathbf{F}^\perp\mathbf{B}' + \mathbf{E}\end{aligned}$$

d.h. die Transformation von  $\mathbf{L}$  wirkt sich nur auf  $\mathbf{F}$  aus, nicht aber auf die Zerlegung von  $\mathbf{Z}$  bzw.  $\mathbf{Y}$ . Da eine Transformation von  $\mathbf{L}$  darüber hinaus keine Veränderung der Kovarianzmatrix  $\mathbf{\Sigma}$  zur Folge hat (sich der Fit-Wert also nicht durch die Transformation von  $\mathbf{L}$  verändert), können die Transformationen auch nach der Schätzung des nicht-transformierten  $\mathbf{L}$  durchgeführt werden (Kasper, 2006, S. 117). Warum aber sollte  $\mathbf{L}$  überhaupt transformiert werden?

Die Antwort auf diese Frage ist relativ einfach: Durch eine Transformation von  $\mathbf{L}$  kann erreicht werden, dass jede Variable nur mit einem Faktor hoch korreliert und jeder Faktor mit einigen Variablen hoch und mit allen anderen Variablen niedrig korreliert, d.h. in  $\mathbf{L}$  sind spaltenweise nur einige Einträge sehr hoch und zeilenweise nur ein Eintrag hoch. Schematisch stellt sich dieses nach Thurstone (1947) benannte Einfachstrukturkriterium im Beispiel von  $p = 6$  Variablen und  $k = 3$  Faktoren wie folgt da

$$\mathbf{L} = \begin{pmatrix} \times & & & & & \\ \times & & & & & \\ & \times & & & & \\ & \times & & & & \\ & & \times & & & \\ & & & \times & & \\ & & & & \times & \\ & & & & & \times \end{pmatrix}$$

wobei  $x$  für eine hohe Ladung steht und niedrige Ladungen zur Wahrung der Übersicht weggelassen wurden. Durch das Einfachstrukturkriterium wird also die inhaltliche Interpretation der Faktoren erleichtert: Die Faktoren können als Surrogat der sie bezeichnenden Variablen verstanden werden.

Zur Bestimmung der Transformationsmatrix  $\mathbf{Q}$  stehen eine Reihe verschiedener Verfahren zur Verfügung (einen Überblick bietet z.B. Bortz, 1997), die sich in zwei Klassen einteilen lassen: Die Transformationen der einen Klasse erhalten die orthogonale Struktur der Faktoren, d.h. nach der Transformation von  $\mathbf{L}$  gilt (weiterhin)  $\text{cov}(\mathbf{f}^\perp) = \mathbf{I}$ , weswegen diese Verfahren auch orthogonale Rotationstechniken bezeichnet werden; die Verfahren der anderen Klasse, die sogenannten *obliquen* Rotationstechniken, führen zu korrelierten Faktoren, d.h. nach der Transformation von  $\mathbf{L}$  gilt  $\text{cov}(\mathbf{f}^\perp) = \mathbf{O} \neq \mathbf{I}$ .

Zu den orthogonalen Rotationstechniken zählen die Quartimaxmethode nach Carroll (1953), Ferguson (1954) bzw. Neuhaus und Wrigley (1954) sowie das Varimaxverfahren nach Kaiser (1958). Während die Quartimaxmethode auf eine Maximierung der quadrierten Zeilensummen von  $\mathbf{L}$  abzielt, d.h. auf eine Maximierung der Varianzaufklärung in den Variablen durch Berücksichtigung mehrerer Faktoren, werden mit der Varimaxmethode die quadrierten Spaltensummen von  $\mathbf{L}$  maximiert, d.h. jeder Faktor soll möglichst viel Varianz der Variablen erklären.

Gemeinsam ist beiden Verfahren, dass eine Matrix  $\mathbf{Q}$  bestimmt wird, so dass

$$\mathbf{B} = \mathbf{L}\mathbf{Q}$$

gilt mit

$$\mathbf{Q} = \mathbf{Q}_1\mathbf{Q}_2 \dots \mathbf{Q}_l$$

und  $l = k(k-1)/2$ . Die einzelnen  $\mathbf{Q}_i$  sind dabei von der Form (Harman, 1967)

$$\mathbf{Q}_i = \begin{pmatrix} \cos(g \pm \varphi) & \pm \sin(g \pm \varphi) \\ \pm \sin(g \pm \varphi) & \cos(g \pm \varphi) \end{pmatrix}$$

d.h. Rotationen werden jeweils nur für zwei Faktoren gleichzeitig betrachtet, aber für alle möglichen Paare  $l = k(k-1)/2$  wird ein  $\mathbf{Q}_i$  bestimmt. Während die Form von  $\mathbf{Q}_i$  für die beiden orthogonalen Rotationstechniken identisch ist, wird  $\varphi$  jeweils anders bestimmt.

Bei der Quartimaxmethode gilt für  $\mathbf{L}\mathbf{Q}_i$

$$\tan 4\varphi = \frac{2 \sum_{j=1}^p (2l_{jt}l_{jm})(l_{jt}^2 - l_{jm}^2)}{\sum_{j=1}^p [(l_{jt}^2 - l_{jm}^2)^2 - (2l_{jt}l_{jm})^2]} \quad (2.36)$$

d.h. für je zwei Faktoren  $t$  und  $m$  ( $t, m = 1, \dots, k; t \neq m$ ) ist der Winkel  $\varphi$  welcher die Zeilensummen maximiert eindeutig aus  $\mathbf{L}$  zu bestimmen. Für  $\mathbf{Q}_i$  gilt dann in Abhängigkeit von den Vorzeichen des Zählers und Nenners der Gleichung 2.36

$$\mathbf{Q}_i = \begin{pmatrix} \cos \varphi & -\sin \varphi \\ \sin \varphi & \cos \varphi \end{pmatrix}$$

wenn Zähler und Nenner positiv sind,

$$\mathbf{Q}_i = \begin{pmatrix} \cos(45^\circ - \varphi) & -\sin(45^\circ - \varphi) \\ \sin(45^\circ - \varphi) & \cos(45^\circ - \varphi) \end{pmatrix}$$

bei positiven Zähler und negativen Nenner,

$$\mathbf{Q}_i = \begin{pmatrix} \cos \varphi & \sin \varphi \\ -\sin \varphi & \cos \varphi \end{pmatrix}$$

für negativen Zähler und positiven Nenner, und

$$\mathbf{Q}_i = \begin{pmatrix} \cos(45^\circ - \varphi) & \sin(45^\circ - \varphi) \\ -\sin(45^\circ - \varphi) & \cos(45^\circ - \varphi) \end{pmatrix}$$

sofern Zähler und Nenner negativ sind. Die durch die  $\mathbf{Q}_i$  Matrizen beschriebene Rotation kann nun wiederum an  $\mathbf{B}$  durchgeführt werden, bis das Quadrat der Zeilensummen konvergiert.

Bei der Varimaxmethode wird  $\varphi$  für  $\mathbf{LQ}_i$  mittels der Formel

$$\tan 4\varphi = \frac{Z - 2DS/p}{W - (D^2 - S^2)/p}$$

bestimmt, wobei

$$\begin{aligned} Z &= 2 \sum u_j v_j & D &= \sum u_j^2 \\ S &= \sum v_j^2 & W &= \sum u_j^2 - v_j^2 \\ u_j &= x_j^2 - y_j^2 & v_j &= 2x_j y_j \\ x_j &= l_{jt}/h_j & y_j &= l_{jm}/h_j \end{aligned}$$

und  $j = 1, \dots, p$  gilt. Die einzelnen  $\mathbf{Q}_i$  sind mit diesem Winkel wie bei der Quartimaxrotation definiert.

Eine oblique Rotationstechnik ist die Quartimin Transformation von Carroll (1953). Demnach wird eine Transformationsmatrix  $\mathbf{Q}$  gesucht, so dass

$$\mathbf{V} = \mathbf{LQ}$$

gilt und

$$N = \sum_{i=1}^p \sum_{l=2}^k v_{i(l-1)}^2 v_{il}^2 \quad (2.37)$$

minimiert wird. Zum Auffinden des Minimums von  $N$  werden die Ausdrücke

$$\begin{aligned} N_h &= \mathbf{q}'_h \mathbf{P} \mathbf{q}_h \\ &= \mathbf{q}'_h \mathbf{L}' \mathbf{W} \mathbf{L} \mathbf{q}_h \end{aligned}$$

( $h = 1, \dots, k$ ) um die Nebenbedingung  $\mathbf{q}'_h \mathbf{q}_h = 1$  erweitert, und partiell nach  $\mathbf{q}_h$  abgeleitet

$$(\mathbf{L}' \mathbf{W} \mathbf{L} - \lambda \mathbf{I}) \mathbf{q}_h = \mathbf{0}$$

d.h.  $N$  wird pro Spalte von  $\mathbf{Q}$  minimiert, wobei  $\mathbf{W}$  eine Diagonalmatrix mit den Elementen

$$w_i = \sum_{p \neq h=1}^k v_{ip}^2$$

in der Diagonalen ist, d.h. es werden Zeilenweise alle Elemente der  $\mathbf{V}$  Matrix im Quadrat summiert außer jene, die zu der Spalte gehören, die durch  $h$  bezeichnet ist. Der Eigenvektor  $\mathbf{q}_h$  zum kleinsten Eigenwert von  $\mathbf{L}' \mathbf{W} \mathbf{L}$  ist schließlich jener, der  $N_h$  minimiert.

Das beschriebene Vorgehen ist nun iterativ:

1. Schätzung von  $\mathbf{L}$  beispielsweise über die Varimaxmethode.
2. Ermittlung von  $\mathbf{W}$  auf der Basis von  $\mathbf{V}$  und einer (im ersten Durchgang beliebigen) Matrix  $\mathbf{Q}_n$ .
3. Bestimmung der kleinsten Eigenwerte und der zugehörigen Eigenvektoren pro  $\mathbf{W}$  Matrix.
4. Zusammenfassung der Eigenvektoren zur Matrix  $\mathbf{Q}_{n+1}$ .
5. Ermittlung von  $\mathbf{V} = \mathbf{L} \mathbf{Q}_{n+1}$  und  $N$ .
6. Falls  $N$  nicht konvergiert ist, erneut mit dem zweiten Schritt beginnen wobei  $\mathbf{Q}_n \leftarrow \mathbf{Q}_{n+1}$  gesetzt wird.

Die Quartiminmethode zählt zu einer Klasse von Transformationen, die von Carroll (s. Harman, 1967) als Oblimin bezeichnet wurden. Nach einem ähnlichen Vorgehen wie beim Quartiminverfahren wird nicht der Ausdruck 2.37 minimiert, sondern

$$B = \sum_{l=2}^k \left( p \sum_{i=1}^p v_{i(l-1)}^2 v_{il}^2 - \gamma \sum_{i=1}^p v_{i(l-1)}^2 \sum_{i=1}^p v_{il}^2 \right)$$

wobei  $\gamma = \beta/(\alpha+\beta)$  frei wählbar ist. Wenn  $\beta = 0$  ist, resultiert die Quartimin Rotation, und wenn  $\alpha = 0$  ist, dann resultiert die Covarimin Transformation. Da sich erstere als extrem schiefwinklig, letztere als zu orthogonal erwiesen, wurde von Carroll die Biquartimin Rotation empfohlen, in der  $\alpha = \beta$  ist, also  $\gamma = 0.5$ .

## 2.5 Kriterien zur Bestimmung der faktoriellen Struktur

In diesem Kapitel wird dargestellt, wie die faktorielle Struktur in der Faktorenanalyse bestimmt wird. Wie bereits oben erwähnt, entspricht die faktorielle Struktur der Anzahl der Faktoren bzw. Hauptkomponenten, deren korrelative Beziehungen untereinander sowie der Korrelation der einzelnen Testaufgaben mit den Faktoren bzw. Hauptkomponenten, d.h. insgesamt lassen sich drei Kriterien zur Bestimmung der faktoriellen Struktur unterscheiden. Auf jedes der drei Kriterien wird im nachfolgenden gesondert eingegangen.

Die Anzahl der Faktoren bzw. Hauptkomponenten entspricht der Anzahl der Spalten von  $\mathbf{F}$ , und im Strukturmodell hat  $\mathbf{F}$  genau  $p$ -Spalten. Allerdings werden in der Regel nicht alle  $p$  Faktoren bzw. Hauptkomponenten in der empirischen Forschung berücksichtigt, sondern nur die ersten  $k$  Komponenten ( $k < p$ ) ermittelt. Zur Bestimmung von  $k$  wird dabei meistens auf eine von drei Entscheidungsregeln zurückgegriffen:

1. Nach dem Eigenwertkriterium oder auch Kaiser-Guttman Kriterium (Guttman, 1954; Kaiser & Dickman, 1959) werden alle Faktoren bzw. Hauptkomponenten mit einem Eigenwert  $\lambda_j \geq 1$  berücksichtigt, d.h. es werden alle Faktoren bzw. Hauptkomponenten ausgewählt, die mindestens die durchschnittliche Varianz aller Variablen erklären. Allerdings hat sich gezeigt, dass gerade bei großen Variablenzahlen das Kaiser-Guttman Kriterium zu einer Überschätzung der tatsächlichen Faktorenanzahl führt (Zwick & Velicer, 1986).

2. Der Scree-Test nach Cattell (1966) geht vom Eigenwerteverlauf aus. Der Eigenwerteverlauf ergibt sich, wenn man die Eigenwerte einer Korrelationsmatrix der Größe nach auf der Abszisse eines Koordinatensystems anordnet und auf der Ordinate der Wert der Eigenwerte abgetragen wird. Es hat sich gezeigt, dass der Eigenwerteverlauf von Korrelationsmatrizen normalverteilter und unabhängiger Zufallsvariablen eher einen flachen Verlauf um eins hat während der typische Eigenwerteverlauf einer Korrelationsmatrix gemessener abhängiger Daten zunächst einen steilen, dann einen flachen Verlauf hat (s. Abbildung 2.2). Dementsprechend schlägt Cattell vor, nur jene Faktoren bzw. Hauptkomponenten zu extrahieren, deren Eigenwerte links vor dem Knick des Eigenwerteverlauf liegen (in der Abbildung 2.2 also die ersten drei Eigenwerte von links), weil die dann noch übrigen Eigenwerte einen Verlauf zeigen, der typisch ist für unabhängige Zufallsvariablen.

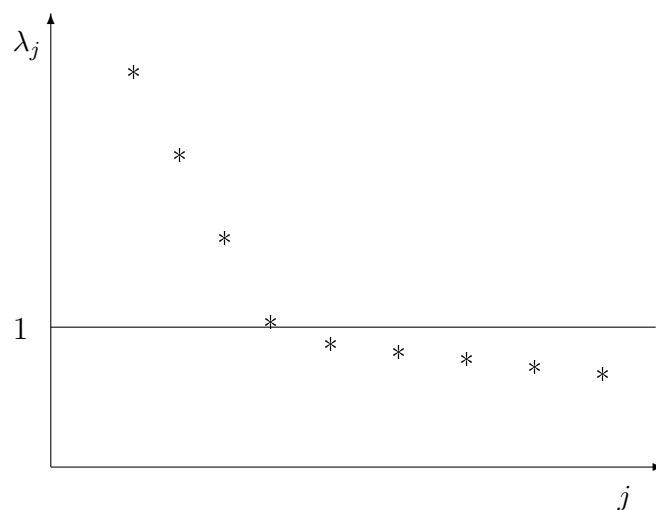


Abbildung 2.2: Eigenwerteverlauf abhängiger Daten

3. Kritisch an dem Scree-Test ist, dass er nicht immer zu objektiv nachvollziehbaren Ergebnissen führt und zwar insbesondere dann nicht, wenn der Eigenwerteverlauf weniger eindeutig ist als beispielsweise in Abbildung 2.2 dargestellt. Dieses Problem wird in der Parallelanalyse nach Horn (1965) umgangen, indem genau  $V$  Datensätze (meistens über 100 Datensätze) mit  $n$  Beobachtungen und  $p$  Variablen auf der Basis einer multivariaten Normalverteilung mit unabhängigen Variablen simuliert werden. Für jeden Datensatz wird eine Korrelationsmatrix und die  $p$  Eigenwerte berechnet. Anschließend werden die Eigenwer-

te für jeden Datensatz absteigend nach ihrer Größe geordnet so dass  $\lambda_{1v} \geq \lambda_{2v} \geq \dots \geq \lambda_{pv}$  ( $v = 1, \dots, V$ ) gilt. Unter der Annahme, dass die Variablen unabhängig sind, bildet die Verteilung des Eigenwerts  $\lambda_{jv}$  über die  $V$  Datensätze dann eine Schätzung für die Verteilung des  $j$ -ten Eigenwerts ab. Im Weiteren werden dann für die beobachteten Daten die Korrelationsmatrix sowie die  $p$  Eigenwerte bestimmt (die wiederum absteigend nach ihrer Größe zu ordnen sind) und es ist der  $j$ -te Faktor bzw. die  $j$ -te Hauptkomponente zu extrahieren, wenn der  $j$ -te Eigenwert  $\lambda_j$  größer ist als beispielsweise 95% der simulierten Eigenwerte  $\lambda_{jk}$ .<sup>13</sup>

Für die explorative Faktorenanalyse existiert zusätzlich ein  $\chi^2$ -Test um die Faktorenanzahl zu bestimmen, denn wie Jöreskog (1967), Jöreskog (1972) und Browne (1974) zeigen konnten, sind die Ausdrücke  $\gamma(\mathbf{L}, \mathbf{D}) - \ln|\mathbf{S}| + p$  und  $F(\mathbf{S}, \boldsymbol{\Sigma}|\mathbf{S})$  approximativ  $\chi^2$ -Verteilt mit  $df = [(p - k)^2 - (p + k)]/2$  Freiheitsgraden. Da der  $\chi^2$ -Test allerdings nicht für alle betrachteten Faktorenmodelle zur Verfügung steht, wird er bei den weiteren Analysen nicht berücksichtigt.

Nach der Bestimmung der Faktorenanzahl bzw. der Anzahl der Hauptkomponenten (auch als **Extraktion** bezeichnet) beträgt die Anzahl der Faktoren bzw. Hauptkomponenten nicht mehr  $p$  sondern  $k$ , wobei  $k \leq p$  gilt. Die Korrelation zwischen diesen extrahierten Faktoren bzw. Hauptkomponenten stehen in der Matrix  $\mathbf{F}'\mathbf{F}$ , wobei die Faktoren bzw. Hauptkomponenten so geschätzt werden dass  $\mathbf{F}'\mathbf{F} = \mathbf{I}$  gilt, d.h. die Faktoren bzw. Hauptkomponenten sind unkorreliert. Korrelierte Dimensionen können durch eine entsprechendes Rotationsverfahren sinnvoll modelliert werden.

Die Korrelationen der einzelnen Testaufgaben mit den Faktoren bzw. Hauptkomponenten stehen in der Ladungsmatrix  $\mathbf{L}'$ , denn  $\mathbf{F}'\mathbf{Z} = \mathbf{F}'\mathbf{F}\mathbf{L}' = \mathbf{L}'$ . Hohe Ladungskoeffizienten weisen auf einen starken Zusammenhang zwischen der jeweiligen Variable und dem Faktor bzw. der Hauptkomponente hin; die entsprechende Variable hat also einen relativ starken Einfluss auf den entsprechenden Testgesamtwert bzw. das latente Merkmal. Niedrige Ladungskoeffizienten verweisen auf Variablen, die zum Testgesamtwert bzw. zur latenten Dimension relativ wenig beitragen.

---

<sup>13</sup>In der Praxis haben sich mittlerweile verschiedene regressionsanalytische Verfahren etabliert, um die auf Simulation basierenden kritischen Größen für die Eigenwerte zu schätzen (z.B. Allen & Hubbard, 1986; Lautenschlager, Lance & Flaherty, 1989; Longman, Cota, Holden & Fekken, 1989; Wang, 2001).

## 2.6 Verteilungseigenschaften von $\mathbf{F}$ und $\mathbf{Z}$

Wie Micceri (1989) zeigen konnte, sind die Testdaten in  $\mathbf{Z}$  häufig rechtssteil verteilt; außerdem ist es im Falle von kriteriumsorientierten Tests durchaus nicht unplausibel, die Fähigkeitswerte in  $\mathbf{F}$  als nicht normalverteilt anzunehmen. Genau genommen kann angenommen werden, dass  $\mathbf{F}$  im Falle von kriteriumsorientierten Tests rechtssteil verteilt sein sollte. Warum ist das so? Nun, durch die Festlegung eines Lehrzieles bei kriteriumsorientierten Tests wird der getesteten Fähigkeit quasi ein maximaler Wert zugeordnet, wobei eine vollständige Erreichung des Lehrziels diesem maximal erreichbaren Wert bzw. dieser maximal getesteten Fähigkeitsausprägung entspricht. Weil die meisten Personen aufgrund des vorangestellten intensiven Trainings das Lehrziel in der Regel zumindestens annähernd erreichen, sollten die meisten Personen auch annähernd den maximalen Fähigkeitswert zugeordnet bekommen. Größere Abweichungen davon sollten vor allem nach unten hin möglich sein, weil die getestete Fähigkeitsdimension nur nach oben, nicht aber nach unten begrenzt wurde. In der Konsequenz resultiert damit eine rechtssteile Verteilung der Fähigkeitswerte.

Bevor klassische Faktorenmodelle zur Überprüfung der faktoriellen Struktur bei rechtssteil verteilten Testdaten bzw. bei kriteriumsorientierten Tests eingesetzt werden, ist also zu klären, welche Verteilungseigenschaften die klassische Faktorenanalyse bezüglich  $\mathbf{Z}$  und  $\mathbf{F}$  aufweist. Dabei soll zunächst  $\mathbf{F}$  und danach  $\mathbf{Z}$  betrachtet werden, wobei in beiden Fällen zwischen den wahren Werten  $\mathbf{F}$  und  $\mathbf{Z}$  und den aufgrund des Modells geschätzten Werten  $\hat{\mathbf{F}}$  und  $\hat{\mathbf{Z}}$  zu unterscheiden ist.

Wie an den Voraussetzungen der klassischen Faktorenanalyse zu erkennen, wird bezüglich der Verteilungseigenschaften von  $\mathbf{F}$  lediglich gefordert, dass  $E(\mathbf{f}) = \mathbf{0}$  bzw.  $\text{cov}(\mathbf{f}\mathbf{e}') = \mathbf{0}$  gilt. Aussagen über die Verteilungsform von  $\mathbf{F}$  werden damit nicht getroffen. Dennoch lassen sich Erwartungen bezüglich der Verteilungsform von  $\hat{\mathbf{F}}$  ableiten, denn nach Formel 2.4 (S. 16), Gleichung 2.33 (S. 50) und Gleichung 2.35 (S. 51) ist  $\hat{\mathbf{F}}$  als gewichtete Summe von Zufallsvariablen darstellbar, und in Anlehnung an den zentralen Grenzwertsatz nähern sich derartige gewichtete Summenwerte mit zunehmender Anzahl an Variablen der Normalverteilung an (Feller, 1971; Klartag, 2007; Lindeberg, 1922; Stroock, 1993).

In Bezug auf die Verteilung von  $\mathbf{Z}$  ist bei Verwendung der allgemeinen Diskrepanzfunktion zu fordern, dass die vierten Kumulanten von  $\mathbf{z}$  Null sind bzw. im Falle der Maximum-Likelihood-Methode, dass  $\mathbf{z}$  normalverteilt ist mit  $\mathbf{z} \sim N(\mathbf{0}, \mathbf{\Sigma})$ . In allen anderen vorgestellten Modellen werden keine Anforderungen an die Verteilungsform von  $\mathbf{Z}$  gestellt. Dennoch gilt auch



hier, dass die geschätzten Werte  $\hat{\mathbf{Z}}$  approximativ einer Normalverteilung folgen, wenn die Anzahl der Faktoren gegen unendlich strebt, weil auch die geschätzten Werte  $\hat{\mathbf{Z}}$  als gewichtete Summe von Zufallsvariablen darstellbar sind ( $\hat{\mathbf{Z}} = \hat{\mathbf{F}}\hat{\mathbf{L}}'$  bzw.  $\hat{\mathbf{Z}} = \hat{\mathbf{F}}\hat{\mathbf{L}}' + \hat{\mathbf{E}}$ ) und damit den Bedingungen des zentralen Grenzwertsatzes genügen.

Die klassische Faktorenanalyse scheint also nur begrenzt geeignet zu sein, rechtssteil verteilte Testwerte  $\mathbf{Z}$  bzw. Fähigkeitswerte  $\mathbf{F}$  zu schätzen, weil erwartet werden kann, dass sich die geschätzten Testwerte  $\hat{\mathbf{Z}}$  bzw. Fähigkeitswerte  $\hat{\mathbf{F}}$  mit zunehmender Anzahl der Faktoren bzw. Testitems der Normalverteilung annähern. Über das Ausmaß der mit der Normalverteilungsapproximation von  $\hat{\mathbf{Z}}$  bzw.  $\hat{\mathbf{F}}$  verbundenen Verschätzung der faktoriellen Struktur ist allerdings wenig bekannt (s. nachfolgendes Kapitel), weswegen diesem Sachverhalt in der Simulationsstudie nachgegangen wird.

## 2.7 Empirischer Forschungsstand

Wie in den vorherigen Kapiteln gezeigt, sind die Modelle der klassischen Faktorenanalyse an bestimmte Voraussetzungen gebunden und implizieren (im Falle der Maximum-Likelihood-Methode sogar explizieren) eine bestimmte (nämlich normalverteilte) Verteilungsform von  $\mathbf{Z}$  und  $\mathbf{F}$ . In diesem Kapitel sollen empirische Forschungsarbeiten aufgeführt werden, die der Frage nachgegangen sind, welche Auswirkungen die Verletzung dieser Voraussetzungen und Verteilungseigenschaft auf die Schätzgüte der klassischen Faktorenanalyse hat. Im Fokus der bisher durchgeführten empirischen Forschungsarbeiten stand dabei die Annahme intervallskalierter Testdaten und die mit der Faktorenanalyse verbundene Verteilungseigenschaft von  $\mathbf{Z}$ . Genau genommen wurde der Frage nachgegangen, wie genau die klassische Faktorenanalyse die faktorielle Struktur von Testdaten schätzt, wenn  $\mathbf{Z}$  nicht intervallskaliert bzw. nicht normalverteilt ist.

Eine erste Untersuchung zu diesem Thema legten W. S. Martin, Fruchter und Mathis (1974) vor. Diese Autoren untersuchten den Effekt der Intervallanzahl von Indikatoren auf die Ergebnisse der Hauptkomponentenmethode. Als Datengrundlage dienten die an 110 männlichen Schülern des zehnten Jahrgangs gemessenen Werte auf 11 stetig skalierten Merkmalen (z.B. systolischer und diastolischer Blutdruck). Neben einer Hauptkomponentenzerlegung der kontinuierlich skalierten Variablen untersuchten die Autoren dieselben Indikatoren, wenn diese auf neun, sieben, fünf, vier, drei oder zwei Intervalle skaliert wurden. Als Rotationskriterium wurde die Varimaxmethode verwendet.

Die Ergebnisse der Analysen zeigten, dass der Eigenwert des ersten Faktors mit der Reduktion der Intervallanzahl kleiner wurde, während der Eigenwert des dritten extrahierten Faktors mit der Reduktion der Intervallanzahl an Größe zunahm (die weiteren Eigenwerte wurden nicht berichtet). Dieser Effekt führte dazu, dass bei Verwendung des Kaiser-Guttman Kriteriums im Falle von dichotomen Indikatoren mehr Faktoren extrahiert werden mussten als im Falle von stetig skalierten Merkmalen: Während bei stetigen Merkmalen zwei Faktoren zur Erklärung der Testergebnisse ausreichten, mussten bei dichotomen Indikatoren drei Faktoren angenommen werden.

Bezüglich  $\mathbf{L}$  konnten die Autoren einen positiven Zusammenhang zwischen der Größe der Ladungen und der Intervallanzahl der Indikatoren feststellen, d.h. bei stetigen Variablen waren die Faktorladungen im Allgemeinen größer als bei dichotomen Indikatoren. Die Korrelation der Indikatoren mit den Faktoren nahm also mit zunehmender Reduktion der Intervallanzahl der Indikatoren ab. Dieser Effekt fiel für kleine Ladungen (etwa  $|l| \leq .30$ ) allerdings nicht so deutlich aus wie für große Ladungen (etwa  $|l| \geq .70$ ).

Zusammenfassend weisen die Ergebnisse von W. S. Martin et al. (1974) darauf hin, dass mit der klassischen Faktorenanalyse mehr Faktoren extrahiert werden, wenn  $\mathbf{Z}$  nicht stetig sondern dichotom skaliert ist. Außerdem ist zu erwarten, dass die Korrelationen der manifesten Merkmale mit den latenten Faktoren im Allgemeinen um so geringer ausfallen, je weniger Intervalle die manifesten Merkmale aufweisen. Kritisch Anzumerken ist allerdings, dass die wahre Dimensionalität und die wahre Ladungsmatrix  $\mathbf{L}$  der Testdaten nicht bekannt war, d.h. die den Daten tatsächlich zugrundeliegende faktorielle Struktur war unbekannt. Deswegen kann nicht abgeschätzt werden, ob die mit der Reduktion der Intervallanzahl verbundene Veränderung von  $k$  bzw.  $\mathbf{L}$  zu einer genaueren oder ungenaueren Schätzung der wahren faktoriellen Struktur der Testdaten geführt hat.

Da die wahre faktorielle Struktur empirischer Testdaten in der Regel nicht bekannt ist, und deswegen mit Faktorenmodellen geschätzt werden soll, werden zur Überprüfung der Schätzgüte verschiedener Faktorenmodelle häufig Simulationsstudien durchgeführt. Im Rahmen dieser Simulationsstudien wird eine wahre faktorielle Struktur auf der Basis eines ausgewählten Faktorenmodells vorgegeben, um dann ausgehend von dieser wahren Struktur über die entsprechende Strukturgleichung des Faktorenmodells die dazugehörige Datenmatrix  $\mathbf{Z}$  zu erzeugen. Diese Datenmatrix wird dann mit verschiedenen Faktorenmodellen analysiert, wobei sowohl das Extraktionskriterium als auch die Rotationstechnik variieren können. Die Ergebnisse dieser Analysen, also die geschätzte faktorielle Struktur, kann dann mit der wahren faktoriellen Struktur verglichen werden.

Das Vorgehen von derartigen Simulationsstudien kann auch am Beispiel der explorativen Faktorenanalyse erklärt werden. Die wahre faktorielle Struktur wird hier durch die Vorgabe von Werten für  $k$ ,  $\mathbf{L}$ ,  $\mathbf{F}$  und  $\mathbf{E}$  spezifiziert. Sind diese Werte festgelegt, kann über die Strukturgleichung  $\mathbf{Z} = \mathbf{F}\mathbf{L}' + \mathbf{E}$  die dazugehörige Datenmatrix  $\mathbf{Z}$  erzeugt werden. Diese Datenmatrix wird dann mit verschiedenen Faktorenmodellen analysiert, wobei sowohl das Extraktionskriterium als auch die Rotationstechnik variieren können. Die Ergebnisse dieser Analysen, also die geschätzten Werte für  $k$ ,  $\mathbf{L}$ ,  $\mathbf{F}$  und  $\mathbf{E}$  können dann mit den vorgegebenen Werten verglichen werden.

In Simulationsstudien wird der beschriebene Prozess häufig wiederholt durchgeführt, wobei die Strukturparameter (d.h. jene Werte, die die wahre faktorielle Struktur definieren) fixiert werden, während die konkreten Werte über die Wiederholungen variieren. Ziel dieser Wiederholungen ist es, die stichprobenbedingte Streuung, d.h. den Standardfehler, der Analyseergebnisse abschätzen zu können. Im Falle des explorativen Faktorenmodells würden also beispielsweise  $k$  und  $\mathbf{L}$  fixiert sein, während für  $\mathbf{F}$  und  $\mathbf{E}$  pro Wiederholung neue Werte generiert werden.<sup>14</sup> Anstelle einer einzelnen Datenmatrix  $\mathbf{Z}$  resultieren also  $v$  Datenmatrizen, wobei  $v$  der Anzahl der Wiederholungen entspricht. Alle  $v$  Datenmatrizen werden im Anschluss mit den ausgewählten Faktorenmodellen analysiert und die resultierenden Verteilungen der Schätzer beschrieben.<sup>15</sup>

Die vorzugebenden Werte für die Simulation einer wahren faktoriellen Struktur bzw. den darauf basierenden Testwerten variieren natürlich etwas mit dem zur Konstruktion der Daten ausgewählten Faktorenmodell, an dem prinzipielle Vorgehen ändert sich dadurch allerdings nichts. Dies muss erwähnt werden, weil viele der nachfolgend aufgeführten empirischen Forschungsarbeiten zur Simulation der wahren faktoriellen Struktur bzw. der dazugehörigen Testwerte auf IRT-Modelle zurückgegriffen haben.

Die Auswahl von IRT-Modellen ist deswegen plausibel, weil die Studien auf die Simulation einer Datenmatrix  $\mathbf{Z}$  mit dichotom skalierten Merkmalen fokussieren, und IRT-Modelle bestens dazu geeignet sind, eine solche Datenmatrix zu konstruieren (s. dazu die Ausführungen in Kapitel 2.1).<sup>16</sup> Die auf

<sup>14</sup>Die Werte für  $\mathbf{F}$  und  $\mathbf{E}$  stellen in der Regel eine Zufallsauswahl aus vorher spezifizierten Verteilungen (meistens Normalverteilungen) dar.

<sup>15</sup>Um den Überblick zu wahren, wird bei den weiteren Ausführungen zu den empirischen Forschungsergebnissen der Ablauf nur in Bezug auf eine beliebige Replikation  $\mathbf{Z}$  beschrieben. Soweit nicht anders vermerkt, ist der Wiederholungsprozess dabei mitzudenken.

<sup>16</sup>Auf eine Darstellung von IRT-Modellen wird verzichtet, da eine Kenntnis dieser Modelle für das Verständnis der weiteren Ausführungen nicht notwendig ist. Der interessierte Leser sei auf Bock und Lieberman (1970), Christoffersson (1975), Christoffersson (1977), Muthén (1978), Muthén und Christoffersson (1981) oder Muthén (1983, 1984) verwiesen.

Basis des ausgewählten IRT-Modells simulierte Datenmatrix  $\mathbf{Z}$  wird sodann mit klassischen Faktorenmodellen analysiert, wobei die Frage im Vordergrund steht, wie genau die klassischen Faktorenmodelle die wahre Dimensionalität schätzen, wenn die Daten nicht intervall- sondern dichotom skaliert sind.

Ein solches Vorgehen wählte beispielsweise Green (1983), der ein mehrdimensionales dreiparametrisches Birnbaummodell zur Simulierung von Testdatensätzen verwendete. Neben der Anzahl der wahren Dimensionen ( $k = 0$ ,  $k = 1$  oder  $k = 2$ ) variierte Green auch die Anzahl der Aufgaben, für die Daten simuliert wurden ( $p = 10$ ,  $p = 15$ ,  $p = 20$ ,  $p = 25$ ,  $p = 30$ ,  $p = 35$ ,  $p = 40$ ,  $p = 45$  oder  $p = 50$ ) und die Stichprobengröße ( $n = 100$ ,  $n = 200$  oder  $n = 500$ ). Insgesamt simulierte Green für jede der damit resultierenden  $3 \times 9 \times 3 = 81$  verschiedenen Konstellationen jeweils gut 27 Testdatensätze, womit eine Variation der Diskriminationsparameter  $a_{jt}$  ( $t = 1, \dots, k$ ), der Schwierigkeitsparameter  $b_j$  und der Rateparameter  $c_j$  berücksichtigt wurde.<sup>17</sup> Jeder dieser Testdatensätze wurde sodann mit Hilfe der Hauptkomponentenmethode und der Hauptfaktorenanalyse untersucht, wobei für die Hauptkomponentenmethode der Scree-Test und für die Hauptfaktorenanalyse die Parallelanalyse als Extraktionskriterium gewählt wurde.

Wie die Untersuchungen der Testdatensätze zeigten, konnte eine wahre Dimensionalität von  $k = 0$  mit der Hauptfaktorenanalyse als Faktorenmodell und der Parallelanalyse als Extraktionskriterium in allen Kombinationen und Replikationen entdeckt werden. Wenn hingegen die wahre Dimensionalität  $k = 1$  oder  $k = 2$  betrug, dann führte die Hauptfaktorenanalyse mit anschließender Parallelanalyse bei nahezu allen Kombinationen und Replikationen zu einer Überschätzung der wahren Dimensionalität, und zwar wurde jeweils ein Faktor mehr extrahiert, als den Daten in Wahrheit zugrundelag, d.h. im Falle eines wahren  $k = 1$  wurde  $k = 2$  angenommen und im Falle eines wahren  $k = 2$  wurde  $k = 3$  angenommen. Die Ergebnisse zur Hauptkomponentenmethode in Verbindung mit dem Scree-Test werden von Green nicht weiter erläutert.

Die Ergebnisse von Green (1983) sprechen also dafür, dass die Hauptfaktorenanalyse in Verbindung mit der Parallelanalyse zu einer Überschätzung der wahren Faktorenanzahl führt, wenn dichotom skalierte Aufgaben analysiert werden. Dies korrespondiert mit den Befunden von W. S. Martin et al. (1974), demnach bei dichotom skalierten Indikatoren mehr Faktoren extrahiert werden als bei intervallskalierten Merkmalen, wenn die Hauptkomponentenmethode in Verbindung mit dem Kaiser-Guttman Kriterium auf die

<sup>17</sup>Die Diskriminationsparameter  $a_{jt}$  entsprechen  $l_{jt}$  im klassischen Faktorenmodell; für die Schwierigkeitsparameter  $b_j$  und Rateparameter  $c_j$  des IRT-Modells gibt es kein entsprechendes Äquivalent in den klassischen Faktorenmodellen.

Daten angewendet wird. Wie kann dieser Effekt der Überschätzung der Faktorenanzahl bei dichotom skalierten Aufgaben erklärt werden?

Nach Ferguson (1941) handelt es sich bei der Überschätzung um ein methodisches Artefakt. Ferguson konnte zeigen, dass die von W. S. Martin et al. (1974) und Green (1983) verwendete Produkt-Moment-Korrelation zwischen dichotomen Aufgaben, der sogenannte Phi-Koeffizient  $\phi$ , von der Schwierigkeit der Aufgaben abhängt: Je größer der Unterschied zwischen der Schwierigkeit von Aufgaben ist, desto kleiner wird die maximal erreichbare Korrelation zwischen den Aufgaben. Dieser Zusammenhang führt im Extremfall dazu, dass Faktoren alleine aufgrund der unterschiedlichen Schwierigkeit der Aufgaben extrahiert werden, weswegen diese Faktoren auch als *difficulty factors* bezeichnet werden (Roznowski, Tucker & Humphreys, 1991).

Der Phi-Koeffizient kann im Rahmen der klassischen Faktorenanalyse also zu einer Überschätzung der wahren Faktorenanzahl führen, weil die maximal erreichbare Korrelation zwischen den Aufgaben die wahre Korrelation zwischen den Aufgaben unterschätzt, wenn sich die Aufgaben in ihrer Schwierigkeit unterscheiden. Deswegen schlägt Carroll (1945) vor, für faktorenanalytische Studien nicht auf  $\phi$  zurückzugreifen, sondern den tetrachorischen Korrelationskoeffizienten  $r_{tet}$  zu berechnen. Der tetrachorische Korrelationskoeffizient kommt zur Anwendung, wenn Aufgaben intervallskaliert und die durch die Aufgaben erfassten Merkmale normalverteilt sind, die Testwerte der Aufgaben aber dichotomisiert wurden, d.h. es handelt sich um künstliche Dichotomien. In diesem Fall schätzt die tetrachorische Korrelation zwischen den dichotomen Aufgaben die wahre Korrelation zwischen den intervallskalierten Aufgaben.

In einer Simulationsstudie vergleichen Collins, Cliff, McCormick und Zatkin (1986) die Ergebnisse der Hauptkomponentenmethode wenn  $\phi$ -Koeffizienten faktorisiert wurden mit den Ergebnissen der Hauptkomponentenmethode angewendet auf tetrachorische Korrelationen. Als Extraktionskriterium diente eine algebraische Variante des Scree-Test. Zur Konstruktion der Testdaten verwendeten die Autoren das zweiparametrische Birnbaummodell, d.h. die analysierten Daten waren dichotom skaliert.

Insgesamt simulierten die Autoren 12 Bedingungen, wobei neben der Trennschärfe der Aufgaben (moderat, hoch, sehr hoch) die Verteilung der Aufgabenschwierigkeiten (mittelschwer Aufgaben mit Normalverteilung der Aufgabenschwierigkeiten, schwere Aufgaben mit Normalverteilung der Aufgabenschwierigkeiten, mittelschwer Aufgaben mit uniformer Verteilung der Aufgabenschwierigkeiten, extrem leichte und extrem schwere Aufgaben mit bimodaler Verteilung der Aufgabenschwierigkeiten) und die Verteilung der latenten Fähigkeitswerte (Normalverteilung, Uniform) variiert wurde, obgleich

die Fähigkeitsverteilung nicht unabhängig von der Verteilung der Aufgabenschwierigkeiten verändert wurde. Für jede der 12 Bedingungen generierten die Autoren fünf Testdatensätze mit jeweils  $k = 3$  Dimensionen und  $n = 500$  Fällen.

Wie Collins et al. (1986) zeigen konnten, führten beide Vorgehensweisen (Faktorisierung von  $\phi$  vs. Faktorisierung von  $r_{tet}$ ) zu der genauesten Schätzungen bezüglich der wahren Faktorenanzahl, wenn die Aufgaben mittelschwer und sowohl die Aufgabenschwierigkeiten als auch die Fähigkeitswerte normalverteilt waren (Entdeckungsrate unabhängig vom faktorisierten Korrelationskoeffizienten etwa 70 Prozent), d.h. wenn die Daten sich in gewisser Weise ideal verhalten haben. Die ungenaueste Schätzung der wahren Faktorenanzahl resultierte hingegen dann, wenn die Aufgabenschwierigkeiten bimodal verteilt waren ( $\phi$ ; Entdeckungsrate etwa 20 Prozent) bzw. wenn ausschließlich schwere Aufgaben vorlagen ( $r_{tet}$ ; Entdeckungsrate etwa 7 Prozent), d.h. wenn die Testdaten  $\mathbf{Z}$  nicht nur dichotom skaliert, sondern auch schief verteilt waren. Über alle 12 Bedingungen hinweg war es im Falle von  $\phi$  möglich, in rund 45 Prozent der Datensätze die wahre Dimensionalität zu entdecken; wurde hingegen  $r_{tet}$  faktorisiert, dann betrug die Entdeckungsrate 33 Prozent.

Die Ergebnisse von Collins et al. (1986) sprechen dafür, bei klassischen Faktorenanalysen von dichotom skalierten Aufgaben nicht (in jedem Fall) wie von Carroll (1945) empfohlen auf den tetrachorischen Korrelationskoeffizienten zurückzugreifen. Vielmehr scheint die Faktorisierung der tetrachorischen Korrelation eine noch größere Verschätzung der faktoriellen Struktur zu implizieren als die Faktorisierung des  $\phi$ -Koeffizienten. Gegen die Verwendung der tetrachorischen Korrelation im Rahmen von klassischen Faktorenanalysen sprechen aber noch ganz andere Gründe: Zum einen führen tetrachorische Korrelationen nicht selten zu Heywood Cases, d.h. einige Kommunalitäten fallen deutlich größer aus als Eins (Collins et al., 1986), zum anderen sind Varianz-Kovarianzmatrizen bzw. Korrelationsmatrizen von tetrachorischen Korrelationen häufig nicht positiv definit (Green, 1983; Weng & Cheng, 2005), was dazu führt, dass explorative Faktorenanalysen nicht durchgeführt werden können (Fahrmeir et al., 1996).

Tetrachorische Korrelationen sind im Allgemeinen also nicht besser geeignet als  $\phi$ -Koeffizienten, um explorative Faktorenanalysen mit dichotom skalierten Aufgaben durchzuführen. Darauf deuten auch die Ergebnisse der Simulationsstudie von Roznowski et al. (1991) hin. Diese Autoren überprüften welche Methode am besten geeignet ist, um Eindimensionalität festzustellen. Verglichen wurden dabei die drei Methoden local independence-Indize (LII), Ladungsstruktur der Aufgaben auf dem zweiten Faktor (LS) und eine algebraische Variante des Scree-Test (ST), wobei vor dem Scree-Test eine Hauptfaktoren-

analyse mit den  $\phi$ -Koeffizienten, den tetrachorischen Korrelationen oder mit der Varianz-Kovarianzmatrix durchgeführt wurde.

Um zu überprüfen, welche der drei Methoden bei unterschiedlichen Bedingungen am besten geeignet ist Eindimensionalität festzustellen, variierten die Autoren die Anzahl der wahren Faktoren ( $k = 1, k = 2, k = 3, k = 4$  und  $k = 5$ ), die Korrelation zwischen den Faktoren (schwacher Zusammenhang, mittlerer Zusammenhang und starker Zusammenhang), die Anzahl der Aufgaben ( $p = 20, p = 30, p = 40, p = 50$  und  $p = 60$ ), die Verteilung der Aufgabenschwierigkeiten (schwere Aufgaben mit relativ großer Streuung der Aufgabenschwierigkeiten, mittelschwere Aufgaben mit mittlerer Streuung der Aufgabenschwierigkeiten und leichte Aufgaben mit relativ kleiner Streuung der Aufgabenschwierigkeiten) und die Stichprobengröße ( $n = 125, n = 500$  und  $n = 2.000$ ). Schließlich generierten Roznowski et al. (1991) auf der Basis des explorativen Faktorenmodells für jede der 675 Bedingungen 100 Testdatensätze. Nach einer Dichotomisierung der generierten kontinuierlichen  $z_j$  wurden die Testdatensätze mit allen drei Methoden analysiert.

In den Ergebnissen der Studie zeigte sich, dass der LII über alle Bedingungen hinweg eindimensionale Testdatensätze am genauesten identifizieren konnte: Je größer die Stichprobe bzw. je kleiner die Korrelation zwischen den Faktoren war, desto deutlicher unterschied sich der LII, angewendet auf eindimensionale Testdatensätze, von dem LII, angewendet auf mehrdimensionale Testdatensätze. Mit der LS-Methode konnten eindimensionale Testdatensätze über alle Bedingungen hinweg zwar ähnlich genau identifiziert werden wie mit LII, aber die Leistungsfähigkeit dieser Methode war von der Interaktion zwischen der Stichprobengröße, der Korrelation der Faktoren und der Verteilung der Aufgabenschwierigkeit abhängig. Schließlich konnten eindimensionale Testdatensätze am wenigsten genau von mehrdimensionalen Testdatensätzen unterschieden werden, wenn der ST verwendet wurde. Eine genaue Unterscheidung zwischen eindimensionalen und mehrdimensionalen Testdatensätzen war dabei vor allem dann nicht möglich, wenn tetrachorische Korrelationen faktorisiert wurden.

Eine ganz andere Erklärung für die Überschätzung der Dimensionalität bei dichotomen Daten liefern Weng und Cheng (2005). In einer Simulationsstudie variierten die Autoren die Anzahl der Aufgaben ( $p = 8$  und  $p = 20$ ), die Größe der Faktorladungen ( $|l_{max}| = .45, |l_{max}| = .70$  und  $|l_{max}| = .90$ ), die relative Anzahl der Lösungen bei dichotomen Aufgaben (50/50, 60/40, 70/30, 80/20 und 90/100) und die Stichprobengröße ( $n = 50, n = 100, n = 200, n = 500$  und  $n = 1.000$ ). Sodann konstruierten Weng und Cheng für die daraus resultierenden 150 Bedingungen jeweils 500 eindimensionale Testdatensätze auf der Basis eines explorativen Faktorenmodells. Wie Green (1983)

verwendete die Autoren die Parallelanalyse zur Bestimmung der Faktorenanzahl der Testdatensätze, jedoch wurden die Eigenwerte der zufallsbasierten Korrelationsmatrizen über einen anderen Algorithmus geschätzt, der sich in einer vergleichenden Studie als zuverlässiger erwiesen hat (Wang, 2001).

Mit diesem Algorithmus als Extraktionsmethode führten sowohl  $\phi$  als auch  $r_{tet}$  nur zu einer Verschätzung der wahren Dimensionalität, wenn:

- die Faktorladungen sehr klein waren ( $\phi$ -Koeffizienten und tetrachorische Korrelation) oder
- die Stichprobengröße sehr klein war ( $\phi$ -Koeffizienten) oder
- die Aufgaben sehr leicht waren (tetrachorische Korrelation).

Demnach sind es nicht die methodischen Eigenschaften der Korrelationskoeffizienten *per se*, welche bei der Faktorisierung von dichotom skalierten Aufgaben zu einer Überschätzung der Dimensionalität führen, sondern die verwendete Extraktionsmethode.

Zusammenfassend kann festgestellt werden, dass klassische Faktorenmodelle offensichtlich sensibel auf die Skalierung von  $\mathbf{Z}$  reagieren: Sind die einzelnen Aufgaben nicht intervall- sondern dichotom skaliert, dann wird die wahre Dimensionalität tendenziell überschätzt. Dieser Effekt ist bei der Faktorisierung von tetrachorischen Korrelationen etwas größer als bei der Faktorisierung von  $\phi$ -Koeffizienten und er kann scheinbar vermieden werden, wenn die Parallelanalyse in der Modifikation wie bei Weng und Cheng (2005) als Extraktionskriterium verwendet wird. Die Studie von Collins et al. (1986) legt außerdem nahe, dass die Verteilungseigenschaft von  $\mathbf{Z}$  bzw.  $\mathbf{F}$  einen Einfluss auf die Schätzgüte klassischer Faktorenmodelle hat. Die wahre Faktorenanzahl kann demnach am genauesten geschätzt werden, wenn die einzelnen  $\mathbf{f}_j$  einer Normalverteilung folgen und die einzelnen  $\mathbf{z}_j$  keine Schiefe aufweisen.

Der zuletzt genannte Befund kann allerdings nicht eindeutig interpretiert werden, weil es sich um einen konfundierten Effekt handelt: Der Faktor der Aufgabenschwierigkeit bzw. Verteilung von  $\mathbf{z}_j$  wurde nicht unabhängig von der Verteilung von  $\mathbf{f}_j$  variiert. Damit bleibt es letztlich unklar, welchen Einfluss alleine schief verteilte Fähigkeitswerte auf die Verteilung von  $\mathbf{Z}$  bzw. auf die Anzahl extrahierter Dimensionen haben. Unklar bleibt auch, inwiefern andere Maße zur Bestimmung der Schätzgüte von klassischen Faktorenmodellen (etwa die Übereinstimmung zwischen wahren  $\mathbf{L}$  und geschätzten  $\mathbf{L}$  oder die Verteilung von wahren  $\mathbf{F}$  und geschätzten  $\mathbf{F}$ ) von schief verteilten Fähigkeitswerten beeinflusst werden, da keine der vorgenannten Simulationsstudien diese Effekte untersucht hat. Deswegen wird diesen Aspekten in den nachfolgenden Kapiteln nachgegangen.



# Kapitel 3

## Methodische Grundlagen

### 3.1 Forschungsfragen

Wie aus den vorherigen Ausführungen deutlich wurde, ist die klassische Faktorenanalyse an spezifische Voraussetzungen gebunden. Nicht alle dieser Voraussetzungen werden von diskret skalierte Messinstrumenten im Allgemeinen, und von kriteriumsorientierte Tests im Besonderen erfüllt. Insbesondere die Annahme normalverteilter manifester Daten bzw. normalverteilter Fähigkeitswerte scheint nicht in jedem Fall plausibel zu sein.

Forschungsergebnisse darüber, welchen Einfluss die Verletzung dieser Voraussetzungen auf die Schätzgenauigkeit klassischer Faktorenanalysen hat liegen nur zum Teil vor, und können aufgrund von konfundierten Effekten nicht eindeutig interpretiert werden. Deswegen ergeben sich im Kontext der Analyse von kriteriumsorientierten Tests im Allgemeinen, und schief verteilter Daten im Besonderen, die dieser Dissertation zugrundeliegenden offenen Forschungsfragen.<sup>1</sup>

1. Welcher Zusammenhang besteht zwischen rechtssteil verteilten manifesten Testdaten und der zugrundeliegenden Faktorstruktur bzw. der Verteilung der latenten Fähigkeitswerte?

Manifeste Testdaten von diskret skalierte Messinstrumenten im Allgemeinen und von kriteriumsorientierte Tests im Besonderen sind nicht selten rechtss-

---

<sup>1</sup>Alle aufgeführten Forschungsfragen und die nachfolgend aufgestellten Hypothesen beziehen sich ausschließlich auf die in Kapitel 2.3 vorgestellten klassischen Faktorenmodelle. Um die Übersichtlichkeit zu wahren, wird bei den einzelnen Aussagen auf diesen Hinweis verzichtet.

teil verteilt (Micceri, 1989). Sollen klassische Faktorenanalysen zur Überprüfung der faktoriellen Validität dieser Testdaten angewendet werden ergibt sich damit die Frage, ob die Verteilung der manifesten Daten bereits eine bestimmte wahre faktorielle Struktur der Testdaten impliziert oder ob jede (beliebige) faktorielle Struktur bei gegebenen Daten extrahiert werden kann.

2. Welcher Zusammenhang besteht zwischen der Schätzgenauigkeit von klassischen Faktorenanalysen und dem Grad an Rechtssteile der Fähigkeitswerte?

Wird angenommen, dass die Fähigkeitswerte nicht normalverteilt sondern rechtssteil verteilt sind, dann stellt sich die Frage, inwiefern die Schätzgenauigkeit klassischer Faktorenanalysen mit dem Grad an Rechtssteile der Fähigkeitswerte variiert.

3. Lassen sich bestimmte Kriterien der faktoriellen Struktur identifizieren, hinsichtlich derer die klassische Faktorenanalyse robust gegenüber dem Ausmaß an Rechtssteile der Fähigkeitswerte ist?

Sollte ein Zusammenhang zwischen der Schätzgenauigkeit von klassischen Faktorenanalysen und dem Grad an Rechtssteile der Fähigkeitswerte existieren, dann stellt sich die Frage, ob alle Kriterien zur Bestimmung der faktoriellen Struktur eines Tests von diesem Zusammenhang in gleicher Weise betroffen sind, oder ob einige Kriterien der faktoriellen Struktur robust gegenüber dem Grad an Rechtssteile der Fähigkeitswerte sind.

4. Hängt die Schätzgenauigkeit klassischer Faktorenmodelle bei rechtssteil verteilten Fähigkeitswerten von dem Extraktionskriterium zur Bestimmung der Faktorenanzahl ab?

Ein zentraler Bestandteil zur Bestimmung der faktoriellen Struktur ist die Anzahl der mit einer Faktorenanalyse extrahierten Dimensionen. Allerdings existieren verschiedene Verfahren, um die Anzahl der zu extrahierenden Dimensionen zu bestimmen. Deswegen erhebt sich die Frage, ob das verwendete Extraktionskriterium einen Einfluss auf die Schätzgenauigkeit klassischer Faktorenmodelle bei rechtssteil verteilten Fähigkeitswerten hat.

## 3.2 Hypothesen

Zu jeder der aufgeworfenen Fragen lässt sich eine Hypothese aus den theoretischen Grundlagen bzw. dem bisherigen Forschungsstand ableiten. Diese Hypothesen werden nachfolgend formuliert und anschließend anhand von theoretischen Herleitungen sowie Datensimulationen überprüft.

Hypothese I: Es wird angenommen, dass schief verteilte Daten bestimmte Faktorstrukturen bzw. bestimmte Verteilungen der Fähigkeitswerte implizieren.

Collins et al. (1986) konnten zeigen, dass die Anzahl extrahierter Dimensionen (als ein Indikator für die Faktorstruktur) bei der klassischen Faktorenanalyse mit dem Grad der Schiefe der manifesten Daten variierte, wobei gleichzeitig die Schiefe der latenten Werte manipuliert wurde (s. Kapitel 2.7). Dies spricht für einen generellen Zusammenhang zwischen der Schiefe der manifesten Daten, der Schiefe der latenten Fähigkeitswerte und der extrahierten Faktorstruktur. Die Annahme eines Zusammenhangs zwischen manifesten Datenverteilungen und Verteilungen der latenten Fähigkeitswerte wird auch durch das zentrale Grenzwerttheorem untermauert (s. Kapitel 2.6). Die durch die Marginalisierung bedingte Verteilung der manifesten Daten müsste mit zunehmender Anzahl der latenten Dimensionen einer Normalverteilung folgen. Umgekehrt implizieren schief verteilte manifeste Daten damit schief verteilte latente Werte auf wenigen Dimensionen.

Hypothese II: Es wird angenommen, dass mit zunehmender Rechtssteile der Fähigkeitswerte die Schätzgenauigkeit der Faktorenmodelle abnimmt.

Wie an den Formeln 2.4 (S. 16), 2.33 (S. 50) und 2.35 (S. 51) erkennbar, sind die geschätzten Fähigkeitswerte als gewichtete Summe von Zufallsvariablen darstellbar, und folgen damit (mit zunehmender Anzahl an manifesten Variablen  $z_j$ ) einer Normalverteilung (Grenzwerttheorem), und zwar unabhängig davon, wie die wahren Fähigkeitswerte verteilt sind. Damit ist anzunehmen, dass die Diskrepanz zwischen der Verteilung der geschätzten Fähigkeitswerte (tendenziell Normalverteilt) und der wahren Fähigkeitswerte (ggf. rechtssteil Verteilt) mit zunehmender Rechtssteile der wahren Fähigkeitswerte zunimmt, also die Schätzgenauigkeit der Faktorenmodelle abnimmt.

Hypothese III: Es wird angenommen, dass nicht alle Kriterien zur Beurteilung der Schätzgenauigkeit von Faktorenmodellen gleich sensibel auf das Ausmaß an Rechtssteile der Fähigkeitswerte reagieren.

Hypothese IIIa: Die Kriterien „Anzahl der extrahierten Faktoren“ und „Fit der geschätzten Ladungsmatrize“ sollten relativ robust gegenüber dem Ausmaß an Rechtssteile der Fähigkeitswerte sein.

Hypothese IIIb: Die „geschätzte Schiefe der Fähigkeitsverteilungen“ sollte relativ sensibel gegenüber dem Ausmaß an Rechtssteile der Fähigkeitswerte sein.

Für die Hypothese IIIa sprechen zum einen die empirischen Befunde, nach denen die Anzahl der wahren Dimensionen relativ robust unter verschiedenen Verteilungsbedingungen der Fähigkeitswerte geschätzt werden konnte, wenn die Parallelanalyse als Extraktionskriterium verwendet wurde (Weng & Cheng, 2005), zum anderen basieren die vorgestellten Extraktionskriterien und die Schätzung  $\hat{\mathbf{L}}$  auf der Eigenwertzerlegung von  $\mathbf{Z}$  bzw.  $\mathbf{R}$ , und wie an den theoretischen Ausführungen zur Eigenwertbestimmung zu erkennen (s. Kapitel 2.3.2.3), werden dabei keinerlei Verteilungsannahmen bezüglich  $\mathbf{Z}$  oder  $\mathbf{R}$  getroffen. Hypothese IIIb ergibt sich wiederum aus der Anwendung des zentralen Grenzwerttheorems auf die Schätzer  $\hat{\mathbf{F}}$ : Demnach ist für  $\hat{\mathbf{F}}$  zu erwarten, dass es unabhängig von der Verteilung von  $\mathbf{F}$  aufgrund von linearen Kombinationen tendenziell normalverteilt mit einer Schiefe von null ist.

Hypothese IV: Es wird angenommen, dass die Parallelanalyse nach Horn (1965) zu genaueren Ergebnissen hinsichtlich der zugrundeliegenden Faktorenanzahl führt als der Scree-Test (Cattell, 1966) oder das Kaiser-Guttman-Kriterium (Guttman, 1954; Kaiser & Dickman, 1959), wenn die Fähigkeitswerte rechtsteil verteilt sind.

Hypothese IV folgt direkt aus den empirischen Forschungsbefunden, die unter der Bedingung normalverteilter Fähigkeitswerte eine generelle Überlegenheit der Parallelanalyse gegenüber dem Scree-Test bzw. gegenüber dem Kaiser-Guttman-Kriterium aufgezeigt haben (s. Kapitel 2.7).

### 3.3 Mattson-Ansatz

Zur Beantwortung der Forschungsfragen und Überprüfung der parallel aufgestellten Hypothesen wird auf den Ansatz von Mattson (1997) zurückgegriffen. Dieser Ansatz stellt zum einen auf Basis des Strukturgleichungsmodells (Jöreskog, 1966, 1969; Browne, 1974; Jöreskog, 1978) den theoretischen Zusammenhang zwischen manifesten Datenverteilungen und latenten Fähigkeitsverteilungen heraus; zum anderen können auf Basis des Ansatzes manifeste Datenverteilungen für Simulationsstudien erzeugt werden.

Grundlage des Mattson-Ansatzes ist dabei die Zerlegung des  $p \times 1$  Vektors  $\mathbf{z}$  in eine  $p \times (p+k)$  Gewichtungsmatrix  $\mathbf{A}$  und einen  $(p+k) \times 1$  Vektor mit latenten Zufallsvariablen  $\boldsymbol{\nu}$

$$\mathbf{z} = \mathbf{A}\boldsymbol{\nu} \quad (3.1)$$

wobei

$$\boldsymbol{\nu} = \mathbf{T}\boldsymbol{\omega}$$

gilt, sodass

$$\mathbf{z} = \mathbf{AT}\boldsymbol{\omega}$$

resultiert.  $\mathbf{T}$  ist eine untere Dreiecksmatrix der Ordnung  $(p+k) \times (p+k)$  mit konstanten Elementen  $t_{11}, t_{21}, t_{22}, \dots, t_{(p+k)(p+k)}$ , und  $\boldsymbol{\omega}$  ein  $(p+k) \times 1$  Vektor von wechselseitig unabhängigen Zufallsvariablen  $\omega_1, \omega_2, \dots, \omega_{p+k}$  mit den ersten vier Momente  $\mu_{1i}, \mu_{2i}, \mu_{3i}, \mu_{4i}$  ( $i = 1, \dots, p+k$ ).

Von  $\boldsymbol{\omega}$  wird angenommen, dass es standardisiert ist (also  $\mu_{1i} = 0, \mu_{2i} = 1$ ), sodass

$$\mathbb{E}(\boldsymbol{\omega}\boldsymbol{\omega}') = \mathbf{I}$$

und

$$\begin{aligned} \mathbb{E}(\boldsymbol{\nu}) &= \mathbf{T}\mathbb{E}(\boldsymbol{\omega}) \\ &= \mathbf{0} \end{aligned}$$

mit

$$\begin{aligned} \mathbb{E}(\boldsymbol{\nu}\boldsymbol{\nu}') &= \mathbf{T}\mathbb{E}(\boldsymbol{\omega}\boldsymbol{\omega}')\mathbf{T}' \\ &= \mathbf{T}\mathbf{T}' \\ &= \boldsymbol{\Sigma} \end{aligned}$$

gilt, d.h. die Varianzkovarianzmatrix der latenten Variablen  $\boldsymbol{\nu}$  lässt sich vollständig aus  $\mathbf{T}$  reproduzieren.

Für die ersten vier Momente von  $\mathbf{z}$  folgt damit:

$$\begin{aligned} \mathbf{E}(\mathbf{z}) &= \mathbf{E}(\mathbf{A}\boldsymbol{\nu}) \\ &= \mathbf{A}\mathbf{T}\mathbf{E}(\boldsymbol{\omega}) \\ &= \mathbf{0} \end{aligned}$$

$$\begin{aligned} \mathbf{E}(\mathbf{z}\mathbf{z}') &= \mathbf{E}(\mathbf{A}\boldsymbol{\nu}\boldsymbol{\nu}'\mathbf{A}') \\ &= \mathbf{A}\mathbf{T}\mathbf{E}(\boldsymbol{\omega}\boldsymbol{\omega}')\mathbf{T}'\mathbf{A}' \\ &= \mathbf{A}\mathbf{T}\mathbf{T}'\mathbf{A}' \\ &= \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}' \end{aligned}$$

und

$$\mathbf{E}(z_i^3) = \sum_{o=1}^{p+k} \gamma_{io}^3 \mu_{3o}$$

bzw.

$$\mathbf{E}(z_i^4) = \sum_{o=1}^{p+k} \gamma_{io}^4 \mu_{4o} + 6 \sum_{o=2}^{p+k} \sum_{n=1}^{o-1} \gamma_{io}^2 \gamma_{in}^2$$

mit

$$\boldsymbol{\gamma}_i = (\mathbf{a}'_i \mathbf{T})'$$

und  $\mathbf{a}'_i$  als  $i$ -te Zeile von  $\mathbf{A}$ .

Die Schiefe  $\sqrt{\beta_{3i}}$  und Kurtosis  $\beta_{4i}$  von  $z_i$  sind als direkte Funktion der ersten vier Momente von  $z_i$  bzw. von  $\boldsymbol{\omega}$  darstellbar

$$\begin{aligned} \sqrt{\beta_{3i}} &= \frac{\mathbf{E}(z_i^3)}{[\mathbf{E}(z_i^2)]^{3/2}} = \frac{\sum_{o=1}^{p+k} \gamma_{io}^3 \mu_{3o}}{\sqrt{\sum_{o=1}^{p+k} \gamma_{io}^6}} \\ \beta_{4i} &= \frac{\mathbf{E}(z_i^4)}{[\mathbf{E}(z_i^2)]^2} = \frac{\sum_{o=1}^{p+k} \gamma_{io}^4 \mu_{4o} + 6 \sum_{o=2}^{p+k} \sum_{n=1}^{o-1} \gamma_{io}^2 \gamma_{in}^2}{\sum_{o=1}^{p+k} \gamma_{io}^4} \end{aligned}$$

und weil für extrem schiefe manifeste Daten  $\mathbf{E}(z_i^3) - [\mathbf{E}(z_i^2)]^{3/2} \gg 0$  im Allgemeinen und  $\mathbf{E}(z_i^4) - [\mathbf{E}(z_i^2)]^2 \gg 0$  im Besonderen gilt, kann für die Schiefe von  $z_i$  auch vereinfachend

$$\sqrt{\beta_{3i}} \approx \frac{\mathbf{E}(z_i^3)}{\mathbf{E}(z_i^2)} = \frac{\sum_{o=1}^{k+p} \gamma_{io}^3 \mu_{3o}}{\sum_{o=1}^{k+p} \gamma_{io}^2} \quad (3.2)$$

geschrieben werden, d.h. extrem schiefe manifeste Daten implizieren, dass wenigstens für ein  $\mu_{3o} \gg 0$  gilt, dass also wenigstens ein  $\omega_i$  eine deutliche Schiefe aufweist.

Um den Mattson-Ansatz nutzen zu können, muss das in Gleichung 3.1 aufgeführte Strukturmodell konkretisiert werden. Dabei scheint das explorative Faktorenmodell als am geeignetsten, weil auf Seiten der Strukturgleichung das Hauptfaktorenmodell identisch mit dem explorativen Faktorenmodell ist und das Hauptkomponentenmodell lediglich ein Spezialfall des explorativen Faktorenmodells darstellt (das Hauptkomponentenmodell resultiert aus dem explorativen Faktorenmodell, wenn  $\mathbf{E} = \mathbf{0}$  angenommen wird). Damit sind alle Ausführungen zum explorativen Faktorenmodell die sich nicht auf den zugrundeliegenden Schätzalgorithmus beziehen ohne großen Mehraufwand auch auf die Hauptfaktorenanalyse und die Hauptkomponentenmethode übertragbar.

Das explorative Faktorenmodell für  $\phi(\mathbf{0}, \mathbf{I})$  standardisierte manifeste Variablen lautet  $\mathbf{z} = \mathbf{L}\mathbf{f} + \mathbf{e}$ . Wenn  $\mathbf{A} = (\mathbf{L} \ \mathbf{I})$  mit  $\mathbf{I}$  von der Ordnung  $p \times p$ , und  $\boldsymbol{\nu}' = (\mathbf{f}' \ \mathbf{e}')$  angenommen wird, dann lässt sich das explorative Faktorenmodell auch in Form der Mattson-Gleichung darstellen

$$\begin{aligned} \mathbf{z} &= \mathbf{A}\boldsymbol{\nu} \\ &= (\mathbf{L} \ \mathbf{I}) \begin{pmatrix} \mathbf{f} \\ \mathbf{e} \end{pmatrix} \\ &= \mathbf{L}\mathbf{f} + \mathbf{e} \end{aligned} \tag{3.3}$$

mit:

$$\boldsymbol{\nu} = \begin{pmatrix} \mathbf{f} \\ \mathbf{e} \end{pmatrix} = \mathbf{T}\boldsymbol{\omega}$$

Wird im weiteren angenommen, dass  $\mathbf{E}(\mathbf{f}\mathbf{f}') = \mathbf{I}$ ,  $\mathbf{E}(\mathbf{f}\mathbf{e}')$  und  $\mathbf{E}(\mathbf{e}\mathbf{e}') = \mathbf{I}$  gilt, d.h. dass die Voraussetzungen des orthogonalen explorativen Faktorenmodells erfüllt sind, dann folgt für die Kovarianzmatrix  $\boldsymbol{\Sigma} = \mathbf{T}\mathbf{T}'$  der latenten Variablen  $\boldsymbol{\nu}$

$$\begin{aligned} \mathbf{T}\mathbf{T}' &= \mathbf{E}(\boldsymbol{\nu}\boldsymbol{\nu}') \\ &= \begin{pmatrix} \mathbf{E}(\mathbf{f}\mathbf{f}') & \mathbf{E}(\mathbf{f}\mathbf{e}') \\ \mathbf{E}(\mathbf{e}\mathbf{f}') & \mathbf{E}(\mathbf{e}\mathbf{e}') \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \\ &= \mathbf{I} \end{aligned}$$

sodass für  $\mathbf{T} = \mathbf{I}$  angenommen werden kann, denn  $\mathbf{T}\mathbf{T}' = \mathbf{I}\mathbf{I} = \mathbf{I} = \mathbf{\Sigma}$ .

Wenn  $\mathbf{T} = \mathbf{I}$  angenommen wird, dann folgt für  $\boldsymbol{\gamma}'_i = \mathbf{a}'_i\mathbf{T} = \mathbf{a}'_i = (l'_{ij} \ \mathbf{1}'_i)$  (wobei  $\mathbf{1}_i$  ein Vektor mit einer 1 an der  $i$ -ten Stelle ist, und sonst Nullen), d.h. die Schiefe  $\sqrt{\beta_{3i}}$  und Kurtosis  $\beta_{4i}$  von  $z_i$  hängt im orthogonalen explorativen Faktorenmodell (nur) von den entsprechenden Momenten der latenten Variablen  $\omega_i$  und den Ladungskoeffizienten  $l_{ij}$  ab:

$$\sqrt{\beta_{3i}} = \frac{\sum_{j=1}^k l_{ij}^3 \mu_{3j} + \mu_{3i}}{\sqrt{1 + \sum_{j=1}^k l_{ij}^6}} \quad (3.4)$$

$$\beta_{4i} = \frac{\sum_{j=1}^k l_{ij}^4 \mu_{4j} + \mu_{4i} + 6 \left( \sum_{o=2}^k \sum_{n=1}^{o-1} l_{io}^2 l_{in}^2 + \sum_{j=1}^k l_{ij} \right)}{1 + \sum_{j=1}^k l_{ij}^4} \quad (3.5)$$

Wird die in Gleichung 3.2 aufgeführte vereinfachende Formel für die Schiefe der manifesten Variablen auf das explorative Faktorenmodell übertragen, resultiert

$$\sqrt{\beta_{3i}} \approx \frac{\sum_{j=1}^k l_{ij} \mu_{3j} + \mu_{3i}}{1 + \sum_{j=1}^k l_{ij}^2}$$

d.h. auch im explorativen Faktorenmodell implizieren extrem schiefe manifeste Daten, dass wenigstens für ein  $\mu_3 \gg 0$  gilt, dass also wenigstens ein  $\omega_i$  eine deutliche Schiefe aufweist. Und weil  $\mathbf{T} = \mathbf{I}$  angenommen wird, gilt auch

$$\boldsymbol{\nu} = (\mathbf{f} \ \mathbf{e}) = \mathbf{T}\boldsymbol{\omega} = \boldsymbol{\omega} \quad (3.6)$$

d.h. die ersten  $\omega_j$  ( $j = 1, \dots, k$ ) entsprechen den Komponenten  $f_j$  und die letzten  $\omega_i$  ( $i = 1, \dots, p$ ) den Komponenten  $e_i$  bzw. die ersten  $\mu_{3j}$  bezeichnen die dritten Momente der Komponenten  $f_j$  und die letzten  $\mu_{3i}$  die dritten Momente der Komponenten  $e_i$ . Wird (wie üblich) angenommen, dass  $e_i \sim N_i(0, 1)$ , dann gilt auch  $\forall_i \mu_{3i} = 0$  sodass

$$\sqrt{\beta_{3i}} \approx \frac{\sum_{j=1}^k l_{ij} \mu_{3j}}{1 + \sum_{j=1}^k l_{ij}^2} \quad (3.7)$$

resultiert, d.h. extreme Schiefen in den manifesten Daten implizieren im Falle des orthogonalen explorativen Faktorenmodells, dass zumindest für ein  $\mu_{3j} \gg$



0 gilt, dass also die Werte auf zumindest einer latenten Fähigkeitsdimension schief verteilt sind.

Der aufgezeigte Zusammenhang zwischen manifester und latenter Verteilungsform kann auf das Hauptkomponentenmodell und die Hauptfaktorenanalyse übertragen werden, denn in der Hauptkomponentenmethode gilt  $e_i = 0$ , sodass  $\boldsymbol{\nu} = \boldsymbol{f} = \boldsymbol{T}\boldsymbol{\omega} = \boldsymbol{\omega}$  resultiert, was bedeutet, dass Gleichung 3.7 unverändert für die Hauptkomponentenmethode übernommen werden kann, und die Hauptfaktorenanalyse basiert auf demselben Strukturmodell wie die explorativen Faktorenanalyse, und Gleichung 3.7 wurde aus dem Strukturmodell der explorativen Faktorenanalyse entwickelt, d.h. Gleichung 3.7 gilt auch für die Hauptfaktorenanalyse.

In Bezug auf die betrachteten klassischen Faktorenmodelle kann also davon ausgegangen werden, dass schief verteilte Daten schief verteilte Fähigkeitswerte auf zumindest einer latenten Dimension implizieren. Aufgrund dieser theoretischen Herleitung kann Hypothese I zumindest partiell bestätigt werden: Es kann angenommen werden, dass schief verteilte Daten bestimmte Verteilungen (nämlich schiefe Verteilungen) der Fähigkeitswerte implizieren. Allerdings bleibt unklar, ob schief verteilte Daten auch bestimmte Faktorstrukturen (etwa im Sinne der Anzahl extrahierbarer Faktoren) implizieren. Ungewiss ist auch der Status der Hypothesen II bis IV, da sich aus der theoretischen Herleitung keine Aussagen über die Gültigkeit dieser Hypothesen ableiten lassen. Deswegen wurden diese Hypothesen in einer Simulationsstudie überprüft.

### 3.4 Simulationsdesign

Ziel der Simulationsstudie war es, den Einfluss von rechtssteil verteilten Fähigkeitswerten auf die Schätzgenauigkeit von klassischen Faktorenanalysen abzuschätzen. Dabei sollte auch überprüft werden, ob die Schätzgütekriterien „geschätzte Schiefe der Fähigkeitswerte“ und „Fit der geschätzten Ladungsmatrix“ sensibler auf das Ausmaß an Rechtssteile der Fähigkeitswerte reagieren als das Schätzgütekriterium „Anzahl der extrahierten Faktoren“, und ob die Parallelanalyse nach Horn (1965) die zugrundeliegende Faktorenanzahl bei rechtssteil verteilten Fähigkeitswerten genauer schätzt als das Kaiser-Guttman-Kriterium (Guttman, 1954; Kaiser & Dickman, 1959) oder der Scree-Test (Cattell, 1966).

Zur Überprüfung der Hypothesen wurden manifeste Datensätze auf Basis des Mattson-Ansatz und eines vorgegebenen orthogonalen explorativen Faktorenmodells konstruiert, und diese Datensätze dann mit den in Kapitel 2.3

vorgestellten Faktorenmodellen unter Verwendung der verschiedenen Extraktionskriterien analysiert. In der Simulationsstudie wurden also manifeste Datensätze auf der Basis eines wahren (bekannten) Faktorenmodells konstruiert, und dann überprüft, wie genau die vorgestellten Faktorenmodelle das wahre Faktorenmodell schätzen.

Wie anhand von Gleichung 3.1 zu erkennen, mussten bestimmte Festlegungen getroffen werden, um manifeste Datensätze  $\mathbf{Z}$  über den Mattson-Ansatz konstruieren zu können:

1. Die Struktur von  $\mathbf{A}$  und  $\boldsymbol{\nu}$  und damit das wahre Strukturmodell für die Simulationsstudie musste spezifiziert werden.
2. Es musste die Anzahl der manifesten Variablen  $z_1, z_2, \dots, z_p$  und die Anzahl der latenten Variablen  $\nu_1, \nu_2, \dots, \nu_{(p+k)}$  und damit  $p$  und  $k$  vorgegeben werden.
3. Die konkreten Werte für  $\mathbf{A}$  und  $\boldsymbol{\nu}$  mussten bestimmt werden, wobei die Werte von  $\boldsymbol{\nu}$  eine anzugebende Schiefe bzw. Kurtosis aufweisen sollten.
4. Es musste die Stichprobengröße  $n$  festgelegt werden.

Ad 1.: Als wahres Strukturmodell wurde ein orthogonales exploratives Faktorenmodell angenommen, d.h.  $\mathbf{A} = (\mathbf{L} \ \mathbf{I})$ ,  $\boldsymbol{\nu}' = (\mathbf{f}' \ \mathbf{e}')$  und  $\mathbf{T} = \mathbf{I}$ .

Ad 2.: Die Anzahl der manifesten Variablen wurde aus ökonomischen Gründen auf  $p = 24$  fixiert. Um die Anzahl der Faktoren zu bestimmen, wurden die technischen Berichte von PIRLS (M. Martin et al., 2007) und PISA (OECD, 2003) ausgewertet, wobei die Anzahl der latenten Dimensionen für die Skalierung der Leistungsdaten als Orientierungsgrößen für die Faktorenanzahl in der Simulationsstudie verwendet wurden. Während in PIRLS 2006 bis zu vier latente Dimensionen angenommen wurden, basierte die Skalierung der Leistungsdaten in PISA 2003 auf sieben Dimensionen. Für die Simulationsstudie wurden deswegen zwei unterschiedliche Anzahlen der Faktoren simuliert: In der ersten Bedingung wurde  $k = 4$  und in der zweiten Bedingung  $k = 8$  angenommen ( $k = 8$  wurde anstelle von  $k = 7$  gewählt, damit auf jedem Faktor gleich viele manifeste Variablen laden konnten).

Ad 3.: In Anlehnung an die bisher vorliegenden empirischen Forschungsarbeiten (s. Kapitel 2.7) wurde für  $\mathbf{L}$  eine Einfachstruktur simuliert, d.h. jede manifeste Variable lud auf nur einem Faktor. Dabei wurde angenommen (wie in der empirischen Bildungsforschung üblich), dass einige Variablen relativ hoch auf dem dazugehörigen Faktor laden, während andere Variablen eher niedrige Ladungen aufweisen, und dass einige Faktoren die dazugehörigen

manifesten Variablen relativ gut erklären während die übrigen Faktoren die dazugehörigen manifesten Variablen nicht so gut erklären. Im Ergebnis dieser Überlegungen wurden die folgenden beiden Ladungsmatrizen (jeweils eine pro Bedingung  $k = 4$  vs.  $k = 8$ ) zur Erzeugung von Datenmatrizen verwendet:

$$\mathbf{L}_4 = \begin{pmatrix} .9 & 0 & 0 & 0 \\ .8 & 0 & 0 & 0 \\ .7 & 0 & 0 & 0 \\ .6 & 0 & 0 & 0 \\ .5 & 0 & 0 & 0 \\ .4 & 0 & 0 & 0 \\ 0 & .8 & 0 & 0 \\ 0 & .7 & 0 & 0 \\ 0 & .6 & 0 & 0 \\ 0 & .5 & 0 & 0 \\ 0 & .4 & 0 & 0 \\ 0 & .3 & 0 & 0 \\ 0 & 0 & .6 & 0 \\ 0 & 0 & .6 & 0 \\ 0 & 0 & .5 & 0 \\ 0 & 0 & .4 & 0 \\ 0 & 0 & .4 & 0 \\ 0 & 0 & .3 & 0 \\ 0 & 0 & 0 & .6 \\ 0 & 0 & 0 & .5 \\ 0 & 0 & 0 & .5 \\ 0 & 0 & 0 & .4 \\ 0 & 0 & 0 & .3 \\ 0 & 0 & 0 & .3 \end{pmatrix} \quad \mathbf{L}_8 = \begin{pmatrix} .9 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ .8 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ .7 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & .8 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & .8 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & .7 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & .8 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & .7 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & .6 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & .7 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & .7 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & .7 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & .7 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & .6 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & .6 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & .6 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & .6 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & .5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & .5 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & .4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & .4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & .4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & .4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & .3 \end{pmatrix}$$

Wie aus Gleichung 3.6 ersichtlich, können Werte für  $\boldsymbol{\nu}' = (\mathbf{f}' \ \mathbf{e}')$  mit bestimmten Verteilungsparametern über die Festlegung von Werten und entsprechenden Verteilungsparametern für  $\boldsymbol{\omega}' = (\mathbf{f}' \ \mathbf{e}')$  spezifiziert. Um eine empirische Basis für die Auswahl von Schiefe- und Kurtosisparametern für  $\boldsymbol{\omega}$  zu erhalten, wurden die freigegeben Leistungsitems aus PIRLS 2006 analysiert, wobei der Datensatz für Deutschland gewählt wurde (IEA, 2007). In diesem Datensatz sind die Antworten der deutschen Testteilnehmer (über alle Aufgaben hinweg etwa 7899 Schülerinnen und Schüler der vierten Klasse) auf 126 Fragen kodiert. Bevor die Schiefe- und Kurtosiswerte für diese 126 Items bestimmt wurden, wurden Mehrfachwahlaufgaben dichotomisiert, wobei 0 für eine falsche Antwort und 1 für eine richtige Antwort stand.

Die durchschnittliche Schiefe der 126 Items liegt bei  $\bar{b}_3 = -1.00$  mit einem Standardfehler von  $s_e = 1.20$ . Der minimale Wert liegt bei  $b_{3min} = -5.75$  und der maximale Wert bei  $b_{3max} = 1.59$ . Ungefähr 20 Prozent der Aufgaben weisen eine Schiefe von  $b_3 = -0.074$  oder höher auf, und gut 50 Prozent der Items haben eine Schiefe von  $b_3 = -0.83$  oder kleiner. Die durchschnittliche Kurtosis der 126 Aufgaben liegt bei  $\bar{b}_4 = 0.60$  mit einem Standardfehler von  $s_e = 4.5$ . Der minimale Wert liegt bei  $b_{4min} = -2.00$  und der maximale Wert bei  $b_{4max} = 31.05$ . Rund 20 Prozent der Aufgaben weisen eine Kurtosis von  $b_4 = 1.33$  oder höher auf, und etwa 50 Prozent der Items haben eine Kurtosis von  $b_4 = -0.73$  oder kleiner.

Auf der Basis dieser Befunde wurde entschieden, drei Bedingungen für die Verteilung von  $\omega$  zu simulieren: In der ersten Bedingung (normalverteilt) ist  $f_m$  ( $m = 1, \dots, k$ ) normalverteilt mit  $\mu_{1m} = 0$ ,  $\mu_{2m} = 1$ ,  $\mu_{3m} = 0$  und  $\mu_{4m} = 3$ , in der zweiten Bedingung (leicht rechtssteil) ist  $f_m$  leicht rechtssteil verteilt mit  $\mu_{1m} = 0$ ,  $\mu_{2m} = 1$ ,  $\mu_{3m} = -0.20$  und  $\mu_{4m} = 3$ , und in der dritten Bedingung (stark rechtssteil) ist  $f_m$  stark rechtssteil verteilt mit  $\mu_{1m} = 0$ ,  $\mu_{2m} = 1$ ,  $\mu_{3m} = -2$  und  $\mu_{4m} = 9$ . Für die Fehlerterme, d.h. für  $e_h$  ( $h = k + 1, \dots, k + p$ ) wurde stets eine Normalverteilung mit  $\mu_{1h} = 0$ ,  $\mu_{2h} = 1$ ,  $\mu_{3h} = 0$  und  $\mu_{4h} = 3$  angenommen.

Die erwartete Schiefe und Kurtosis für die manifesten Variablen  $z_i$  unter den drei Verteilungsbedingungen der latenten Fähigkeitswerte können auf der Basis der Formel 3.4 und Gleichung 3.5 bestimmt werden, und sind in Tabelle 3.1 (Seite 80) und Tabelle 3.2 (Seite 81) abgetragen. Wie zu sehen, variiert die erwartete Schiefe für die manifesten Variablen in der Bedingung „leicht rechtssteil“ zwischen  $\sqrt{\beta_3} = -0.060$  und  $\sqrt{\beta_3} = -0.005$ . Dies entspricht in etwa den Schiefewerten von rund 20 Prozent der analysierten Aufgaben aus PIRLS. In der Bedingung „stark rechtssteil“ variiert die erwartete Schiefe der manifesten Variablen zwischen  $\sqrt{\beta_3} = -0.599$  und  $\sqrt{\beta_3} = -0.047$ ; ein Wertebereich, in den die Schiefewerte von rund 30 Prozent der analysierten Aufgaben aus PIRLS fallen. Insgesamt deckt die Simulation also etwa 50 Prozent der in PIRLS 2006 zu beobachtenden Schiefewerte ab. Die erwartete Kurtosis für die manifesten Variablen liegt in nahezu allen Fällen bei dem Wert für die Normalverteilung  $\beta_4 = 3$ . Nur wenn der erwartete Wert für die Schiefe extrem niedrig ist, erreicht die erwartete Kurtosis einen Wert von ungefähr  $\beta_4 \approx 4$ .

Wie Reinartz, Echambadi und Chin (2002) zeigen konnten, weichen die nach der Simulation von manifesten Daten zu beobachtenden Schiefe- und Kurtosiswerte von  $z_i$  von den erwarteten Werten praktisch dann nicht ab, wenn die generalisierte Lambda-Verteilung nach Ramberg, Tadikamalla, Dudewicz und Mykytka (1979) genutzt wird, um Werte für  $\omega$  zu generieren. Deswegen

Tabelle 3.1: Erwartete Schiefe  $\sqrt{\beta_{3i}}$  und Kurtosis  $\beta_{4i}$  für  $z_i$  ( $k_1 = 4$ )

Manifeste Variable $z_i$	Verteilung der latenten Variablen $f_m$					
	Normal <sup>a</sup> $\sqrt{\beta_{3i}}$	$\beta_{4i}$	Leicht rechtssteil <sup>b</sup> $\sqrt{\beta_{3i}}$	$\beta_{4i}$	Stark rechtssteil <sup>c</sup> $\sqrt{\beta_{3i}}$	$\beta_{4i}$
$z_1$	0	3	-0.060	3	-0.599	4.202
$z_2$	0	3	-0.049	3	-0.488	3.914
$z_3$	0	3	-0.038	3	-0.377	3.649
$z_4$	0	3	-0.027	3	-0.272	3.420
$z_5$	0	3	-0.018	3	-0.179	3.240
$z_6$	0	3	-0.010	3	-0.102	3.114
$z_7$	0	3	-0.049	3	-0.488	3.914
$z_8$	0	3	-0.038	3	-0.377	3.649
$z_9$	0	3	-0.027	3	-0.272	3.420
$z_{10}$	0	3	-0.018	3	-0.179	3.240
$z_{11}$	0	3	-0.010	3	-0.102	3.114
$z_{12}$	0	3	-0.005	3	-0.047	3.041
$z_{13}$	0	3	-0.027	3	-0.272	3.420
$z_{14}$	0	3	-0.027	3	-0.272	3.420
$z_{15}$	0	3	-0.018	3	-0.179	3.240
$z_{16}$	0	3	-0.010	3	-0.102	3.114
$z_{17}$	0	3	-0.010	3	-0.102	3.114
$z_{18}$	0	3	-0.005	3	-0.047	3.041
$z_{19}$	0	3	-0.027	3	-0.272	3.420
$z_{20}$	0	3	-0.018	3	-0.179	3.240
$z_{21}$	0	3	-0.018	3	-0.179	3.240
$z_{22}$	0	3	-0.010	3	-0.102	3.114
$z_{23}$	0	3	-0.005	3	-0.047	3.041
$z_{24}$	0	3	-0.005	3	-0.047	3.041

<sup>a</sup>:  $\mu_{1m} = 0$ ,  $\mu_{2m} = 1$ ,  $\mu_{3m} = 0$  und  $\mu_{4m} = 3$

<sup>b</sup>:  $\mu_{1m} = 0$ ,  $\mu_{2m} = 1$ ,  $\mu_{3m} = -0.20$  und  $\mu_{4m} = 3$

<sup>c</sup>:  $\mu_{1m} = 0$ ,  $\mu_{2m} = 1$ ,  $\mu_{3m} = -2$  und  $\mu_{4m} = 9$

wurden die konkreten Werte in der Simulation für  $\omega' = (\mathbf{f}' \ \mathbf{e}')$  über die generalisierte Lambda-Verteilung erzeugt. Gemäß dieser Verteilung kann

$$\omega_o = \lambda_1 + \frac{u_o^{\lambda_3} - (1 - u_o)^{\lambda_4}}{\lambda_2}$$

geschrieben werden, wobei  $u_o$  einer Gleichverteilung mit  $(0, 1)$  folgt,  $\lambda_1$  der Lokationsparameter,  $\lambda_2$  der Skalierungsparameter, und  $\lambda_3$  sowie  $\lambda_4$  die Formparameter sind ( $o = 1, \dots, (k + p)$ ).

Damit  $\omega$  einer Verteilung mit den gewünschten Eigenschaften (normalverteilt, leicht rechtssteil, stark rechtssteil) folgt, müssen für die Parameter  $\lambda_1$ ,  $\lambda_2$ ,  $\lambda_3$  und  $\lambda_4$  entsprechende Werte festgelegt werden. Ramberg et al. (1979) haben für unterschiedliche Wertekombinationen von  $\mu$  die dazugehörigen Werte für  $\lambda_1$ ,  $\lambda_2$ ,  $\lambda_3$  und  $\lambda_4$  tabelliert. Die in der Simulation verwendeten Koeffizienten sind in Tabelle 3.3 (Seite 82) aufgeführt.

Tabelle 3.2: Erwartete Schiefe  $\sqrt{\beta_{3i}}$  und Kurtosis  $\beta_{4i}$  für  $z_i$  ( $k_2 = 8$ )

Manifeste Variable $z_i$	Verteilung der latenten Variablen $f_m$					
	Normal <sup>a</sup> $\sqrt{\beta_{3i}}$	$\beta_{4i}$	Leicht rechtssteil <sup>b</sup> $\sqrt{\beta_{3i}}$	$\beta_{4i}$	Stark rechtssteil <sup>c</sup> $\sqrt{\beta_{3i}}$	$\beta_{4i}$
$z_1$	0	3	-0.060	3	-0.599	4.202
$z_2$	0	3	-0.049	3	-0.488	3.914
$z_3$	0	3	-0.038	3	-0.377	3.649
$z_4$	0	3	-0.049	3	-0.488	3.914
$z_5$	0	3	-0.049	3	-0.488	3.914
$z_6$	0	3	-0.038	3	-0.377	3.649
$z_7$	0	3	-0.049	3	-0.488	3.914
$z_8$	0	3	-0.038	3	-0.377	3.649
$z_9$	0	3	-0.027	3	-0.272	3.420
$z_{10}$	0	3	-0.038	3	-0.377	3.649
$z_{11}$	0	3	-0.038	3	-0.377	3.649
$z_{12}$	0	3	-0.038	3	-0.377	3.649
$z_{13}$	0	3	-0.038	3	-0.377	3.649
$z_{14}$	0	3	-0.027	3	-0.272	3.420
$z_{15}$	0	3	-0.027	3	-0.272	3.420
$z_{16}$	0	3	-0.027	3	-0.272	3.420
$z_{17}$	0	3	-0.027	3	-0.272	3.420
$z_{18}$	0	3	-0.018	3	-0.179	3.240
$z_{19}$	0	3	-0.018	3	-0.179	3.240
$z_{20}$	0	3	-0.010	3	-0.102	3.114
$z_{21}$	0	3	-0.010	3	-0.102	3.114
$z_{22}$	0	3	-0.010	3	-0.102	3.114
$z_{23}$	0	3	-0.010	3	-0.102	3.114
$z_{24}$	0	3	-0.005	3	-0.047	3.041

<sup>a</sup>:  $\mu_{1m} = 0, \mu_{2m} = 1, \mu_{3m} = 0$  und  $\mu_{4m} = 3$

<sup>b</sup>:  $\mu_{1m} = 0, \mu_{2m} = 1, \mu_{3m} = -0.20$  und  $\mu_{4m} = 3$

<sup>c</sup>:  $\mu_{1m} = 0, \mu_{2m} = 1, \mu_{3m} = -2$  und  $\mu_{4m} = 9$

Mit Bezug auf  $u_o$  haben Reinartz et al. (2002) zeigen können, dass die Prozedur RANUNI der Statistiksoftware SAS zu nahezu unverzerrten Ergebnissen bezüglich der Verteilung von  $\omega$  führt. Deswegen wurden alle simulierten Datensätze mit dem Programm SAS 9.0 konstruiert. Im Anhang A (Seite 104) findet sich als Beispiel für eine Syntax zur Datengenerierung die Syntax für die Bedingung normalverteilter latenter  $f_m$  bei vier Faktoren. Um die Übersichtlichkeit zu wahren, wurde die vollständige implementierte Syntax für die Wiedergabe in diesem Buch so gekürzt, dass sie nur eine Beobachtung und nur einen Datensatz erzeugt. Die Erweiterung dieser im Anhang aufgeführten Basissyntax für mehr als eine Beobachtung und mehr als einen Datensatz erfolgte durch das hinzufügen von Schleifen zur Basissyntax.

Ad 4.: Für die Stichprobengröße wurden zwei Bedingungen simuliert. In der ersten Bedingung betrug die Stichprobengröße  $N = 200$ , und in der zweiten

Tabelle 3.3: Gewichtskoeffizienten der generalisierten Lambdaverteilung nach Ramberg et al. (1979)

Bedingung	Gewichtskoeffizienten der Lambdaverteilung			
	$\lambda_1$	$\lambda_2$	$\lambda_3$	$\lambda_4$
Normal <sup>a</sup>	0	0.1974	0.1349	0.1349
Leicht rechtssteil <sup>b</sup>	0.237	0.193	0.1672	0.1065
Stark rechtssteil <sup>c</sup>	0.993	-0.001 081	-0.001 076	-0.000 040 7

<sup>a</sup>:  $\mu_{1m} = 0$ ,  $\mu_{2m} = 1$ ,  $\mu_{3m} = 0$  und  $\mu_{4m} = 3$

<sup>b</sup>:  $\mu_{1m} = 0$ ,  $\mu_{2m} = 1$ ,  $\mu_{3m} = -0.20$  und  $\mu_{4m} = 3$

<sup>c</sup>:  $\mu_{1m} = 0$ ,  $\mu_{2m} = 1$ ,  $\mu_{3m} = -2$  und  $\mu_{4m} = 9$

Bedingung  $N = 600$ . Damit wird sowohl der Fall einer eher kleinen Stichprobe abgedeckt, als auch die Möglichkeit geboten, das Grenzwertverhalten der Auswertungsstatistiken zu beobachten.

Insgesamt besteht das Simulationsdesign also aus  $2 \times 3 \times 2 = 12$  Bedingungen, wobei für jede Bedingung jeweils 100 Datensätze simuliert wurden. Tabelle 3.4 gibt einen Überblick über das Simulationsdesign.

Tabelle 3.4: Simulationsdesign im Überblick

$N$	$k$	Verteilung der latenten Fähigkeitswerte		
		Normal	Leicht rechtssteil	Stark rechtssteil
200	4	100	100	100
	8	100	100	100
600	4	100	100	100
	8	100	100	100

### 3.5 Auswertungsmethoden

An allen 1.200 Datensätzen wurde eine Hauptkomponentenanalyse, eine explorative Faktorenanalyse mit Maximum-Likelihood-Schätzung, und eine Hauptfaktorenanalyse durchgeführt, um  $\mathbf{F}$  und  $\mathbf{L}$  zu schätzen. Zur Bestimmung der zu extrahierenden Dimensionen wurden bei allen Analysen die drei Extraktionskriterien (Kaiser-Guttman-Kriterium, Scree-Test, Parallelanalyse) angewendet.

Als Kriterien für die Schätzgüte der Faktorenmodelle wurde die „Anzahl der extrahierten Dimensionen“ die „geschätzte Schiefe der Fähigkeitswerte“ und der „Fit der geschätzten Ladungsmatrize“ gewählt. Die beiden letzten Kriterien wurden dabei unter der Bedingung bestimmt, dass die wahre Faktorenanzahl bekannt ist, d.h. die geschätzten  $\mathbf{F}$  und  $\mathbf{L}$  enthielten jeweils  $k$  Spalten, und vor den weiteren Analysen wurde eine Varimax-Rotation durchgeführt, um die Einfachstruktur von  $\mathbf{L}$  zu modulieren.

Zur Bestimmung der „geschätzten Schiefe der Fähigkeitswerte“ wurde für die Werte jeder Spalte der 1.200 geschätzten  $\mathbf{F}$  Matrizen der Schiefewert berechnet. Um den „Fit der geschätzten Ladungsmatrize“ zu quantifizieren, wurde pro simulierter Bedingung für jeden der 100 Datensätze ein Abweichungskoeffizient  $AQ_{Li}$  berechnet

$$AQ_{Li} = \frac{\mathbf{1}'_{24} \text{abs} \left[ \left( \hat{\mathbf{L}}_i - \mathbf{L} \right) \right]_k \mathbf{1}}{kp}$$

worin  $\hat{\mathbf{L}}_i$  die geschätzte varimax-rotierte Ladungsmatrize des  $i$ -ten Datensatz,  $\mathbf{L}$  die wahre Ladungsmatrize,  $\mathbf{1}$  ein Vektor mit lauter Einsen und  $\text{abs}[\cdot]$  eine Funktion ist, die den Absolutwert von jedem Element einer Matrix zurückgibt, also beispielsweise:

$$\mathbf{A} = \begin{pmatrix} -1 & 2 \\ 3 & -4 \end{pmatrix} \quad \text{abs}[\mathbf{A}] = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$$

Aus den 100 Abweichungskoeffizienten pro simulierter Bedingung wurde ein durchschnittlicher Abweichungskoeffizient

$$\overline{AQ_L} = \sum_{i=1}^{100} \frac{AQ_{Li}}{100}$$

und eine Varianz der Abweichungskoeffizienten

$$\sigma_{AQ_L}^2 = \sum_{i=1}^{100} \frac{(AQ_{Li} - \overline{AQ_L})^2}{100 - 1}$$

bestimmt.

Der durchschnittliche Abweichungskoeffizient ist Null, wenn pro simulierter Bedingung alle 100 geschätzten Ladungsmatrizen der wahren Ladungsmatrize entsprechen, und er wird umso größer, je mehr die geschätzten Ladungsmatrizen von der wahren Ladungsmatrize abweichen. Darüber hinaus entspricht



der durchschnittliche Abweichungskoeffizient der durchschnittlichen absoluten Differenz zwischen den geschätzten Ladungskoeffizienten und den wahren Ladungskoeffizienten, denn:

$$\begin{aligned}
 \overline{AQ}_L &= \sum_{i=1}^{100} \frac{AQ_{Li}}{100} \\
 &= \sum_{i=1}^{100} \frac{\mathbf{1}'_{24} \text{abs} \left[ \left( \hat{\mathbf{L}}_i - \mathbf{L} \right) \right]_k \mathbf{1}}{100kp} \\
 &= \sum_{i=1}^{100} \frac{\sum_{p=1}^{24} \sum_{q=1}^k |\hat{l}_{ipk} - l_{pk}|}{100kp}
 \end{aligned}$$

# Kapitel 4

## Ergebnisse

### 4.1 Anzahl der extrahierten Dimensionen

In Abbildung 4.1 auf der folgenden Seite ist die relative Häufigkeit abgebildet, mit der eine bestimmte Anzahl an Dimensionen extrahiert wurde, wenn die Stichprobengröße  $N = 200$  und die wahre Faktorenanzahl  $k = 4$  war. Bei Verwendung des Kaiser-Guttman-Kriteriums wurde die wahre Faktorenanzahl unter dieser Bedingung überschätzt (Hauptkomponentenmethode) oder (tendenziell) unterschätzt (exploratives Faktorenmodell bzw. Hauptfaktorenanalyse). Mit dem Scree-Test wurden zwar mehrheitlich vier Dimensionen extrahiert, die Extraktionsergebnisse streuten aber erheblich, was auf den uneindeutigen und damit schwer zu interpretierenden Funktionsverlauf der Eigenwertgrafik verweist. Durch Anwendung der Parallelanalyse konnte die Streuung der Extraktionsergebnisse gegenüber dem Scree-Tests deutlich reduziert werden, sodass mit dieser Methode im Allgemeinen die wahre Faktorenanzahl von  $k = 4$  am genauesten geschätzt werden konnte.

Ein Zusammenhang zwischen der Verteilung der latenten Fähigkeitswerte (Normal, Leicht rechtssteil, Stark rechtssteil) und der Anzahl extrahierter Faktoren ist unter der Bedingung  $N = 200$  und  $k = 4$  nicht zu erkennen, wird einmal von der leichten Tendenz zur Überschätzung der wahren Faktorenanzahl (Kaiser-Guttman-Kriterium und Hauptkomponentenmethode; Scree-Test und exploratives Faktorenmodell bzw. Hauptfaktorenanalyse) bzw. zur Unterschätzung der wahren Faktorenanzahl (Kaiser-Guttman-Kriterium und exploratives Faktorenmodell; Scree-Test und Hauptkomponentenmethode, Parallelanalyse) bei stark rechtssteil verteilten Fähigkeitswerten abgesehen.

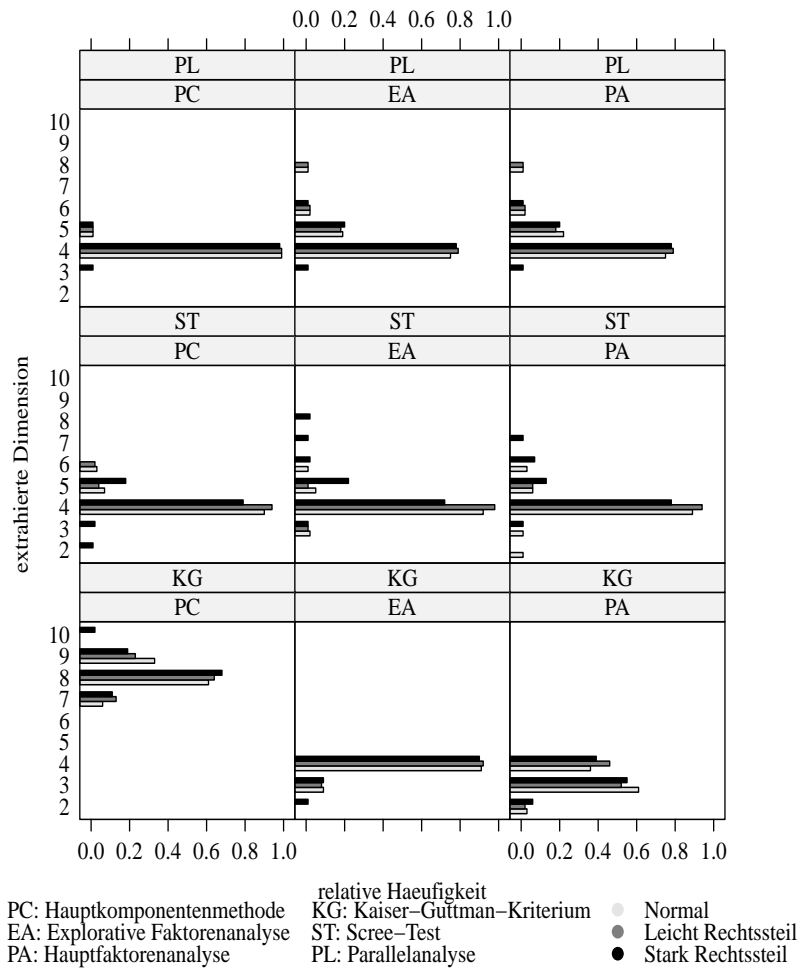


Abbildung 4.1: Anzahl der extrahierten Dimensionen ( $N = 200$ ;  $k = 4$ )

Wenn die Stichprobengröße von  $N = 200$  auf  $N = 600$  erhöht wurde, dann zeigte sich ein gewisses Konvergenzverhalten der Extraktionsmethoden (Abbildung 4.2, Seite 87). Die Extraktionsergebnisse streuten deutlich weniger, und mit dem Kaiser-Guttman-Kriterium wurde die wahre Faktorenanzahl wiederum am ungenauesten geschätzt. Im Unterschied zu kleinen Stichproben scheint bei großen Stichproben neben der Parallelanalyse auch der Scree-Test zu genauen Schätzungen der wahren Faktorenanzahl zu führen, und die Abhängigkeit der Schätzgenauigkeit in Bezug auf die Anzahl der extrahierten Dimensionen von der Verteilung der latenten Fähigkeitswerte scheint sich aufzulösen.

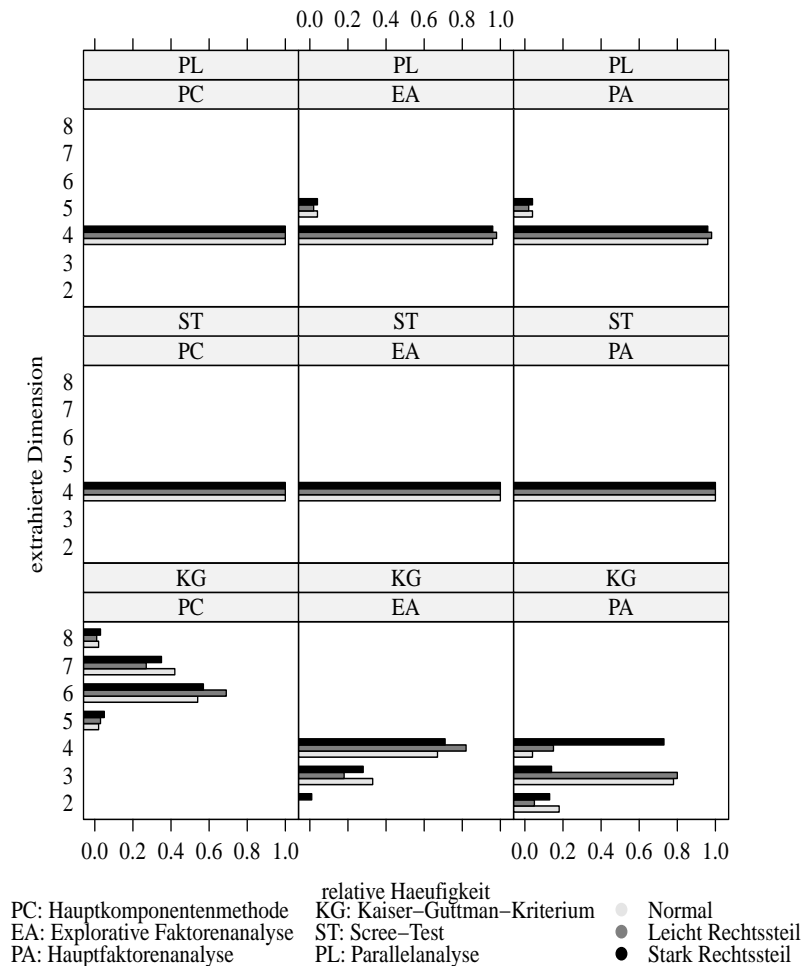


Abbildung 4.2: Anzahl der extrahierten Dimensionen ( $N = 600; k = 4$ )

Generell lassen sich die beobachteten Ergebnisse auf den Fall von  $k = 8$  übertragen. Auch hier führte das Kaiser-Guttman-Kriterium relativ häufig zu einer Über- bzw. Unterschätzung der wahren Faktorenanzahl, streuten die Extraktionsergebnisse des Scree-Tests erheblich und schätzte die Parallelanalyse die wahre Faktorenanzahl am genauesten (Abbildung 4.3, Seite 88). Ein Einfluss der Verteilung der latenten Fähigkeitswerte auf die Anzahl der extrahierten Dimensionen konnte nicht beobachtet werden.

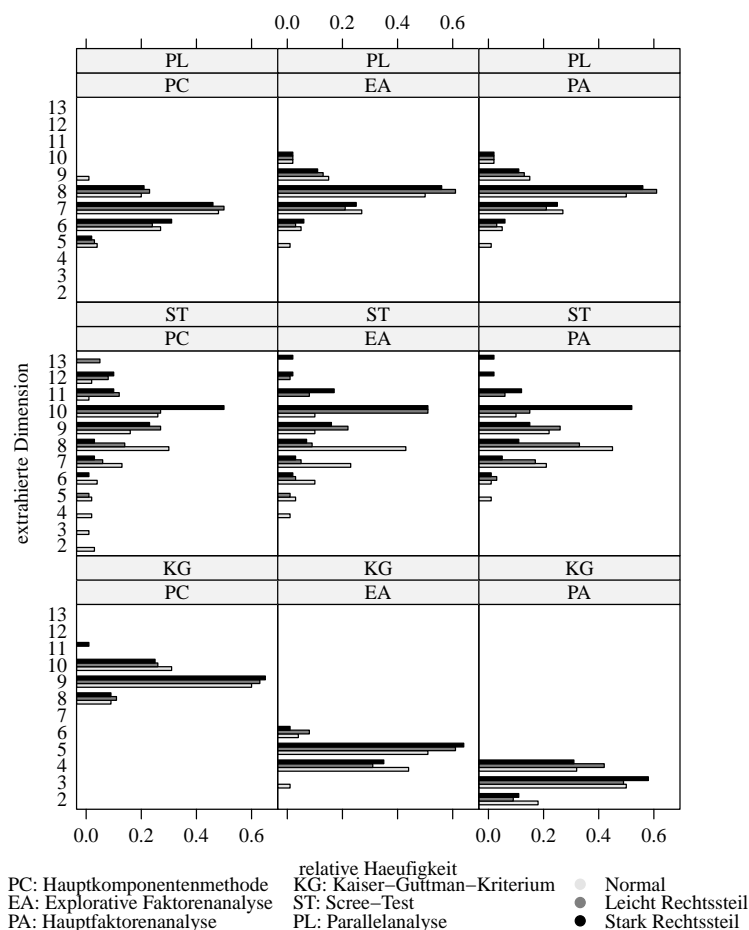


Abbildung 4.3: Anzahl der extrahierten Dimensionen ( $N = 200$ ;  $k = 8$ )

Die Erhöhung der Stichprobe von  $N = 200$  auf  $N = 600$  brachte eine Reduzierung der Streuung der Extraktionsergebnisse mit sich (Abbildung 4.4, Seite 89). Prinzipiell konnte aber nur die Parallelanalyse unter allen Faktorenmodellen die wahre Faktorenanzahl relativ genau schätzen. Ein Zusammenhang zwischen der Verteilung der latenten Fähigkeitswerte und der Anzahl extrahierter Faktoren konnte auch unter dieser Bedingung nicht erkannt werden. Insgesamt kann damit festgestellt werden, dass das Kriterium „Anzahl der extrahierten Faktoren“ relativ robust gegenüber dem Ausmaß an Rechtsteile der Fähigkeitswerte ist, und dass die Parallelanalyse nach Horn (1965) im Allgemeinen zu genaueren Ergebnissen hinsichtlich der zugrundeliegenden Faktorenanzahl führt als der Scree-Test (Cattell, 1966) oder das Kaiser-Guttman-Kriterium (Guttman, 1954; Kaiser & Dickman, 1959).

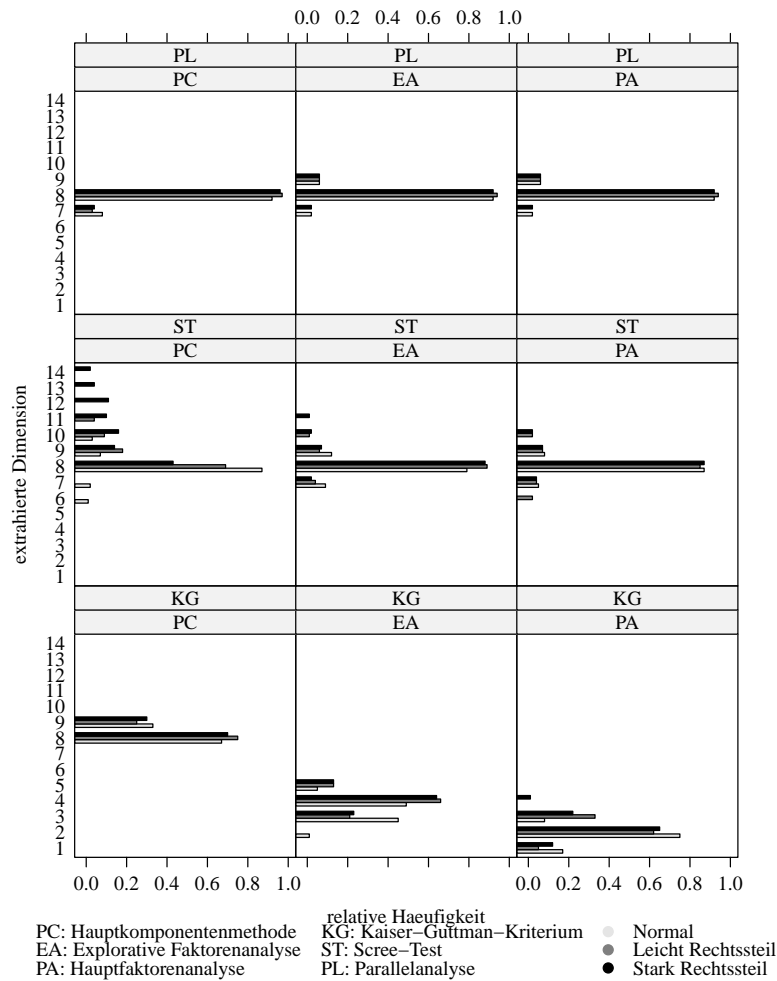


Abbildung 4.4: Anzahl der extrahierten Dimensionen ( $N = 600$ ;  $k = 8$ )

## 4.2 Geschätzte Schiefe der Fähigkeitswerte

Für die folgenden Analysen wurde angenommen, dass die wahre Faktorenanzahl richtig geschätzt wurde. Sodann wurde für jede der 12 simulierten Bedingungen pro simulierten Datensatz die Schiefe der geschätzten Faktorwerte bestimmt. Die Verteilung der geschätzten Schiefewerte in Abhängigkeit des Faktorenmodells und des Faktors für die erste Bedingung (Stichprobengröße  $N = 200$ ; Faktorenanzahl  $k = 4$ ; Verteilung der latenten Fähigkeitswerte „Normal“ d.h. wahre Schiefe der Faktorwerte  $\mu_{3m} = 0$ ) ist in Abbildung 4.5 auf der folgenden Seite dargestellt.

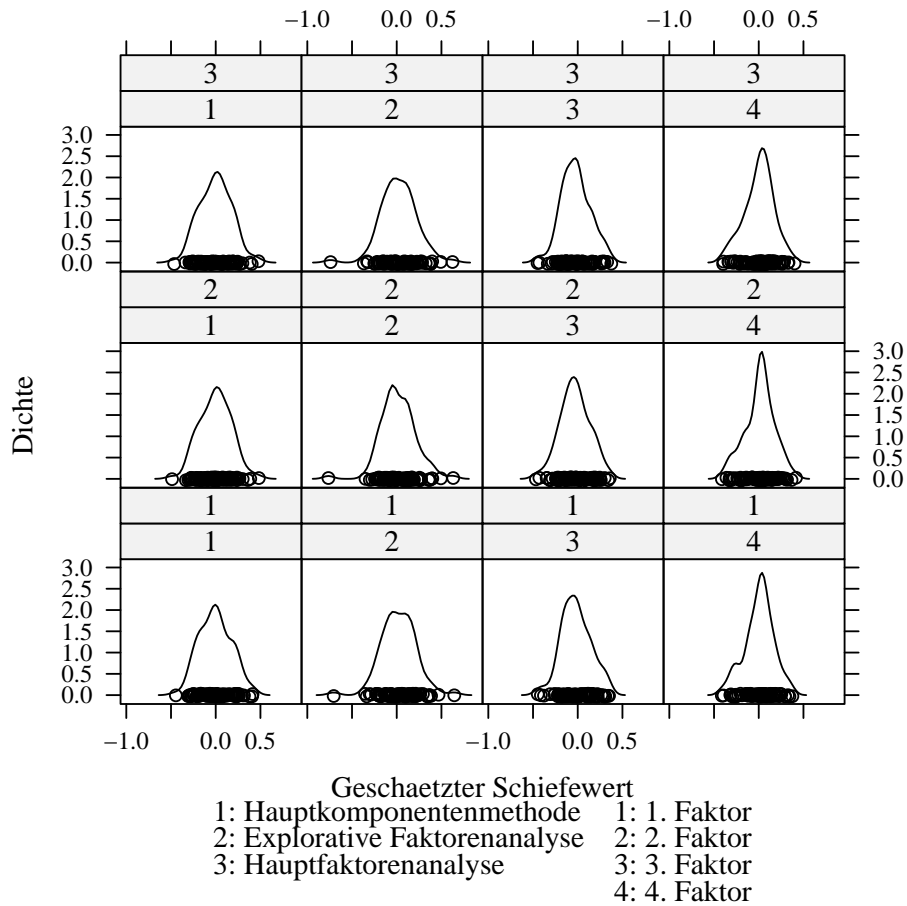


Abbildung 4.5: Verteilung der geschätzten Schiefewerte in Abhängigkeit des Faktorenmodells und des Faktors ( $N = 200$ ;  $k = 4$ ;  $\mu_{3m} = 0$ )

Wie zu erkennen, weisen die geschätzten Fähigkeitswerte mehrheitlich eine Schiefe von  $\sqrt{b} = 0$  auf, d.h. wenn die wahren Fähigkeitswerte nicht schief verteilt sind, dann sind auch die geschätzten Fähigkeitswerte in der Regel nicht schief verteilt. Ein Einfluss des zur Analyse verwendeten Faktorenmodells auf die Schiefe der geschätzten Fähigkeitswerte kann unter dieser Bedingung nicht festgestellt werden, allerdings reduziert sich der Standardfehler der geschätzten Schiefewerte vom ersten zum vierten Faktor deutlich, d.h. die wahre Schiefe der Fähigkeitswerte kann für den vierten Faktor effizienter geschätzt werden als für den ersten Faktor.

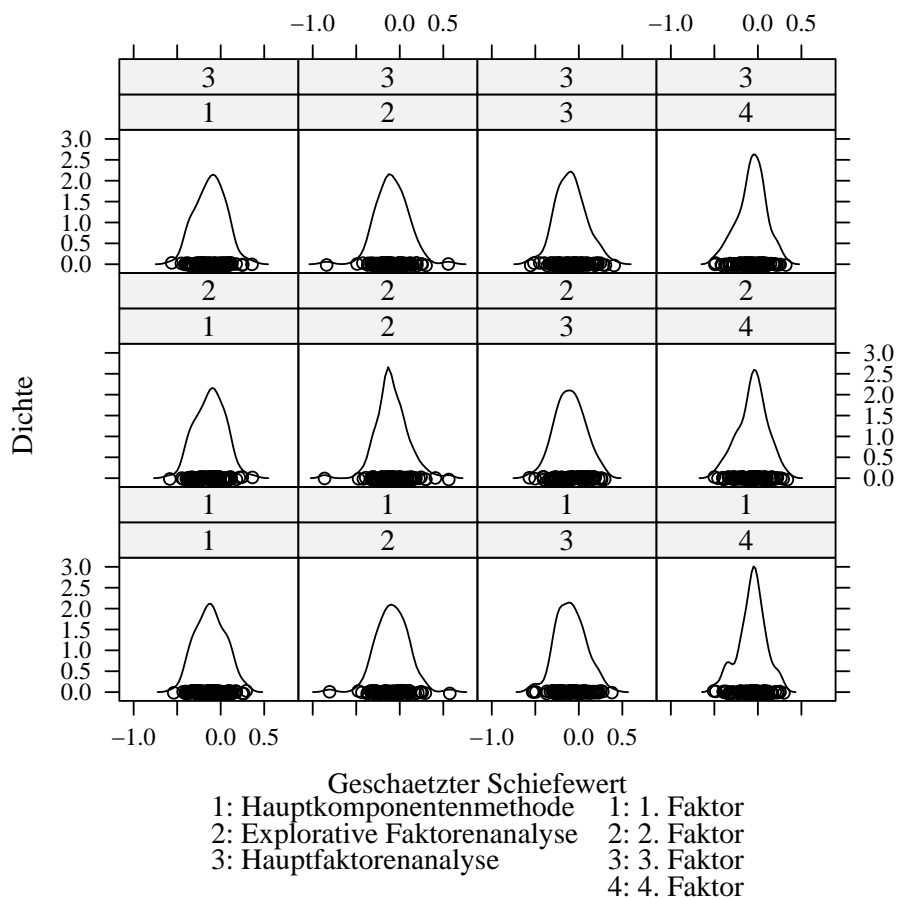


Abbildung 4.6: Verteilung der geschätzten Schiefewerte in Abhängigkeit des Faktorenmodells und des Faktors ( $N = 200$ ;  $k = 4$ ;  $\mu_{3m} = -0.20$ )

Sind die wahren Faktorwerte leicht rechtssteil verteilt, d.h. gilt  $\mu_{3m} = -0.20$ , dann kann diese wahre Schiefe offensichtlich nur für die latenten Werte des ersten und zweiten extrahierten Faktors (nahezu) erwartungstreu geschätzt werden; die latenten Werte des dritten und vierten extrahierten Faktors weisen mehrheitlich eine geschätzte Schiefe von  $\hat{\mu}_3 = 0$  auf, d.h. die wahre Schiefe der latenten Werte des dritten und vierten extrahierten Faktors wird in den meisten Fällen überschätzt (Abbildung 4.6).

Wenn die wahren Faktorwerte eine starke rechtssteile Verteilung aufweisen, d.h. wenn  $\mu_{3m} = -2.00$  gilt, dann kann die wahre Schiefe auch nicht mehr bei den ersten beiden extrahierten Faktoren (näherungsweise) erwartungstreu geschätzt werden. Vielmehr wird die wahre Schiefe unter dieser Bedingung





klassischen Faktorenmodellen nicht erwartungstreu geschätzt werden. Eine Erhöhung der Stichprobengröße von  $N = 200$  auf  $N = 600$  ändert an diesem Befund nichts, weswegen auf eine detaillierte Beschreibung der dazugehörigen Befunde verzichtet wird. Statt dessen finden sich die entsprechenden Abbildungen in Anhang B (Seite 107).

In Abbildung 4.8 ist die Verteilung der geschätzten Schiefewerte in Abhängigkeit des Faktorenmodells und des Faktors für die Bedingung  $N = 200$ ,  $k = 8$  und  $\mu_{3m} = 0$  dargestellt. Wenn die wahren latenten Werte normalverteilt sind, dann kann die Schiefe der Fähigkeitswerte offensichtlich (in etwa) erwartungstreu geschätzt werden, und zwar unabhängig davon, welches Faktorenmodell zur Analyse der Daten gewählt wurde oder welcher Faktor betrachtet wird.

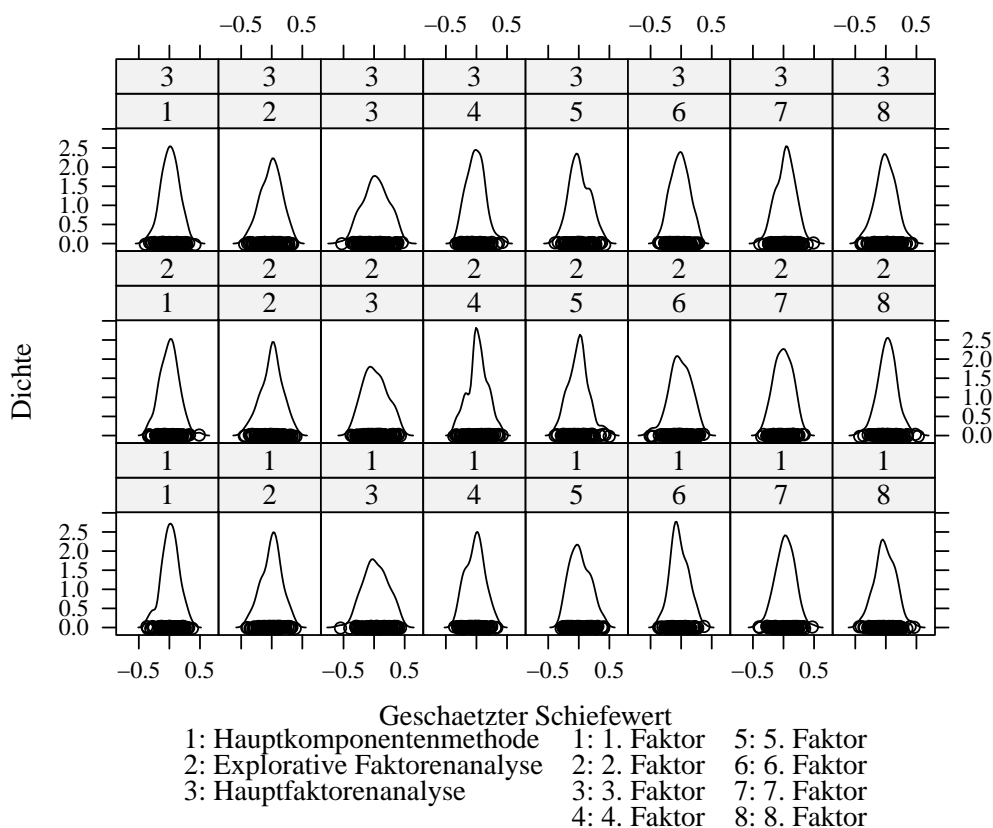


Abbildung 4.8: Verteilung der geschätzten Schiefewerte in Abhängigkeit des Faktorenmodells und des Faktors ( $N = 200$ ;  $k = 8$ ;  $\mu_{3m} = 0$ )

Die Eigenschaft der erwartungstreuen Schätzung der Schiefe scheint allerdings verloren zu gehen, wenn die wahren Fähigkeitswerte leicht rechtssteil verteilt sind (Abbildung 4.9). Zumindest für die zuletzt extrahierten Faktoren wird die wahre Schiefe tendenziell überschätzt, sodass anstelle leicht rechtssteil verteilter Fähigkeitswerte eher latente Werte ohne Schiefe geschätzt werden.

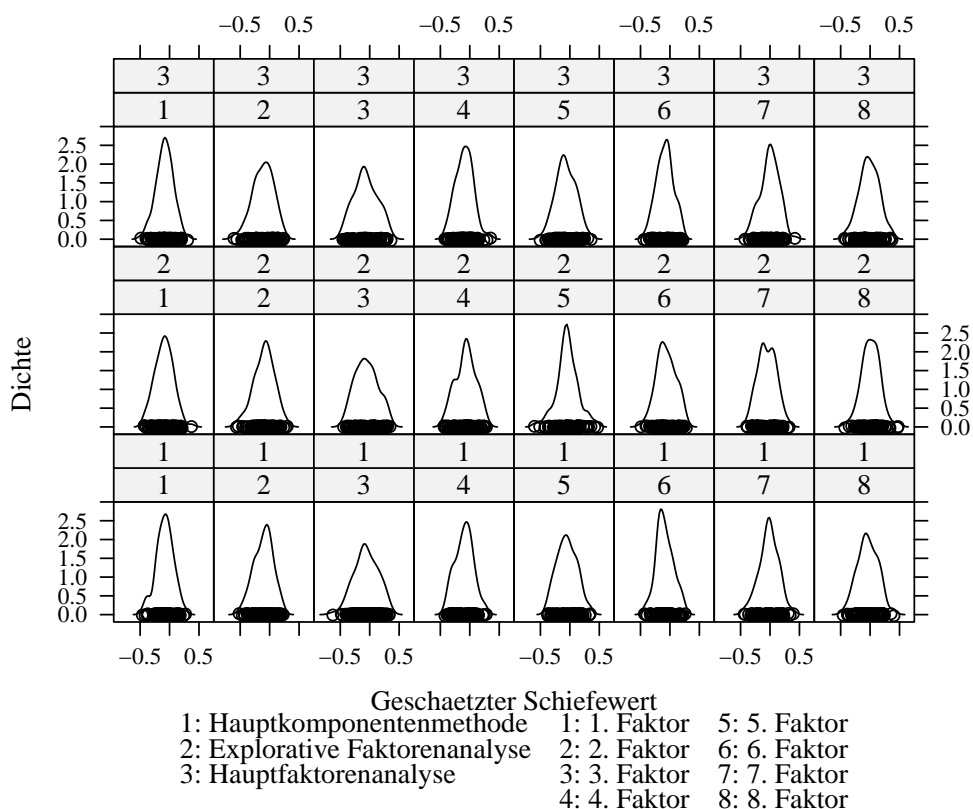


Abbildung 4.9: Verteilung der geschätzten Schiefewerte in Abhängigkeit des Faktorenmodells und des Faktors ( $N = 200$ ;  $k = 8$ ;  $\mu_{3m} = -0.20$ )

Sind die wahren latenten Werte stark rechtssteil verteilt, dann ist die Schätzung der Schiefe auch für die Fähigkeitswerte bei den ersten extrahierten Faktoren nicht mehr erwartungstreu (Abbildung 4.10, Seite 95). Bei stark rechtssteil verteilten Fähigkeitswerten wird die wahre Schiefe der latenten Werte also generell überschätzt, und zwar unabhängig davon, welches Faktorenmodell zur Analyse der Daten verwendet wurde oder welcher Faktor

betrachtet wird. Wird die Stichprobengröße von  $N = 200$  auf  $N = 600$  erhöht, ändert sich an dieser generellen Befundlage wenig, weswegen auf eine detaillierte Beschreibung der entsprechenden Ergebnisse verzichtet wird. Die dazugehörigen Abbildungen finden sich jedoch in Anhang C (Seite 110).

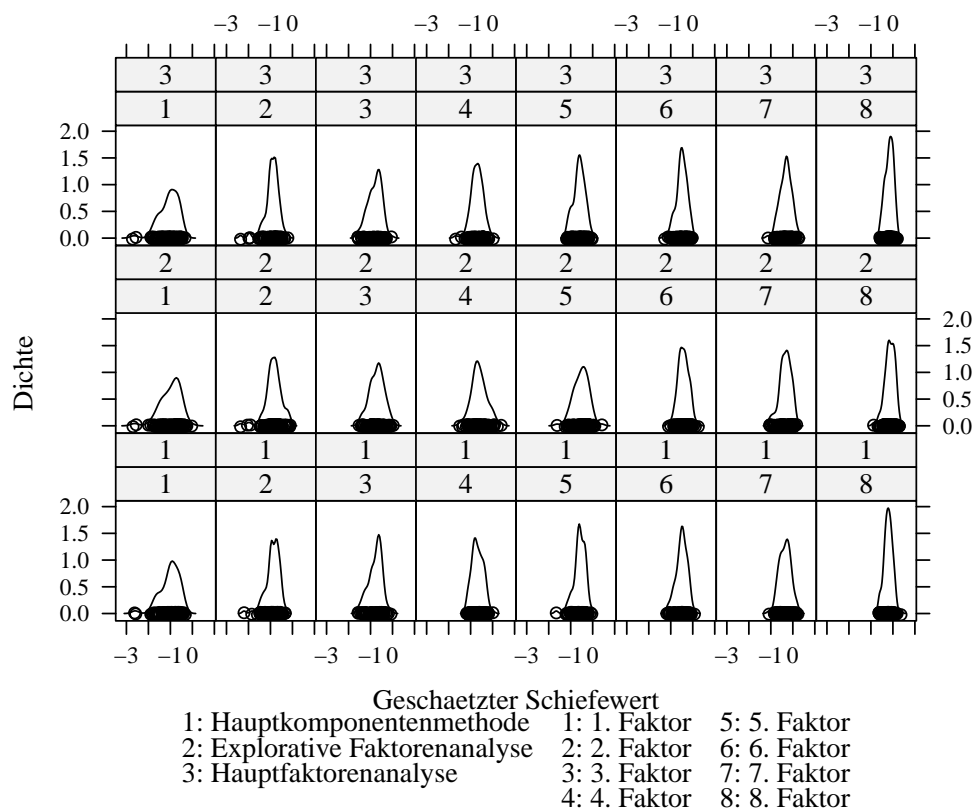


Abbildung 4.10: Verteilung der geschätzten Schiefewerte in Abhängigkeit des Faktorenmodells und des Faktors ( $N = 200$ ;  $k = 8$ ;  $\mu_{3m} = -2.00$ )

Insgesamt fallen die Ergebnisse für den Fall von acht den Daten zugrundeliegenden Faktoren damit in etwa genauso aus wie die Ergebnisse, wenn den Daten vier Faktoren unterliegen. In beiden Fällen reagiert das Kriterium „geschätzte Schiefe der Fähigkeitsverteilung“ eher sensibel auf das Ausmaß an Rechtssteile der wahren Fähigkeitswerte: Mit zunehmender Rechtssteile der wahren Fähigkeitswerte wird die wahre Schiefe der latenten Werte immer deutlicher überschätzt. Dies gilt unabhängig vom dem zur Analyse der Daten verwendeten klassischen Faktorenmodell und dem betrachteten Faktor.

### 4.3 Fit der geschätzten Ladungsmatrizen

In Tabelle 4.1 ist der durchschnittliche Abweichungskoeffizient  $\overline{AQ}_L$  und die Streuung des Abweichungskoeffizienten  $\sigma_{AQ_L}$  in Abhängigkeit der simulierten Bedingung bzw. des zur Auswertung verwendeten Faktorenmodells abgetragen. Die größten durchschnittlichen Abweichungskoeffizienten resultieren unter der Bedingung einer kleinen Stichprobe, einer wahren Faktorenanzahl von  $k = 8$  und rechtssteil verteilten latenten Fähigkeitswerten. Unter dieser Bedingung werden die wahren Ladungskoeffizienten im Durchschnitt also am deutlichsten über- bzw. unterschätzt (durchschnittlicher Abweichungskoeffizient 0.1726, 0.1565 und 0.1429). Nur geringfügige durchschnittliche Abweichungen zwischen geschätzten und wahren Ladungskoeffizienten resultieren, wenn die Stichprobe groß, die wahre Faktorenanzahl  $k = 4$  und die latenten Fähigkeitswerte normal bzw. nur leicht rechtssteil verteilt sind (durchschnittlicher Abweichungskoeffizient 0.0761, 0.0627 und 0.0657 bzw. 0.0750, 0.0622 und 0.0640).

Tabelle 4.1: Durchschnittlicher Abweichungskoeffizient  $\overline{AQ}_L$  und Streuung  $\sigma_{AQ_L}$  in Abhängigkeit der simulierten Bedingung

$N$	$k$	Modell	Verteilung der latenten Fähigkeitswerte					
			Normal		Leicht rechtssteil		Stark rechtssteil	
			$\overline{AQ}_L$	$\sigma_{AQ_L}$	$\overline{AQ}_L$	$\sigma_{AQ_L}$	$\overline{AQ}_L$	$\sigma_{AQ_L}$
200	4	PC <sup>a</sup>	0.1432	0.156	0.1430	0.157	0.1580	0.173
		EA <sup>b</sup>	0.1294	0.142	0.1238	0.137	0.1413	0.154
		PA <sup>c</sup>	0.1278	0.140	0.1236	0.136	0.1369	0.150
600	4	PC	0.0761	0.087	0.0750	0.086	0.0914	0.106
		EA	0.0627	0.072	0.0622	0.073	0.0803	0.095
		PA	0.0657	0.075	0.0640	0.074	0.0820	0.097
200	8	PC	0.1652	0.169	0.1625	0.166	0.1726	0.176
		EA	0.1542	0.157	0.1517	0.155	0.1565	0.159
		PA	0.1350	0.138	0.1343	0.138	0.1429	0.146
600	8	PC	0.1187	0.123	0.1179	0.123	0.1246	0.130
		EA	0.1066	0.112	0.1066	0.112	0.1155	0.120
		PA	0.0969	0.101	0.0949	0.099	0.1016	0.106

<sup>a</sup>PC: Hauptkomponentenmethode; <sup>b</sup>EA: Explorative Faktorenanalyse

<sup>c</sup>PA: Hauptfaktorenanalyse

Allgemein variiert der durchschnittliche Abweichungskoeffizient in Abhängigkeit der wahren Faktorenanzahl und der Stichprobengröße: Die Abweichung zwischen geschätzter und wahrer Ladungsmatrix ist größer, wenn den Daten acht anstelle von vier Faktoren zugrunde liegen, und die Abweichung reduziert sich, wenn große anstelle von kleinen Stichproben analysiert werden. Ebenfalls ist ein schwacher Effekt der Verteilung der latenten Fähigkeitswerte und des zur Auswertung verwendeten Faktorenmodells auf den durchschnittlichen Abweichungskoeffizienten zu erkennen: Bei stark rechtssteil verteilten latenten Fähigkeitswerten nimmt der Fit der geschätzten Ladungsmatrix ab, und die geschätzte Ladungsmatrix weicht von der wahren Ladungsmatrix weniger ab, wenn zur Analyse der Daten das explorative Faktorenmodell oder die Hauptfaktorenanalyse und nicht die Hauptkomponentenmethode gewählt wird.

Um den schwachen Effekt der Verteilung der latenten Fähigkeitswerte und des zur Auswertung verwendeten Faktorenmodells auf den durchschnittlichen Abweichungskoeffizienten differenzierter zu betrachten, wird anstelle der durchschnittlichen Abweichung der geschätzten Ladungsmatrizen die Verteilung der absoluten Differenzen  $|\hat{l}_{ipk} - l_{pk}|$  betrachtet. Pro simulierter Bedingung und angewendetem Faktorenmodell zur Auswertung der Datensätze können  $100pk$  dieser absoluten Differenzen bestimmt und in einem Graphen dargestellt werden. Dadurch sind nicht nur Unterschiede in der durchschnittlichen Abweichung der geschätzten Ladungsmatrizen zwischen den Bedingungen erkennbar, sondern beispielsweise auch die relativen Häufigkeiten, mit der bestimmte absolute Differenzen beobachtet wurden.

In Abbildung 4.11 (Seite 98) ist die Verteilung der absoluten Differenzen  $|\hat{l}_{ipk} - l_{pk}|$  für die Bedingung  $N = 200$  und  $k = 4$  dargestellt.<sup>1</sup> Wie zu erkennen, liegt die Mehrzahl der absoluten Differenzen in dem Bereich von 0 bis etwa 0.2. Größere Abweichungen der geschätzten von den wahren Ladungskoeffizienten sind dagegen eher unwahrscheinlich. Es ist auch ersichtlich, dass sich die neun Verteilungen kaum unterscheiden. Dies spricht eher dafür, den schwachen Effekt der latenten Schiefe und des zur Auswertung verwendeten Faktorenmodells auf den durchschnittlichen Abweichungskoeffizienten nicht zu überinterpretieren, d.h. im Allgemeinen ist davon auszugehen, dass es keinen Zusammenhang zwischen der Verteilung der latenten Fähigkeitswerte bzw. des zur Auswertung verwendeten Faktorenmodells und den Fit der geschätzten Ladungsmatrix gibt.

---

<sup>1</sup>Weil die Verteilungen der absoluten Differenz für die anderen Bedingungen ähnlich ausfallen, wird auf eine detaillierte Beschreibung der dazugehörigen Graphiken verzichtet. Statt dessen sind die entsprechenden Abbildungen in Anhang D (Seite 113) aufgeführt.

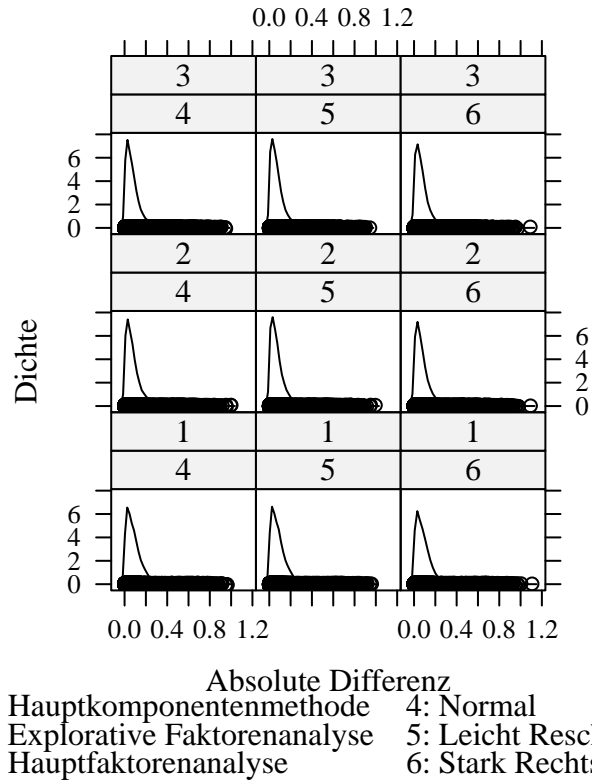


Abbildung 4.11: Verteilung der absoluten Differenzen  $|\hat{l}_{ipk} - l_{pk}|$  in Abhängigkeit des Faktorenmodells und der latenten Schiefe ( $N = 200$ ;  $k = 4$ )

# Kapitel 5

## Zusammenfassung, Schlussfolgerungen und Empfehlungen

Ziel der Dissertation war es zu überprüfen, welches Faktorenmodell die faktorielle Struktur eines Tests am genauesten schätzt, wenn sowohl die Fähigkeitswerte als auch die Testwerte rechtssteil verteilt sind. Dieses Wissen ist von Nutzen, weil zwar für sämtliche Tests (nicht nur) der erziehungswissenschaftlichen Forschung die faktorielle Struktur bestimmt werden muss, um zu überprüfen ob das Gütekriterium der faktoriellen Validität erreicht ist, aber eine Klasse von Tests der empirischen Bildungsforschung häufig rechtssteil verteilte Testdaten aufweist und (insbesondere die kriteriumsorientierten Tests) rechtssteil verteilte Fähigkeiten misst, und damit die zentrale Voraussetzung der zur Überprüfung der faktoriellen Struktur häufig eingesetzten klassischen Faktorenmodelle nicht erfüllt. Wenn also bekannt ist, welches Faktorenmodell die faktorielle Struktur eines Tests am genauesten schätzt, wenn sowohl die Fähigkeitswerte als auch die Testdaten rechtssteil verteilt sind, dann ist auch bekannt, welches Faktorenmodell zur Überprüfung der faktoriellen Validität dieser Klasse von Tests der empirischen Bildungsforschung eingesetzt werden sollte.

Um den Zusammenhang zwischen der Schätzgüte verschiedener klassischer Faktorenmodelle und dem Ausmaß an rechtssteil verteilten Fähigkeitswerten und Testdaten zu bestimmen, wurde 1. der theoretische und empirische Forschungsstand aufgearbeitet, 2. mathematisch-statistische Analysen durchgeführt, und 3. eine Simulationsstudie implementiert. Leitend für die mathematisch-statistischen Analysen und die Simulationsstudie waren vier Forschungsfragen, welche aus der Aufarbeitung des theoretischen und em-



pirischen Forschungsstand abgeleitet wurden (Kapitel 3.1, Seite 68). Korrespondierend zu den vier Forschungsfragen wurden vier Hypothesen aufgestellt (Kapitel 3.2, Seite 70), und die mathematisch-statistischen Analysen sowie die Simulationsstudie dienten zur Überprüfung dieser Hypothesen.

Hypothese I konnte zumindest partiell bestätigt werden: Wie die mathematisch-statistischen Analysen im Rahmen des Mattson-Ansatz zeigten, kann davon ausgegangen werden, dass schief verteilte Daten schief verteilte Fähigkeitswerte auf zumindest einer latenten Dimension implizieren, dass schief verteilte Daten also bestimmte Verteilungen (nämlich schiefe Verteilungen) der Fähigkeitswerte implizieren. Nicht bestätigt werden konnte hingegen die Hypothese, dass schief verteilte Daten auch bestimmte Faktorenstrukturen (etwa im Sinne der Anzahl extrahierbarer Faktoren) implizieren.

Im Unterschied zu Hypothese I, die nur partiell bestätigt werden konnte, konnte Hypothese II vollständig bestätigt werden: Alle betrachteten Faktorenmodelle reagierten sensibel auf rechtsteil verteilte Fähigkeitswerte. Mit zunehmender Rechtssteile der wahren Fähigkeitswerte wich die geschätzte Schiefe der Fähigkeitswerte von der Schiefe der wahren Fähigkeitswerte immer deutlicher ab. Die Schätzgenauigkeit der betrachteten Faktorenmodelle nimmt also mit zunehmender Rechtssteile der Fähigkeitswerte ab.

Die Hypothesen III, IIIa und IIIb konnten ebenfalls vollständig bestätigt werden: Erwartungsgemäß bestand kein bedeutsamer Zusammenhang zwischen dem Ausmaß der Rechtssteile der Fähigkeitsverteilung und der Anzahl extrahierter Faktoren bzw. dem Fit der geschätzten Ladungsmatrizen, wohl aber wurde die Schiefe der latenten Werte bei stark rechtsteil verteilten Fähigkeitswerten stärker überschätzt als bei leicht rechtsteil oder normal verteilten Fähigkeitswerten. Das Kriterium „geschätzte Schiefe der Fähigkeitsverteilung“ reagierte also sensibler auf das Ausmaß an Rechtssteile der Fähigkeitswerte als die Kriterien „Anzahl der extrahierten Faktoren“ und „Fit der geschätzten Ladungsmatrize“.

Schließlich konnte Hypothese IV insofern bestätigt werden, als dass sich eine generelle Überlegenheit der Parallelanalyse gegenüber dem Scree-Test oder dem Kaiser-Guttman-Kriterium gezeigt hat, und zwar unabhängig davon, welches Faktorenmodell zur Analyse der Daten angewendet wurde.

Insgesamt sprechen die Ergebnisse der Dissertation dafür, dass das Ausmaß an Rechtssteile der wahren Fähigkeitswerte zwar einen Einfluss auf die Schätzung der Fähigkeitswerte hat, nicht aber auf den Fit der Ladungsmatrize oder auf die Anzahl der extrahierten Faktoren. Mit Bezug auf die Zielstellung der Dissertation bleibt damit festzuhalten, dass die betrachteten klassischen Faktorenmodelle gleich robust gegenüber dem Ausmaß an Rechtssteile

der wahren Fähigkeitswerte sind, oder anders formuliert, die betrachteten klassischen Faktorenmodelle schätzen die faktorielle Struktur bei rechtssteil verteilten Fähigkeitswerte und Testwerte gleich genau.

Das die Schiefe der Fähigkeitsverteilung bei rechtssteil verteilten Fähigkeitswerten überschätzt wird, könnte vor allem dann problematisch sein, wenn die geschätzten Fähigkeitswerte weiter verwendet werden sollen. Dies ist dann der Fall, wenn beispielsweise die geschätzten Fähigkeitswerte das Kriterium in einer Regressionsgleichung sind oder wenn Kompetenzklassen auf der Basis der geschätzten Fähigkeitswerte gebildet werden sollen. Vor einer derartigen Weiterverwendung der geschätzten Fähigkeitswerte wäre also zu klären, ob und gegebenenfalls in welchem Umfang mit einer Verschätzung der Fähigkeitswerte zu rechnen ist und wie sich diese Verschätzung auf die weiteren Analyseschritte auswirken kann bzw. wie diese Verschätzung möglicherweise durch eine Modifikation der weiteren Prozeduren kompensierbar ist.

Wenngleich die Befunde der Dissertation und die darauf basierenden Schlussfolgerungen eindeutig sind, müssen bestimmte Einschränkungen erwähnt werden. So gelten die Ergebnisse dieser Dissertation nur für die betrachteten klassischen Faktorenmodelle. Inwiefern andere klassische Verfahren, wie etwa die Alpha-Faktorisierung (Harman, 1967), ebenfalls robust gegenüber rechtssteil verteilten Fähigkeitswerten sind, müsste erst noch geklärt werden. Auch ist zu bedenken, dass in der Dissertation nur eine Voraussetzung der klassischen Faktorenmodelle (normalverteilte Werte) untersucht wurde. Wie sich die gleichzeitige Verletzung mehrerer Voraussetzungen der klassischen Faktorenanalyse auf die Schätzgüte dieser Modelle auswirkt, bleibt damit unklar. Schließlich ist zu bedenken, dass zwar starke rechtssteil verteilte Fähigkeitswerte simuliert wurden, in der empirischen Praxis aber durchaus noch viel größere Schiefewerte beobachtet werden können. Inwiefern die betrachteten klassischen Faktorenmodelle auch dann noch robust sind, wenn solche Schiefewerte auftreten, muss durch weitere Forschungsarbeiten herausgearbeitet werden.<sup>1</sup>

Gänzlich unberücksichtigt in dieser Dissertation blieben verschiedene Faktorenmodelle der Item-Response-Theorie. Dies konnte methodisch gerechtfertigt werden: Die Modelle der Item-Response-Theorie können vorhandene Schiefe in den manifesten Daten durch entsprechende Fixierung der Item-

---

<sup>1</sup>Bezüglich der Simulierung von extremen Schiefewerten bzw. der dazugehörigen latenten und manifesten Verteilungen ist allerdings zu berücksichtigen, dass derartige extreme Schiefewerte mit dem bisher bestehenden Instrumentarium zum Simulieren von Daten für Modelle mit latenten Variablen nicht systematisch manipuliert werden können, d.h. hier müssen erst die Verfahren zum Simulieren von Daten für Modelle mit latenten Variablen weiterentwickelt werden.

parameter auffangen, und damit die Fähigkeitsverteilung (zumindest in bestimmten Umfang) unabhängig von der Verteilung der manifesten Daten modellieren. Die Art der Fähigkeitsverteilung ist damit bei Modellen der Item-Response-Theorien aus mathematisch-statistischer Sicht in Prinzip beliebig und eine Auswahl für eine bestimmte Art der Fähigkeitsverteilung wäre bestenfalls aufgrund der zugrundeliegenden inhaltlichen Theorie zu begründen. Bisher wird dabei in der Regel die Normalverteilung für die latenten Werte gewählt; ob dass in jedem Fall durch die inhaltliche Theorie zu begründen ist, müsste geklärt werden.

Auf der Grundlage der vorgestellten Befunde und den daraus abgeleiteten Schlussfolgerungen lassen sich Empfehlungen für die wissenschaftliche Forschung und Praxis ableiten. Diese Empfehlungen werden hier enumerativ aufgeführt.

1. Bei der faktorenanalytischen Untersuchung von Testinstrumenten zur Bestimmung der Dimensionalität sollte als Extraktionskriterium der Parallelanalyse nach Horn (1965) der Vorzug gegenüber dem Scree-Test von Cattell (1966) oder dem Kaiser-Guttman-Kriterium nach Guttman (1954) und Kaiser und Dickman (1959) gegeben werden. Die Parallelanalyse erwies sich als zuverlässiger, wenn die Testdaten nicht intervallskaliert waren und dieser Vorteil blieb erhalten, wenn die wahren Fähigkeitswerte stark rechtssteil verteilt waren.
2. Wenn die Parallelanalyse als Extraktionskriterium gewählt wird, und wenn die Voraussetzungen der in dieser Dissertation untersuchten klassischen Faktorenmodelle erfüllt sind, mit Ausnahme der Annahme normalverteilter Testdaten bzw. normalverteilter wahrer Fähigkeitswerte, dann sollten zur Beurteilung der faktoriellen Validität von kriteriumorientierten Tests die hier untersuchten klassischen Faktorenmodelle herangezogen werden. Für diese Modelle konnte in der Dissertation gezeigt werden, dass sie unter den genannten Bedingungen die wahre faktorielle Struktur der Testdaten relativ robust schätzen.
3. Sind die wahren Fähigkeitswerte rechtssteil verteilt, dann sollten die mit einer der in dieser Dissertation untersuchten Faktorenmodelle geschätzten Fähigkeitswerte nicht unreflektiert weiterverwendet werden. Es hat sich gezeigt, dass die Schiefe der Fähigkeitsverteilung bei rechtssteil verteilten wahren Fähigkeitswerten überschätzt wird; eine unreflektierte Verwendung dieser verzerrten Fähigkeitswerte könnte zu falschen Schlussfolgerungen führen.

4. Es sollte geklärt werden, wie sich die gleichzeitige Verletzung von mehreren Voraussetzungen klassischer Faktorenmodelle auf die Schätzgüte dieser Modelle auswirkt. Eine Kenntnis dieser Effekte ist wichtig, weil die Messwerte vieler Testinstrumente der Erziehungswissenschaft nicht selten mehrere Voraussetzungen klassischer Faktorenmodelle nicht erfüllen.
5. Es sollte überprüft werden, wie sich die Schätzgüte der klassischen Faktorenmodelle ändert, wenn die Testdaten bzw. die wahren Fähigkeitswerte noch stärker rechtssteil verteilt sind, als in dieser Dissertation simuliert. Dafür wären zunächst die Methoden zur systematischen Simulation von Daten für Modelle mit latenten Variablen weiterzuentwickeln. Dies erscheint sinnvoll, weil viele Testdaten in der Praxis eine deutlichere Schiefe aufweisen als in dieser Simulation untersucht werden konnte, und weil sich gezeigt hat, dass die Schätzgüte der Faktorenmodelle mit zunehmender Schiefe abnimmt.
6. Generell wäre zu überdenken, ob nicht auch für andere als die in dieser Dissertation betrachteten Faktorenmodelle überprüft werden sollte, wie sich rechtssteil verteilte wahre Fähigkeitswerte auf die Schätzgenauigkeit auswirken. Hier kämen beispielsweise weitere klassische Modelle in Betracht (wie etwa die Alpha-Faktorisierung), oder auch Modelle der Item-Response-Theorie. Wenn Modelle der Item-Response-Theorie betrachtet werden sollen, dann wäre vorher der Zusammenhang zwischen Itemparameterschätzung und Schätzung der latenten Fähigkeitsverteilung aufzuarbeiten.

# Anhang A

## Basissyntax zur Datensimulation

Als Beispiel einer Syntax zur Datengenerierung ist hier die Syntax für die Bedingung normalverteilter latenter  $f_m$  bei vier Faktoren dargestellt. Die Syntax diene als Input in das Programm SAS 9.0, und als Output der Syntax resultiert die simulierte Datenmatrix  $Z$ .

```
proc iml;
start normal;
p = 24;
n = 200;
k = 4;
Seed = 100;
L = {.9 0 0 0,
     .8 0 0 0,
     .7 0 0 0,
     .6 0 0 0,
     .5 0 0 0,
     .4 0 0 0,
     0 .8 0 0,
     0 .7 0 0,
     0 .6 0 0,
     0 .5 0 0,
     0 .4 0 0,
     0 .3 0 0,
     0 0 .6 0,
     0 0 .6 0,
     0 0 .5 0,
     0 0 .4 0,
     0 0 .4 0,
```

```

    0 0 .3 0,
    0 0 0 .6,
    0 0 0 .5,
    0 0 0 .5,
    0 0 0 .4,
    0 0 0 .3,
    0 0 0 .3};
einste = {[24] 1};
einse = einste';
lambda1te = {[24] 0};
lambda1e = lambda1te';
lambda2e = .1974;
lambda3e = .1349;
lambda4e = .1349;
einst = {[4] 1};
eins = einst';
lambda1t = {[4] 0};
lambda1 = lambda1t';
lambda2 = .1974;
lambda3 = .1349;
lambda4 = .1349;

ri = ranuni(Seed) // ranuni(Seed) // ranuni(Seed) // ranuni(Seed);
ri1 = ri##lambda3;
ri2 = eins-ri;
ri2p = ri2##lambda4;
ri3 = ri1-ri2p;
ri4 = ri3/lambda2;
fw = lambda1+ri4;

rie = ranuni(Seed) // ranuni(Seed) // ranuni(Seed) // ranuni(Seed) //
ranuni(Seed) // ranuni(Seed) // ranuni(Seed) // ranuni(Seed) //
ranuni(Seed) // ranuni(Seed) // ranuni(Seed) // ranuni(Seed) //
ranuni(Seed) // ranuni(Seed) // ranuni(Seed) // ranuni(Seed) //
ranuni(Seed) // ranuni(Seed) // ranuni(Seed) // ranuni(Seed) //
ranuni(Seed) // ranuni(Seed) // ranuni(Seed) // ranuni(Seed);
ri1e = rie##lambda3e;
ri2e = einse-rie;
ri2pe = ri2e##lambda4e;
ri3e = ri1e-ri2pe;
ri4e = ri3e/lambda2e;

```

$ew = \lambda_1 e + r_1 4e;$

$FT = fw;$

$F = FT';$

$ET = ew;$

$E = ET';$

$LTT = L';$

$Z = F*LTT + E;$

# Anhang B

## Geschätzte Schiefewerte für $k = 4$

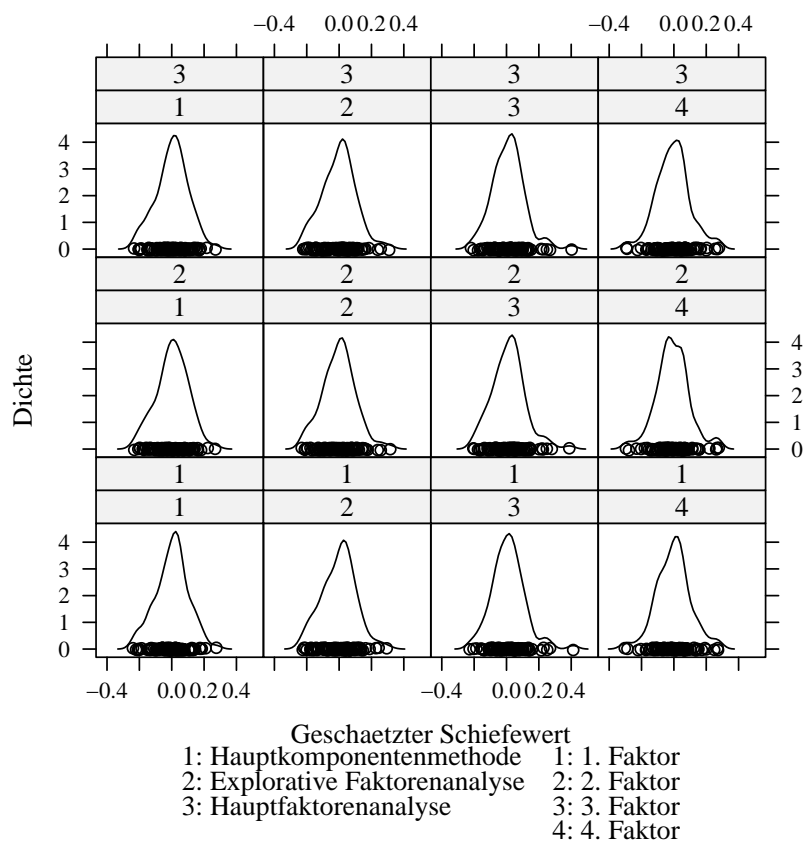


Abbildung B.1: Verteilung der geschätzten Schiefewerte in Abhängigkeit des Faktorenmodells und des Faktors ( $N = 600$ ;  $k = 4$ ;  $\mu_{3m} = 0$ )



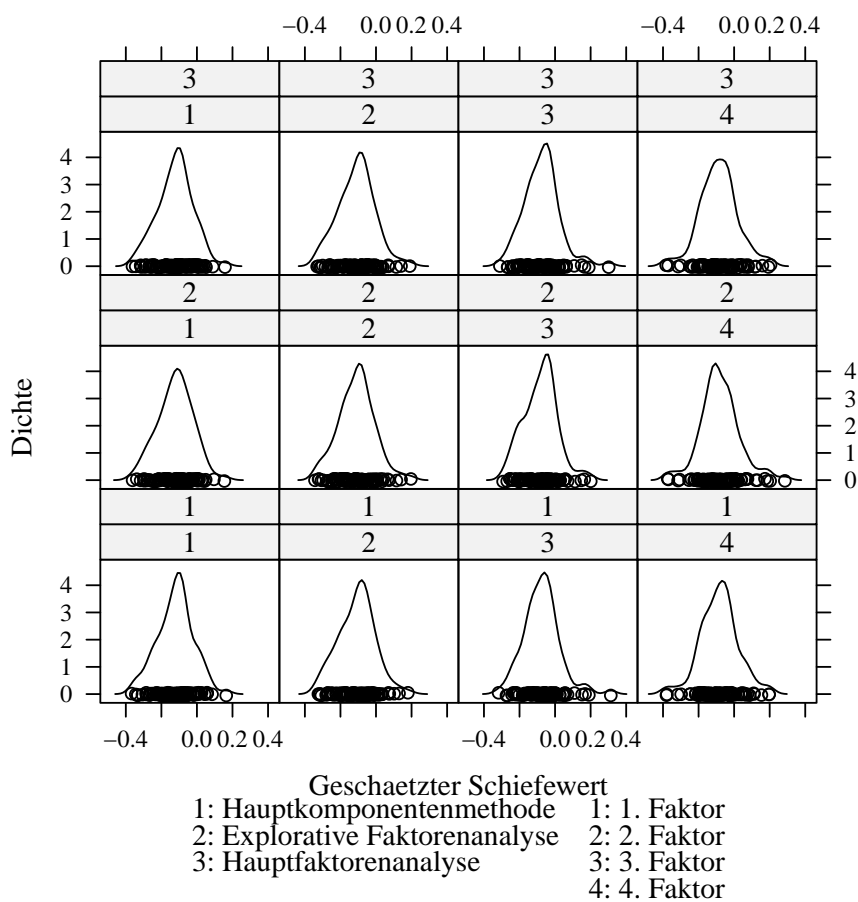


Abbildung B.2: Verteilung der geschätzten Schiefewerte in Abhängigkeit des Faktorenmodells und des Faktors ( $N = 600$ ;  $k = 4$ ;  $\mu_{3m} = -0.20$ )

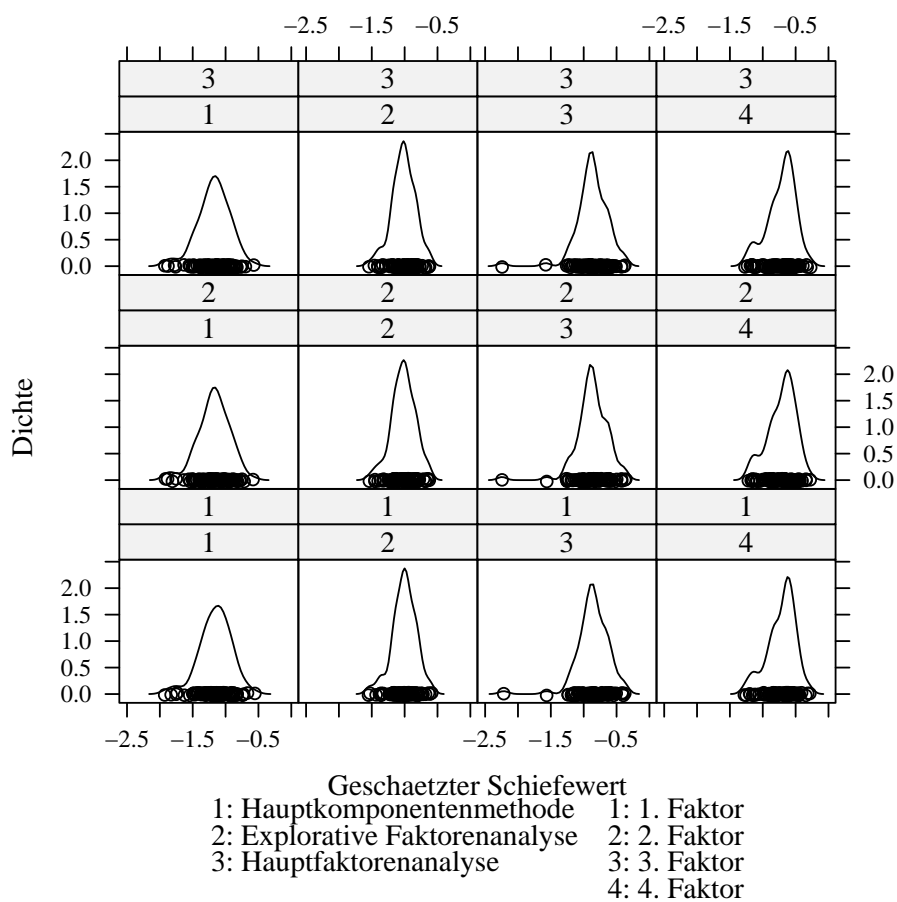


Abbildung B.3: Verteilung der geschätzten Schiefewerte in Abhängigkeit des Faktorenmodells und des Faktors ( $N = 600$ ;  $k = 4$ ;  $\mu_{3m} = -2.00$ )

# Anhang C

## Geschätzte Schiefewerte für $k = 8$

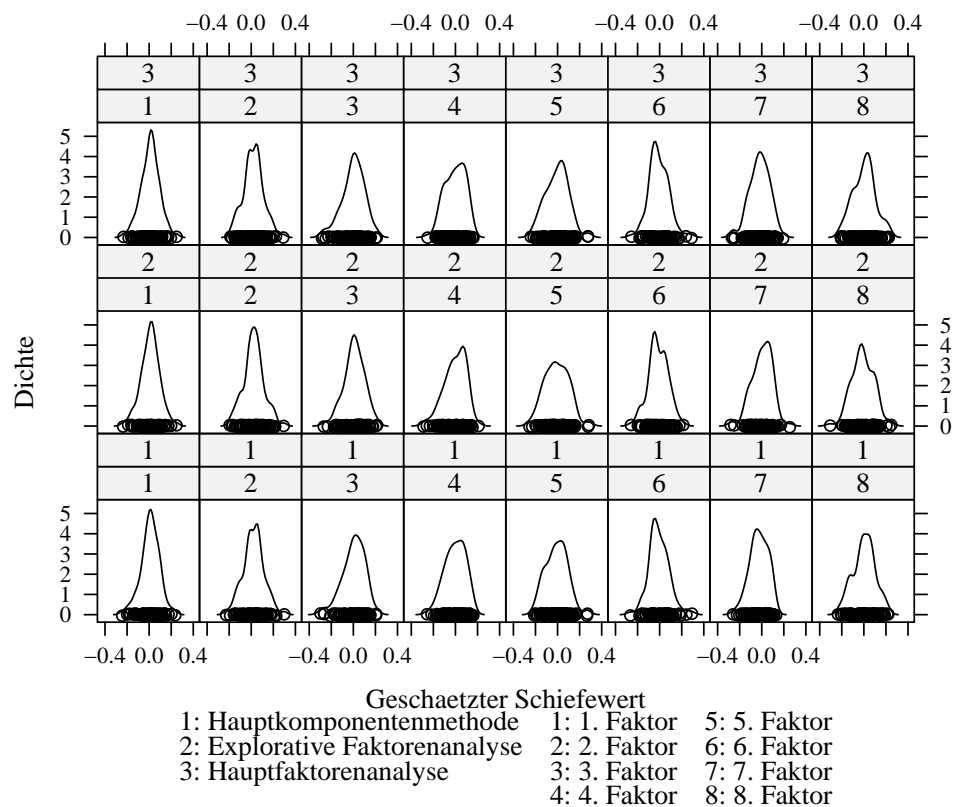


Abbildung C.1: Verteilung der geschätzten Schiefewerte in Abhängigkeit des Faktorenmodells und des Faktors ( $N = 600$ ;  $k = 8$ ;  $\mu_{3m} = 0$ )

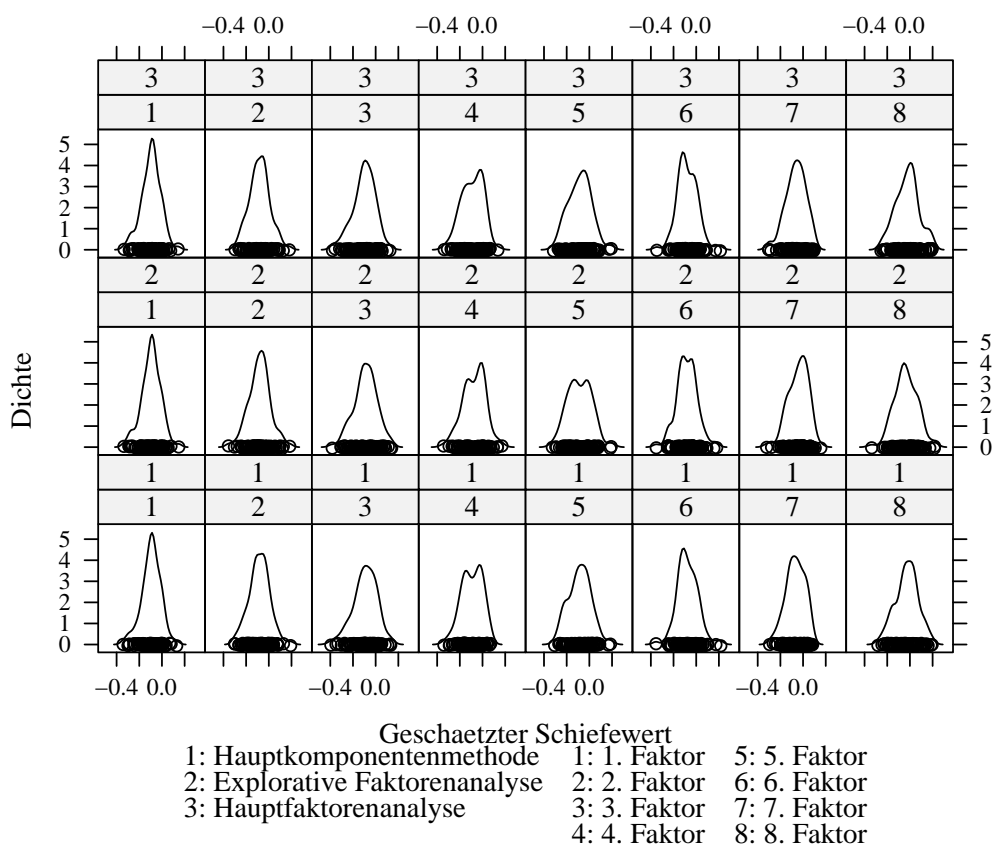


Abbildung C.2: Verteilung der geschätzten Schiefewerte in Abhängigkeit des Faktorenmodells und des Faktors ( $N = 600$ ;  $k = 8$ ;  $\mu_{3m} = -0.20$ )

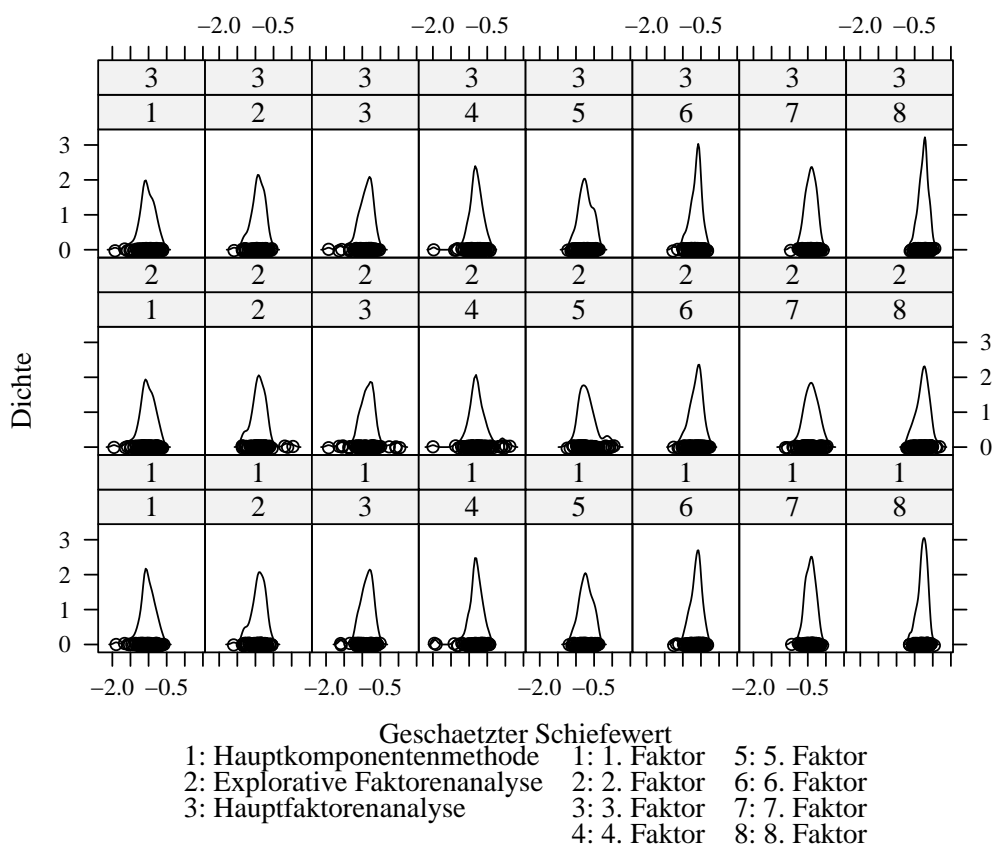


Abbildung C.3: Verteilung der geschätzten Schiefewerte in Abhängigkeit des Faktorenmodells und des Faktors ( $N = 600; k = 8; \mu_{3m} = -2.00$ )

# Anhang D

## Verteilung der absoluten Differenz

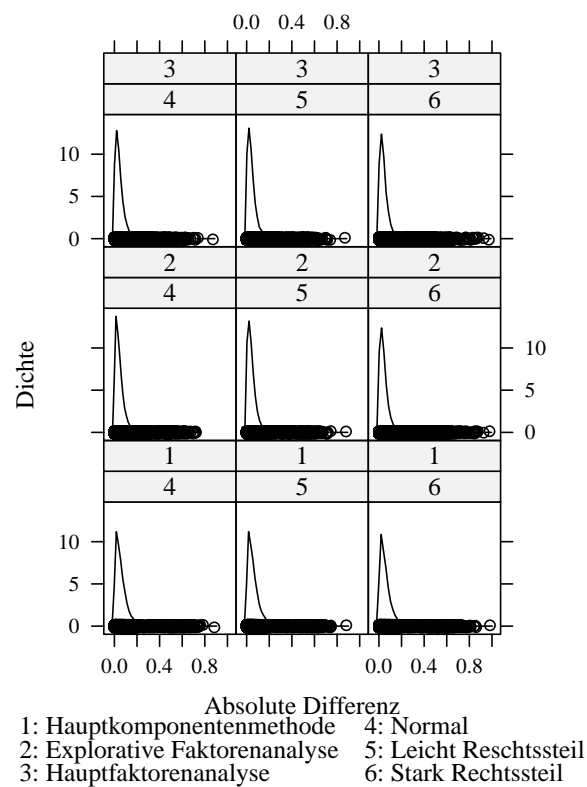


Abbildung D.1: Verteilung der absoluten Differenzen  $|\hat{l}_{ipk} - l_{pk}|$  in Abhängigkeit des Faktorenmodells und der latenten Schiefe ( $N = 600$ ;  $k = 4$ )

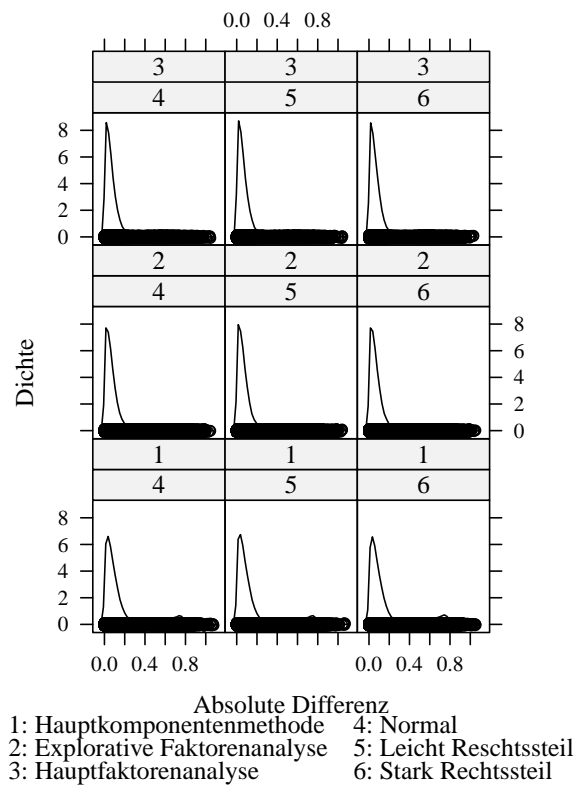


Abbildung D.2: Verteilung der absoluten Differenzen  $|\hat{l}_{ipk} - l_{pk}|$  in Abhängigkeit des Faktorenmodells und der latenten Schiefe ( $N = 200$ ;  $k = 8$ )

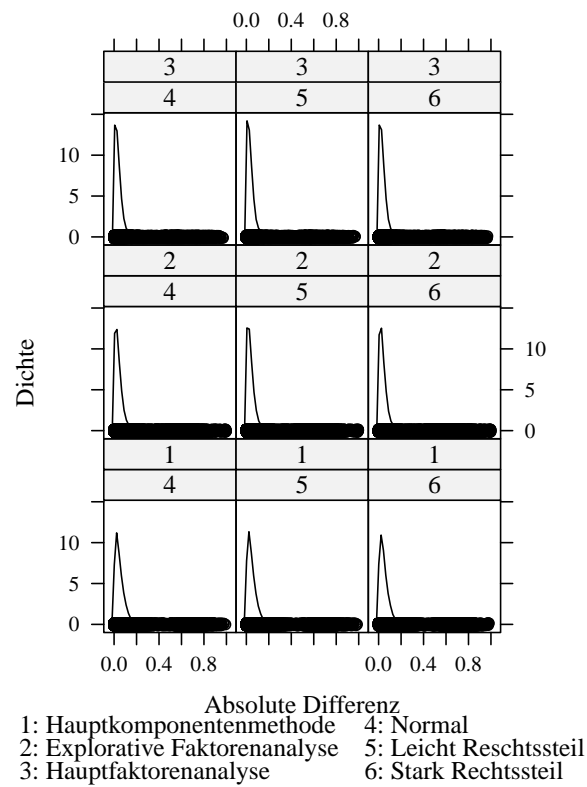


Abbildung D.3: Verteilung der absoluten Differenzen  $|\hat{l}_{ipk} - l_{pk}|$  in Abhängigkeit des Faktorenmodells und der latenten Schiefe ( $N = 600; k = 8$ )



# Anhang E

## Symbolverzeichnis

$\mathbf{A}, \mathbf{B}, \mathbf{C}$	Fettdruck, große Buchstaben: Matrizen
$\mathbf{a}, \mathbf{b}, \mathbf{c}$	Fettdruck, kleine Buchstaben: Spaltenvektoren
$a, b, c$	Normaldruck: Skalare
$n$	Anzahl der Messungen
$p$	Anzahl der Variablen
$s^2$	Varianz in der Stichprobe
$r_{vj}$	Korrelation der Variablen $v$ und $j$ in der Stichprobe
$\text{rg}(\mathbf{A})$	Rang von $\mathbf{A}$
$\det(\mathbf{A}) =  \mathbf{A} $	Determinante von $\mathbf{A}$
$\text{Tr}(\mathbf{A})$	Spur von $\mathbf{A}$
$\text{adj}(\mathbf{A})$	Adjunkte von $\mathbf{A}$
$\mathbf{A}^{-1}$	Inverse von $\mathbf{A}$
$E(\mathbf{A})$	Erwartungswert von $\mathbf{A}$
$\ \mathbf{a}\ $	Norm des Vektors $\mathbf{a}$
$\text{cov}(\mathbf{a})$	Kovarianzmatrix von $\mathbf{a}$
$\mathbf{a}_i$	$i$ -te Zeile der Matrix $\mathbf{A}$
$\mathbf{a}_j$	$j$ -te Spalte der Matrix $\mathbf{A}$
$\mathbf{x} \leftarrow \mathbf{a}$	$\mathbf{x}$ werden die Werte von $\mathbf{a}$ zugeordnet
$\mathbb{R}^{n \times p}$	Menge der reellen Zahlen im $n \times p$ Raum
$\mathbb{C}^{n \times p}$	Menge der komplexen Zahlen im $n \times p$ Raum
	unter der Bedingung, bzw. gegeben, daß...
$\forall_j$	gilt für alle $j$

# Literaturverzeichnis

- Allen, S. & Hubbard, R. (1986). Regression equations for the latent roots of random data correlation matrices with unities on the diagonal. *Multivariate Behavioral Research*, 21, 393–398.
- American Psychological Association. (1954). *Technical recommendations for psychological tests and diagnostics techniques*. Washington, D.C..
- American Psychological Association. (1966). *Standards for educational tests and manuals*. Washington, D.C..
- Anderson, E., Bai, Z., Bischof, C., Blackford, S., Demmel, J., Dongarra, J. et al. (1999). *LAPACK Users' Guide* (Dritte Aufl.). Philadelphia, PA: Society for Industrial and Applied Mathematics.
- Bai, Z. & Demmel, J. (1989). On a block implementation of Hessenberg multishift QR iteration. *International Journal of High Speed Computing*, 1, 97–112.
- Bartholomew, D., Knott, M. & Moustaki, I. (2011). *Latent variable models and factor analysis: A unified approach, 3rd edition*. New York: John Wiley & Sons.
- Basilevsky, A. (1994). *Statistical Factor Analysis and Related Methods*. New York: John Wiley & Sons.
- Bellman, R. (1960). *Introduction to matrix analysis*. New York: McGraw-Hill.
- Bock, R. D. & Lieberman, M. (1970). Fitting a response model for  $n$  dichotomously scored items. *Psychometrika*, 35, 179–197.
- Bortz, J. (1997). *Statistik für Sozialwissenschaftler*. Berlin: Springer Verlag.
- Braman, K., Byers, R. & Mathias, R. (2002). The Multishift QR Algorithm. Part I: Maintaining Well-Focused Shifts and Level 3 Performance. *SIAM Journal on Matrix Analysis and Applications*, 23, 929–947.
- Bravais, A. (1846). Analyse mathématique sur les probabilités des erreurs de situation de point. *Memoires presentes par divers savants a l'Academie des Sciences de l'Institut de France*, 9, 255–332.
- Browne, M. (1974). Generalized least squares estimators in the analysis of covariance structures. *South African Statistical Journal*, 8, 1–24.

- Browne, M. (1982). Covariance Structures. In D. Hawkins (Hg.), *Topics in applied multivariate analysis*. Cambridge: Cambridge University Press.
- Browne, M. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, *37*, 62–83.
- Broyden, C. G. (1967). Quasi-Newton methods and their application to function minimization. *Mathematics of Computation*, *21*, 368–381.
- Broyden, C. G. (1970a). The convergence of a class of double rank minimization algorithms, part I. *Journal of the Institute of Mathematics and its Applications*, *6*, 79–90.
- Broyden, C. G. (1970b). The convergence of a class of double rank minimization algorithms, part II. *Journal of the Institute of Mathematics and its Applications*, *6*, 222–231.
- Bunch, J., Nielsen, C. & Sorensen, D. (1978). Rank-One Modification of the Symmetric Eigenproblem. *Numerische Mathematik*, *31*, 31–48.
- Burt, C. (1909). Experimental tests of general intelligence. *British Journal of Psychology*, *3*, 94–177.
- Campbell, D. & Fiske, D. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*, 81–105.
- Carroll, J. (1945). The Effect of Difficulty and Chance Success on Correlations between Items or between Tests. *Psychometrika*, *10*, 1–19.
- Carroll, J. (1953). An analytical solution for approximating simple structure in factor analysis. *Psychometrika*, *18*, 23–38.
- Cattell, R. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, *1*, 245–276.
- Christoffersson, A. (1975). Factor analysis of dichotomized variables. *Psychometrika*, *40*, 5–32.
- Christoffersson, A. (1977). Two-step weighted least squares factor analysis of dichotomized variables. *Psychometrika*, *42*, 433–438.
- Collins, L., Cliff, N., McCormick, D. & Zarkin, J. (1986). Factor Recovery in Binary Data Sets: A Simulation. *Multivariate Behavioral Research*, *21* (3), 377–391.
- Cronbach, L. & Meehl, P. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*, 281–302.
- Cuppen, J. J. M. (1981). A Divide and Conquer Method for the Symmetric Tridiagonal Eigenproblem. *Numerische Mathematik*, *36*, 177–195.
- Davidon, W. C. (1959). *Variable metric method for minimization*. A.E.C. Research and Development Report ANL-5990.
- Demmel, J. W. (1997). *Applied numerical linear algebra*. Philadelphia: SIAM.

- Ding, J. & Zhou, A. (2007). Eigenvalues of rank-one updated matrices with some applications. *Applied Mathematics Letters*, 20, 1223–1226.
- Duncan, W. J. (1944). LXXVIII. Some devices for the solution of large sets of simultaneous linear equations – With an appendix on the reciprocation of partitioned matrices. *Philosophical Magazine Series 7*, 35, 660–670.
- Fahrmeir, L., Hamerle, A. & Tutz, G. (1996). *Multivariate statistische Verfahren*. Berlin: de Gruyter.
- Feller, W. (1971). *An introduction to probability theory and its applications* (Bd. II). New York: John Wiley & Sons.
- Ferguson, G. (1941). The Factorial Interpretation of Test Difficulty. *Psychometrika*, 6, 323–329.
- Ferguson, G. (1954). The concept of parsimony in factor analysis. *Psychometrika*, 19, 281–290.
- Fischer, G. (2009). *Lineare Algebra: Eine Einführung für Studienanfänger*. Wiesbaden: Vieweg+Teubner.
- Fisher, R. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London. Series A*, 222, 309–368.
- Fletcher, R. (1970). A new approach to variable metric algorithm. *The Computer Journal*, 13, 317–322.
- Fletcher, R. (2006). *Practical Methods of Optimization*. Chichester: John Wiley & Sons.
- Fletcher, R. & Powell, M. J. D. (1963). A rapidly convergent descent method for minimization. *The Computer Journal*, 6, 163–168.
- Fletcher, R. & Reeves, C. (1964). Function minimization by conjugate gradients. *Computer Journal*, 7, 149–154.
- Forster, O. (1999). *Analysis 1*. Braunschweig: Vieweg.
- Francis, J. G. F. (1961). The QR Transformation. A Unitary Analogue to the LR Transformation - Part 1. *The Computer Journal*, 4, 265–271.
- Francis, J. G. F. (1962). The QR Transformation - Part 2. *The Computer Journal*, 4, 332–345.
- Gill, P., Murray, W. & Wright, M. (1981). *Practical Optimization*. London: Academic Press.
- Glaser, R. (1963). Instructional technology and the measurement of learning outcomes: Some questions. *American Psychologist*, 18, 519–521.
- Goldfarb, D. (1970). A family of variable metric methods derived by variational means. *Mathematics of Computation*, 24, 23–26.
- Golub, G. H. & Van der Vorst, H. A. (2000). Eigenvalue computation in the 20th century. *Journal of Computational and Applied Mathematics*, 123, 36–65.
- Green, S. (1983). Identifiability of Spurious Factors Using Linear Factor

- Analysis with Binary Items. *Applied Psychological Measurement*, 7, 139–147.
- Gu, M. & Eisenstat, S. (1992). *A Stable and Efficient Algorithm for the Rank-one Modification of the symmetric Eigenproblem* (Bericht). New Haven (CT): Yale University. Available from <http://www.cs.yale.edu/publications/techreports/tr916.pdf>
- Gu, M. & Eisenstat, S. (1995). A Divide-and-Conquer algorithm for the symmetric tridiagonal eigenproblem. *SIAM Journal on Matrix Analysis and Applications*, 16, 172–191.
- Guttman, L. (1946). Enlargement Methods for Computing the Inverse Matrix. *The Annals of Mathematical Statistics*, 17, 336–343.
- Guttman, L. (1954). Some necessary conditions for common factor analysis. *Psychometrika*, 19, 149–161.
- Hadlock, C. R. (1978). *Field Theory and Its Classical Problems*. Washington, DC: Mathematical Association of America.
- Hager, W. W. (1989). Updating the Inverse of a Matrix. *SIAM Review*, 31, 221–239.
- Harman, H. (1967). *Modern Factor Analysis*. Chicago: The University of Chicago Press.
- Henderson, H. V. & Searle, S. R. (1981). On Deriving the Inverse of a Sum of Matrices. *SIAM Review*, 23, 53–60.
- Horn, J. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30, 179–185.
- Hotelling, H. (1933a). Analysis of a complex of statistical variables into principal components. *The Journal of Educational Psychology*, 24, 417–441.
- Hotelling, H. (1933b). Analysis of a complex of statistical variables into principal components. *The Journal of Educational Psychology*, 24, 498–520.
- Householder, A. S. (1958). Unitary Triangularization of a Nonsymmetric Matrix. *Journal of the ACM (JACM)*, 5, 339–342.
- IEA. (2007). *PIRLS 2006 Assessment*. Boston: TIMSS & PIRLS International Study Center.
- Jöreskog, K. G. (1966). Testing a simple structure hypothesis in factor analysis. *Psychometrika*, 31, 165–178.
- Jöreskog, K. G. (1967). Some contributions to maximum likelihood factor analysis. *Psychometrika*, 32, 443–482.
- Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, 34, 183–202.
- Jöreskog, K. G. (1972). Factor analysis by generalized least squares. *Psychometrika*, 37, 243–260.

- Jöreskog, K. G. (1978). Structural analysis of covariance and correlation matrices. *Psychometrika*, *43*, 443–477.
- Kaiser, H. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, *23*, 187–200.
- Kaiser, H. & Dickman, K. (1959). Analytic determination of common factors. *American Psychologist*, *14*, 425.
- Kasper, D. (2006). *Modelle der Faktorenanalyse. Kovarianzstrukturanalysen mit dem COSAN-Modell*. Berlin: Logos.
- Kelley, T. L. (1935). *Essential traits of mental life*. Cambridge (MA): Harvard University Press.
- Klartag, B. (2007). A central limit theorem for convex sets. *Inventiones mathematicae*, *168*, 91–131.
- Klauer, K. (1987). *Kriteriumsorientierte Tests. Lehrbuch der Theorie und Praxis lehrzielorientierten Messens*. Göttingen: Hogrefe.
- Kölbl, C., Tiedemann, J. & Billmann-Mahecha, E. (2006). Die Bedeutung der Lesekompetenz für Sachfächer. *Psychologie in Erziehung und Unterricht*, *53*, 201–212.
- Kublanovskaya, V. N. (1962). On some algorithms for the solution of the complete eigenvalue problem. *USSR Computational Mathematics and Mathematical Physics*, *1*, 637–657.
- Lautenschlager, G., Lance, C. & Flaherty, V. (1989). Parallel analysis criteria: Revised equations for estimating the latent roots of random data correlation matrices. *Educational and Psychological Measurement*, *49*, 339–345.
- Lawley, D. & Maxwell, A. (1963). *Factor Analysis as a Statistical Method*. London: Butterworths.
- Lazarsfeld, P. (1950). The logical and mathematical foundation of latent structure analysis. In S. Stouffer, L. Guttman, E. Suchman, P. Lazarsfeld, S. Star & J. Clausen (Hg.), *Measurement and Prediction*. New Jersey: Princeton University Press.
- Levenberg, K. (1944). A method for the solution of certain nonlinear problems in least squares. *Quarterly of applied mathematics*, *2*, 164–168.
- Lienert, G. & Raatz, U. (1998). *Testaufbau und Testanalyse*. Weinheim: Psychologie Verlags Union.
- Lindeberg, J. W. (1922). Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung. *Mathematische Zeitschrift*, *15*, 211–225.
- Longman, R., Cota, A., Holden, R. & Fekken, G. (1989). A regression equation for the parallel analysis criterion in principle component analysis: Means and 95th percentile eigenvalues. *Multivariate Behavioral Research*, *24*, 59–69.

- Magnus, J. & Neudecker, H. (1999). *Matrix differential calculus*. Chichester: John Wiley & Sons.
- Martin, M., Mullis, I. & Kennedy, A. (2007). *PIRLS 2006 Technical Report*. Boston: TIMSS & PIRLS International Study Center.
- Martin, W. S., Fruchter, B. & Mathis, W. (1974). An Investigation of the Effect of the Number of Scale Intervals on Principal Components Factor Analysis. *Educational and Psychological Measurement*, 34, 537–545.
- Mattson, S. (1997). How to generate non-normal data for simulation of structural equation models. *Multivariate Behavioral Research*, 32, 355–373.
- Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin*, 105, 156–166.
- Mises, R. von & Pollaczek-Geiring, H. (1929). Praktische Verfahren der Gleichungsauffösung. *Zeitschrift für angewandte Mathematik und Mechanik*, 9, 58–77, 152–164.
- Moore, E. H. (1920). On the reciprocal of the general algebraic matrix. In A. Dresden: The fourteenth western meeting of the American Mathematical Society. *Bulletin of the American Mathematical Society*, 26, 385–396.
- Moustaki, I. & Knott, M. (2000). Generalised latent trait models. *Psychometrika*, 65, 391–411.
- Mullis, I., Martin, M. O., Kennedy, A. M., Trong, K. L. & Sainsbury, M. (2009). *PIRLS 2011 Assessment Framework*. Boston: TIMSS & PIRLS International Study Center.
- Mullis, I., Martin, M. O., Ruddock, G. J., O’Sullivan, C. Y., Arora, A. & Erberber, E. (2005). *TIMSS 2007 Assessment Frameworks*. Boston: TIMSS & PIRLS International Study Center.
- Muthén, B. (1978). Contributions to factor analysis of dichotomous variables. *Psychometrika*, 43, 551–560.
- Muthén, B. (1983). Latent variable structural equation modeling with categorical data. *Journal of Econometrics*, 22, 43–65.
- Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika*, 49, 115–132.
- Muthén, B. & Christoffersson, A. (1981). Simultaneous factor analysis of dichotomous variables in several groups. *Psychometrika*, 46, 407–419.
- Neuenschwander, M. & Kunz-Makarova, E. (2000). *Schule und Identität im Jugendalter III. Kurzdokumentation der Skalen und Stichproben der Lehrerbefragung*. Institut für Pädagogik, Abteilung Pädagogische Psychologie der Universität Bern.
- Neuhaus, J. & Wrigley, C. (1954). The quartimax method: an analytical

- approach to orthogonal simple structure. *British Journal of Statistical Psychology*, 7, 81–91.
- OECD. (2003). *PISA 2003 Technical Report*. Paris: Organisation for Economic Co-operation and Development.
- OECD. (2009a). *PISA 2006 Technical Report*. Paris: Organisation for Economic Co-operation and Development.
- OECD. (2009b). *PISA 2009 Assessment Framework. Key competencies in reading, mathematics and science*. Paris: Organisation for Economic Co-operation and Development.
- Olson, J. F., Martin, M. O. & Mullis, I. V. (2008). *TIMSS 2007 Technical Report*. Boston: TIMSS & PIRLS International Study Center.
- Pearson, K. (1901). LIII. On lines and planes of closest fit to a system of points in space. *Philosophical Magazine Series 6*, 2, 559–572.
- Penrose, R. (1955). A generalized inverse for matrices. *Mathematical Proceedings of the Cambridge Philosophical Society*, 51, 406–413.
- Polak, E. (1971). *Computational Methods in Optimization: A Unified Approach*. New York: Academic Press.
- R Development Core Team. (2010). R: A Language and Environment for Statistical Computing [Software-Handbuch]. Vienna, Austria. Available from <http://www.R-project.org>
- Ramberg, J. S., Tadikamalla, P. R., Dudewicz, E. J. & Mykytka, E. F. (1979). A Probability Distribution and Its Uses in Fitting Data. *Technometrics*, 21, 201–214.
- Reinartz, W. J., Echambadi, R. & Chin, W. W. (2002). Generating Non-normal Data for Simulation of Structural Equation Models Using Mattson's Method. *Multivariate Behavioral Research*, 37, 227–244.
- Roznowski, M., Tucker, L. & Humphreys, L. (1991). Three Approaches to Determining the Dimensionality of Binary Items. *Applied Psychological Measurement*, 15, 109–127.
- Schaback, R. & Wendland, H. (2004). *Numerische Mathematik*. Berlin: Springer Verlag.
- Searle, S. R. (1982). *Matrix algebra useful for statistics*. New York: John Wiley & Sons.
- Shanno, D. F. (1970). Conditioning of quasi-newton methods for function minimization. *Mathematics of Computation*, 24, 647–656.
- Sherman, J. & Morrison, W. J. (1950). Adjustment of an Inverse Matrix Corresponding to a Change in One Element of a Given Matrix. *The Annals of Mathematical Statistics*, 21, 124–127.
- Skondal, A. & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling: Multilevel, longitudinal and structural equation modeling*. London: Chapman & Hall.



- Sparfeldt, J. R., Rost, D. H. & Schilling, S. R. (2004). Schulfachspezifische Interessen - ökonomisch gemessen. *Psychologie in Erziehung und Unterricht*, 51, 213–220.
- Spearman, C. (1904). "General Intelligence," objectively determined and measured. *The American Journal of Psychology*, 15, 201–292.
- Steyer, R. & Eid, M. (1993). *Messen und Testen*. Heidelberg: Springer Verlag.
- Strang, G. (2009). *Introduction to Linear Algebra*. Wellesley: Wellesley-Cambridge Press.
- Stroock, D. W. (1993). *Probability theory, an analytic view*. New York: Cambridge University Press.
- Sturzbecher, D., Kasper, D., Bönninger, J. & Rüdell, M. (2008). *Evaluation der Theoretischen Fahrerlaubnisprüfung. Methodische Konzeption und Ergebnisse des Revisionsprojekts*. Dresden: TÜV|DEKRA arge tp 21.
- Styan, G. P. H. (1973). Hadamard products and multivariate statistical analysis. *Linear Algebra and its Applications*, 6, 217 – 240.
- Thurstone, L. L. (1931). Multiple factor analysis. *Psychological Review*, 38, 406–427.
- Thurstone, L. L. (1947). *Multiple factor analysis*. Chicago: University of Chicago Press.
- Thurstone, L. L. (1965). *Multiple factor analysis*. Chicago (IL): University of Chicago Press.
- Wang, C. (2001). *Effect of number of response categories and scale distribution on factor analysis*. Unveröffentlichte Diplomarbeit, National Taiwan University, Taiwan: Taipei.
- Watkins, D. S. (2007). *The Matrix Eigenvalue Problem. GR and Krylov Subspace Methods*. Philadelphia: Society for Industrial and Applied Mathematics.
- Weng, L. & Cheng, C. (2005). Parallel Analysis with Unidimensional Binary Data. *Educational and Psychological Measurement*, 65, 697–716.
- Werner, J. (1997). *Lineare Statistik. Das Allgemeine Lineare Modell*. Weinheim: Psychologie Verlags Union.
- Wu, M., Adams, R. & Wilson, M. (1997). *Conquest: Generalized item response modeling software*. Victoria, Australia: ACER.
- Zwick, W. & Velicer, W. (1986). Comparison of five rules of determining the number of components to retain. *Psychological Bulletin*, 99, 432–442.

## Erklärung

Gemäß §12 PromO 2008, Absatz 3, Satz 4 versichere ich, dass die Dissertation selbständig verfasst und alle in Anspruch genommenen Quellen und Hilfen in der Dissertation vermerkt wurden. Außerdem erkläre ich gemäß §12 PromO 2008, Absatz 3, Satz 5, dass die Dissertation in der gegenwärtigen oder in einer anderen Fassung nicht bereits an der Technischen Universität Dortmund oder an einer anderen Hochschule im Zusammenhang mit einer staatlichen oder akademischen Prüfung vorgelegt worden ist.

---

Ort, Datum

---

Unterschrift