

ZUR INTERPRETATION VON HAUPTKOMPONENTEN UND FAKTOREN:
EINE WIEDERAUFNAHME DER KRITISCHEN
DISKUSSION FAKTORENANALYTISCHER VERFAHREN¹⁾

Mit 3 Abbildungen und 6 Tabellen

HARALD BATHOLT und WOLF-DIETER ERB

Summary: Interpretation of principal components and factors: new aspects in the critical discussion of factor analysis

The article stresses problems associated with real world applications and interpretations of factor analysis. Although focussed on a spatial context, the results and implications are not only limited to geographical fields. Originally it was aimed to make use of factor analysis to analyse the economic structure of the administrative districts of Hesse. Rotated factor solutions were obtained through the varimax method and factor scores computed accordingly. Surprisingly, many of the rotated factor scores were not consistent with the original variable scores. For instance, the economic performance of Giessen county was considerably underestimated as compared to 11 other counties if factor scores of the corresponding factor were used instead of variable scores of the high-loading variables. Similar problems were identified for each factor in various solutions. If regional policy in Hesse had to be based on the results of the factor analysis presented, wrong conclusions could be drawn from the factor score distributions.

Although the problems identified seem quite confusing, an explanation can be more easily given than might be expected. Inconsistent factor score distributions are directly related to the procedure of orthogonal rotation which results in compensation effects between variables. If variables originally correlating are associated with different factors, a compensation effect will have to take place between the corresponding variables in order to assure orthogonality of the factors extracted. From these characteristics, several implications can be derived. First, factor analysis should not be used without sufficient mathematical background. Secondly, in the case of rotated factor solutions, the interpretation of the factors cannot be solely based on factor loadings as generally recommended. In addition, factor score coefficients have to be used for interpretation purposes to understand the compensation effects between variables. Thirdly, subsequent analysis should be based on surrogate variable scores instead of factor scores in order to assure political applicability of results. Alternatively, oblique rotation procedures might be implemented, as opposed to the orthogonal case.

1 Einführung

Seit Beginn der 60er Jahre zählen Hauptkomponenten- bzw. Faktorenanalysen zu den Standardverfahren im statistischen Methodenspektrum der

Quantitativen Geographie. Nachdem die Verfahren aus dem angloamerikanischen Raum in die deutschsprachige Geographie übertragen worden waren (z. B. STEINER 1965, KILCHENMANN 1968), erlebten sie in den 60er und 70er Jahren geradezu einen Boom. In dieser Zeit wurden sie zu den beliebtesten Verfahren der Quantitativen Geographie im deutschsprachigen Bereich (GIESE 1981, S. 258). Nach einer gewissen Anfangseuphorie mehrten sich jedoch Stimmen, die vor einer allzu kritiklosen Anwendung warnten (vgl. BAHRENBERG u. GIESE 1975b, KEMPER u. SCHMIEDECKEN 1977, GIESE 1978).

In dem vorliegenden Artikel soll die kritische Auseinandersetzung erneut aufgenommen und ein Problemfeld faktorenanalytischer Verfahren diskutiert werden, das nach unserer Meinung bisher zu wenig Beachtung gefunden hat. Im Anschluß an eine kurze Einführung in die methodischen Grundzüge wird eine für die Geographie typische Anwendung von Hauptkomponenten- bzw. Faktorenanalysen vorgestellt: die Analyse der regionalen Wirtschaftsstruktur am Beispiel der hessischen Stadt- und Landkreise. Die Untersuchung führt zu teilweise überraschenden Ergebnissen in der Konstellation von Hauptkomponenten- bzw. Faktorenwerten, die sowohl aus theoretischen Überlegungen als auch aus der Kenntnis der örtlichen Gegebenheiten nicht plausibel sind. Es soll gezeigt werden, worauf diese Abweichungen zurückzuführen sind, um davon ausgehend im abschließenden Teil Empfehlungen für einen sorgfältigeren Umgang mit faktorenanalytischen Verfahren zu geben. Es sei an dieser Stelle ausdrücklich darauf hingewiesen, daß die festgestellten Inkonsistenzen²⁾ nicht auf

¹⁾ Wir danken den Herren Prof. Dr. Fahrmeir (Regensburg), Prof. Dr. Giese (Gießen) und Prof. Dr. Nipper (Köln) für ihre Anregungen zur Überarbeitung einer ersten Fassung des vorliegenden Artikels.

²⁾ Es handelt sich dabei nicht um tatsächliche, sondern um scheinbare Inkonsistenzen, die aus einer unvollständigen Interpretation der extrahierten Hauptkomponenten bzw. Faktoren resultieren. Vgl. im folgenden ausführlich BATHOLT u. ERB (1991).

einem speziell konstruierten Datensatz beruhen, sondern das Nebenprodukt einer sachinhaltlich abgeleiteten Untersuchung darstellen.

2 Idee und Zielsetzung faktorenanalytischer Verfahren

In den Sozial- und Wirtschaftswissenschaften stellt sich häufig die Aufgabe, aus einer Vielzahl von Einflußfaktoren für eine konkrete Problemstellung diejenigen zu extrahieren, die unabhängig voneinander zur Erklärung eines untersuchten Sachverhalts beitragen. Eine solche Fragestellung liegt vor allem dann vor, wenn komplexe Merkmale wie z. B. die regionale Wirtschaftsstruktur analysiert werden sollen, die nicht direkt meßbar sind, sondern nur über eine Vielzahl von Einzelindikatoren erfaßt werden können. In diesem Zusammenhang stellen faktorenanalytische Verfahren ein geeignetes methodisches Instrumentarium dar, um einen Datensatz auf die notwendige Menge von Faktoren zu reduzieren, die unabhängig voneinander zur Beschreibung der Zielgröße beitragen. Faktorenanalytische Verfahren sind zum einen eine Methode zur Datenreduktion (Beschränkung der Variablenzahl), zum anderen eine Methode zur Identifikation von Hintergrundinformationen und schließlich eine Methode zur Erzeugung unabhängiger Einflußgrößen (vgl. ÜBERLA 1971, S. 1 ff.; BACKHAUS et al. 1987, S. 67 ff.; BAHRENBERG u. GIESE 1975 a, S. 205 ff.).

Die in den Ausgangsvariablen enthaltenen Informationen und Zusammenhänge können komprimiert in Form der Korrelationsmatrix R zwischen allen Variablen dargestellt werden. Für den Fall standardisierter Ausgangsvariablen läßt sich die Korrelationsmatrix über ein einfaches Matrizenprodukt schnell berechnen³⁾:

$$R_{m,m} = Z'_{m,n} Z_{n,m} \quad [1]$$

Das Ziel der Hauptkomponentenanalyse (eines Spezialfalls faktorenanalytischer Verfahren) besteht nun darin, die m Ausgangsvariablen (Matrix Z) durch r noch zu identifizierende hypothetische Hauptkomponenten (Matrix F) abzubilden. Entsprechend Gleichung [2] sollen die Ausgangsvariablen durch

eine Linearkombination der Hauptkomponenten vollständig dargestellt werden. Die Hauptkomponenten sollen zwei Bedingungen erfüllen, die in Gleichung [3] zum Ausdruck kommen: Einerseits sollen die Hauptkomponenten standardisiert und andererseits voneinander unabhängig sein. Sind diese Bedingungen erfüllt, so stellen die Koeffizienten in Gleichung [2] (Matrix der Ladungen L) zugleich auch Korrelationskoeffizienten zwischen den Ausgangs- und den Hintergrundvariablen dar.

$$\text{Ziel:} \quad Z_{n,m} = F_{n,r} L'_{r,m} \quad [2]$$

$$\text{Voraussetzung:} \quad I_{r,r} = F'_{r,n} F_{n,r} = \text{Einheitsmatrix} \\ \text{der Ordnung } r \quad [3]$$

Durch Einsetzen der Gleichungen [2] und [3] in Gleichung [1] gelangt man zum sogenannten Fundamentalthem der Faktorenanalyse (siehe Gleichung [4]), das die Beziehungen zwischen den Hintergrundvariablen und der Korrelationsmatrix der Ausgangsvariablen beschreibt (vgl. ÜBERLA 1971, S. 50 ff.; BACKHAUS et al. 1987, S. 74 ff.; BAHRENBERG u. GIESE 1975 a, S. 210 f.):

$$R_{m,m} = Z'_{m,n} Z_{n,m} = (L_{m,r} F'_{r,n}) (F_{n,r} L'_{r,m}) \\ = L_{m,r} L'_{r,m} \quad [4]$$

Ausgangspunkt der Hauptkomponentenanalyse und anderer faktorenanalytischer Verfahren ist also die Korrelationsmatrix der Variablen R . Im Unterschied zur Hauptkomponentenanalyse geht die Faktorenanalyse im engeren Sinn nicht von der Annahme aus, daß die Ausgangsvariablen vollständig durch eine Linearkombination von Hintergrundvariablen beschrieben werden können (siehe Gleichung [2]). Bei einer Faktorenanalyse im engeren Sinn sind deshalb die Hauptdiagonalelemente der Korrelationsmatrix (die sog. Kommunalitäten) nicht gleich Eins (Varianz einer standardisierten Ausgangsvariablen) zu setzen wie im Fall der Hauptkomponentenanalyse, sondern a priori kleiner als Eins.

Das am häufigsten verwendete Verfahren zur Extraktion der Hintergrundvariablen ist die sog. Hauptachsenmethode. Rechnerisch ergeben sich die Ladungen der ersten Hauptkomponente als Eigenvektor des größten Eigenwerts der Korrelationsmatrix R . Der Ladungsvektor der zweiten Hauptkomponente ist identisch mit dem Eigenvektor des zweitgrößten Eigenwerts der Korrelationsmatrix R usw. (vgl. HARMAN 1967, S. 137 ff.; FAHRMEIR u. Hamerle 1984, S. 597 ff.). Der Prozeß der Hauptkomponentenextraktion wird beendet, wenn ein vorgegebenes Abbruchkriterium erfüllt ist (z. B. Varianzanteil, Kaiser-Kriterium oder Scree-Plot; vgl. BAH-

³⁾ Es gelte folgende Notation: n : Anzahl der Raumeinheiten, m : Anzahl der Originalvariablen, r : Anzahl der Hintergrundvariablen (Faktoren bzw. Hauptkomponenten), Z : Standardisierte Datenmatrix, R : Korrelationsmatrix, L : Ladungsmatrix, F : Matrix der Faktoren- bzw. Hauptkomponentenwerte und C : Matrix der Factor Score Coefficients.



Abb. 1: Übersicht der hessischen Stadt- und Landkreise
Administrative districts (Kreise) of Hesse

RENBERG u. GIESE 1975 a, S. 214; BACKHAUS et al. 1987, S. 90 f.; FAHRMEIR u. HAMERLE 1984, S. 603 ff.).

Der nächste Schritt besteht in der Benennung (Interpretation) der Hauptkomponenten, denn die Hintergrundvariablen sind zunächst rein abstrakte, hypothetische Konstrukte ohne inhaltlichen Bezug. Zur Interpretation der Hauptkomponenten verwendet man üblicherweise die berechneten Ladungen, die die Korrelationen zwischen den Hauptkomponenten und den Ausgangsvariablen darstellen. Man versucht mit Hilfe der betragsmäßig am höchsten ladenden Ausgangsvariablen, einen sinnvollen Oberbegriff für jede der extrahierten Hauptkomponenten zu finden (vgl. ÜBERLA 1971, S. 175 ff.; BAH-

RENBERG u. GIESE 1975 a, S. 229 ff.; BACKHAUS et al. 1987, S. 91 f.; FAHRMEIR u. HAMERLE 1984, S. 616 ff.).

Häufig ist eine sinnvolle Interpretation der Hintergrundvariablen anhand der Ladungsmuster jedoch nicht möglich, weil z. B. eine einzelne Hauptkomponente mit fast allen Ausgangsvariablen hoch korreliert, die anderen Hintergrundvariablen dagegen insgesamt vom Betrag her sehr kleine Ladungen besitzen, oder weil sich die Ladungen für eine Variable gleichmäßig auf mehrere Hauptkomponenten aufteilen. Um ein besser interpretierbares Ladungsmuster zu erhalten, führt man im Anschluß an den Extraktionsprozeß in der Regel eine Rotation der Hauptkomponenten (unter Wahrung der Orthogo-

nalitätsbedingung) durch. Das gängigste Rotationsverfahren ist die sog. Varimax-Methode. Bei einer Varimax-Rotation wird eine Einfachstruktur der Ladungsmatrix L angestrebt, so daß die Ladungen je Hauptkomponente entweder möglichst nahe Null oder möglichst nahe Eins liegen (vgl. HARMAN 1967, S. 293 ff.; BAHRENBERG u. GIESE 1975 a, S. 217 ff.; KEMPER 1975).

Der letzte Schritt einer Hauptkomponentenanalyse besteht in einem Rückschluß von den hypothetischen Hintergrundvariablen auf die Untersuchungsobjekte. Im Fall einer Hauptkomponentenanalyse lassen sich die objektbezogenen Werte der extrahierten Hintergrundvariablen (Matrix F) durch algebraische Umformung von Gleichung [2] exakt berechnen. Die Hauptkomponentenwerte ergeben sich durch eine einfache Matrizenmultiplikation der Matrix der standardisierten Variablen Z mit einer Koeffizientenmatrix C in Form von Gleichung [5]:

$$F_{n,r} = Z_{n,m} C_{m,r} = Z_{n,m} L_{m,r} (L'_{r,m} L_{m,r})^{-1} \quad [5]$$

3 Eine typische Anwendung faktorenanalytischer Verfahren in der Wirtschaftsgeographie: Die Analyse der regionalen Wirtschaftsstruktur in Hessen⁴⁾

Für die vorliegende Untersuchung wurden faktorenanalytische Verfahren herangezogen, um eine regionalwirtschaftliche Strukturanalyse auf der Basis der 26 Stadt- und Landkreise in Hessen durchzuführen (vgl. Abb. 1). Von allen Bundesländern bildet Hessen eines der interessantesten Objekte für eine solche Analyse, denn Hessen liegt als zentrales Bundesland sozusagen an der „Nahtstelle“ zwischen dem norddeutschen und dem süddeutschen Raum. Die Hypothese eines großräumigen regionalwirtschaftlichen Süd-Nord-Strukturgefälles in der Bundesrepublik Deutschland (wenn sie auch nicht vollends bestätigt ist) läßt sich in kleinerem Maßstab auf das Bundesland Hessen übertragen.

Die Variablen für die regionalwirtschaftliche Strukturanalyse wurden so ausgewählt, daß die vier Merkmalskategorien (Indikatoren) Beschäftigungssituation, Einkommenssituation, Bildungsniveau

und Verkehrsinfrastruktur mit etwa gleichem Gewicht (gleiche Variablenzahl) erfaßt wurden. In Tab. 1 sind die exakten Definitionen der einbezogenen Variablen mit Quellenangaben aufgelistet. Die in Tab. 2 dargestellte Korrelationsmatrix der Variablen bildete den Ausgangspunkt für eine Hauptkomponentenanalyse und eine Maximum-Likelihood (ML)-Faktorenanalyse. Die extrahierten Hauptkomponenten bzw. Faktoren wurden einer orthogonalen Rotation nach dem Varimax-Prinzip unterworfen, um zu einer sinnvoll interpretierbaren Lösung der Ergebnisse zu gelangen⁵⁾. Für beide angewendeten Verfahren erwies sich eine Lösung mit drei Hauptkomponenten bzw. Faktoren als sinnvoll (vgl. Tab. 3).

In der einschlägigen Fachliteratur (vgl. z. B. FAHRMEIR u. HAMERLE 1984, S. 616 f.; ÜBERLA 1971, S. 175 ff.; COOLEY u. LOHNES 1971, S. 144 f.; HARMAN 1967, S. 309 ff.; BAHRENBERG u. GIESE 1975 a, S. 229 ff.) wird vorgeschlagen, die Interpretation der extrahierten Hauptkomponenten bzw. Faktoren anhand der rotierten Faktorladungsmatrix vorzunehmen. Dazu betrachtet man jede Hauptkomponente (bzw. jeden Faktor) isoliert und versucht, mit Hilfe der hohen Faktorladungen (die ja zugleich Korrelationskoeffizienten mit den Ausgangsvariablen darstellen) einen inhaltlich sinnvollen Oberbegriff für die rechnerisch ermittelten Hintergrundvariablen zu finden. Die für die regionalwirtschaftliche Strukturanalyse der hessischen Kreise ermittelten Ladungen der extrahierten Hauptkomponenten und Faktoren nach einer Varimax-Rotation sind in Tab. 4 dargestellt. Anhand derjenigen Ladungen, die den Wert 0.60 übersteigen (alle hohen Ladungen sind im übrigen positiv), lassen sich die Hauptkomponenten (analog die ML-Faktoren) wie folgt benennen (vgl. Tab. 4):

1. Die Hauptkomponente HK 1 korreliert hoch mit den Variablen Beschäftigtenquote (BQ), Bruttowertschöpfung (BWS), Lohn- und Gehaltssumme (LUG), Autobahndichte (AD) und Straßendichte (SD). Demnach kann die erste Hauptkomponente als Indikator für die *materielle Wirtschaftskraft* der hessischen Kreise angesehen werden. Geeignete Stellvertretervariablen (mit hohen Ladungen und Kommunalitäten) für diese Hauptkomponente

⁴⁾ Die vorliegende Untersuchung schließt an ein Seminar von Prof. Dr. Giese an, der uns freundlicherweise die Daten zur Verfügung stellte und wesentliche Anregungen zur Konzipierung dieses Artikels gab. Herzlicher Dank gilt auch Matthias Höher für die Hilfe bei der Datenaufbereitung sowie Andreas Kristen für die Digitalisierung der Kartengrundlage.

⁵⁾ Die Berechnungen wurden mit den Programmpaketen SPSS/PC +, der SPSS-X Großrechnerversion und mit GEOPRO (einem von Prof. Dr. Nipper und Dr. Erb am Geographischen Institut der Justus-Liebig-Universität Gießen entwickelten Statistikprogramm zur Bearbeitung regionalwirtschaftlicher Fragestellungen) durchgeführt.

Tabelle 1: Indikatorenkatalog

List of indicators

Indikator/Variable	Variablendefinition	Quelle
<i>Beschäftigungssituation</i>		
BQ Beschäftigtenquote	Zahl der Beschäftigten in % der Zahl der Erwerbspersonen (1987)	VZ/AZ 1987, eigene Berechnungen
ALQ Arbeitslosenquote	Zahl der Arbeitslosen in % der Zahl der abhängigen Erwerbspersonen (Sept. 1988)	Landesarbeitsamt Hessen, eigene Berechnungen
ELQ Erwerbslosenquote	Zahl der Erwerbslosen und Arbeitssuchenden in % der Zahl der Erwerbspersonen (1987)	VZ 1987
<i>Einkommenssituation</i>		
BWS Bruttowertschöpfung	Bruttowertschöpfung zu Faktorkosten in DM je Einwohner (1986)	Hess. Kreiszahlen 1988
LUG Lohn- und Gehaltssumme	Lohn- und Gehaltssumme in DM je Arbeitnehmer (1987)	AZ 1987
<i>Bildungsniveau</i>		
ABQ Abiturientenquote	%-Anteil der Wohnbevölkerung mit Hochschul-/ Fachhochschulreife an der Wohnbevölkerung im Alter zwischen 15 und 65 Jahren (1987)	VZ 1987
AKQ Akademikerquote	%-Anteil der Wohnbevölkerung mit Hochschul-/ Fachhochschulabschluss an der Wohnbevölkerung zwischen 15 und 65 Jahren (1987)	VZ 1987
<i>Verkehrsinfrastruktur</i>		
AD Autobahndichte	Länge der Bundesautobahnen in km pro 100 qkm Fläche (1986)	Hess. Kreiszahlen 1986
SD Straßendichte	Länge der Straßen des überörtlichen Verkehrs in km pro 100 qkm Fläche (1986)	Hess. Kreiszahlen 1986

Tabelle 2: Korrelationsmatrix $R_{9,9}$ Correlation matrix $R_{9,9}$

Variable	BQ	BWS	LUG	AD	SD	ABQ	AKQ	ALQ	ELQ
BQ	1,0000	0,9063**	0,6022**	0,6441**	0,5999**	0,5334*	0,4114	0,2630	0,5984**
BWS	0,9063**	1,0000	0,8103**	0,8386**	0,7050**	0,6145**	0,5267*	0,0965	0,4458
LUG	0,6022**	0,8103**	1,0000	0,8130**	0,7245**	0,7420**	0,7461**	-0,0777	0,2116
AD	0,6441**	0,8386**	0,8130**	1,0000	0,7527**	0,4822*	0,4257	0,0362	0,3456
SD	0,5999**	0,7050**	0,7245**	0,7527**	1,0000	0,4772*	0,4578*	0,2598	0,5789**
ABQ	0,5334*	0,6145**	0,7420**	0,4822*	0,4772*	1,0000	0,9649**	-0,1051	0,1373
AKQ	0,4114	0,5267*	0,7461**	0,4257	0,4578*	0,9649**	1,0000	-0,2237	0,0135
ALQ	0,2630	0,0965	-0,0777	0,0362	0,2598	-0,1051	-0,2237	1,0000	0,7881**
ELQ	0,5984**	0,4458	0,2116	0,3456	0,5789**	0,1373	0,0135	0,7881**	1,0000

Anzahl der Kreise: $n = 26$

Signifikanzniveau (einseitiger Test): * = 0,01 und ** = 0,001

- sind die Bruttowertschöpfung und die Autobahndichte.
- Die Hauptkomponente HK 2 korreliert besonders hoch mit den Variablen Abiturientenquote (ABQ) und Akademikerquote (AKQ) und repräsentiert somit das *Qualifikationsniveau* der Kreisbevölkerung.
 - Die Hauptkomponente HK 3 läßt sich aufgrund der hohen Korrelationen mit den Variablen Arbeitslosenquote (ALQ) und Erwerbslosen-

Tabelle 3: Eigenwerte und Varianzanteile für drei Hauptkomponenten/Faktoren einer Hauptkomponenten-/ML-Faktorenanalyse vor Rotation

Eigenvalues and variance shares of three non-rotated principal components/factors of principal component/ML factor analysis

Hauptkomponente/ Faktor	Hauptkomponentenanalyse			ML-Faktorenanalyse		
	Eigenwert	Varianzanteil (%)	kum. Varianzanteil (%)	Eigenwert	Varianzanteil (%)	kum. Varianzanteil (%)
1	5,16940	57,4	57,4	4,97092	55,2	55,2
2	2,06787	23,0	80,4	1,98184	22,0	77,3
3	0,79297	8,8	89,2	0,70462	7,8	85,1

quote (ELQ) als Indikator der *Arbeitslosigkeit* umschreiben.

In der Regel begnügen sich geographische Anwendungen von faktorenanalytischen Verfahren nicht damit, die Anzahl der relevanten Merkmalsdimensionen zu identifizieren und diese inhaltlich zu benennen, sondern versuchen darüber hinaus in einem Rückschluß wieder den Objektbezug herzustellen (vgl. BAHRENBERG u. GIESE 1975a, GIESE 1978). Dementsprechend wurden die kreisspezifischen Hauptkomponenten- bzw. Faktorenwerte für jede Merkmalsdimension berechnet und kartographisch dargestellt (vgl. Abb. 2):

1. Die Werte der Hauptkomponente HK 1 (materielle Wirtschaftskraft) zeigen starke Disparitäten zwischen den Wirtschaftszentren von Hessen und den übrigen Teilräumen. Insgesamt deutet die Hauptkomponente auf Zentrum-Umland-Gefälle hin.
2. Die Hauptkomponente HK 2 als Indikator des Qualifikationsniveaus der Bevölkerung verzeich-

net die höchsten Werte an Hochschulstandorten und in den Umlandkreisen der Wirtschaftszentren des Rhein-Main-Gebiets mit bevorzugten Wohnstandorten hochqualifizierter, wohlhabender Bevölkerungsschichten.

3. Die Hauptkomponente HK 3 zeigt ein ausgeprägtes Nord-Süd-Gefälle.

Als Fazit der bisherigen Untersuchung läßt sich festhalten, daß eine regionalwirtschaftliche Strukturanalyse von Hessen anhand der vorgegebenen Variablen nicht eindimensional vorgenommen werden darf, sondern auf mehreren Merkmalsebenen stattfinden muß. Die drei extrahierten Hauptkomponenten *materielle Wirtschaftskraft*, *Qualifikationsniveau* und *Arbeitslosigkeit* führen zu unterschiedlichen räumlichen Mustern, von denen nur ein einziges (das der Hauptkomponente *Arbeitslosigkeit*) mit der eingangs aufgestellten Hypothese eines Süd-Nord-Gefälles in Hessen korrespondiert. Die Hypothese eines einheitlichen Süd-Nord-Gefälles der hessischen Wirtschaftsstruktur auf Kreisbasis kann nach den Er-

Tabelle 4: Ladungsmatrizen $L_{9,3}$ für drei Hauptkomponenten/Faktoren einer Hauptkomponenten-(HK)/ML-Faktorenanalyse (ML) nach Varimax-Rotation

Matrix of factor loadings $L_{9,3}$ of three varimax-rotated principal components/factors of principal component/ML factor analysis

Variable	HK 1	HK 2	HK 3	ML 1	ML 2	ML 3
BQ	0,72486	0,29077	0,38633	0,83473	0,23722	0,33952
BWS	0,88275	0,32954	0,16185	0,94168	0,30471	0,13926
LUG	0,76336	0,55160	-0,06097	0,67391	0,58856	-0,02718
AD	0,93649	0,16142	0,01549	0,80337	0,23119	0,08290
SD	0,75677	0,26656	0,32780	0,57095	0,36550	0,40395
ABQ	0,33486	0,93086	0,01167	0,36046	0,90674	-0,00996
AKQ	0,28339	0,94583	-0,11288	0,26595	0,95711	-0,11051
ALQ	-0,00734	-0,12138	0,94626	0,02663	-0,14484	0,82968
ELQ	0,35642	0,00911	0,89891	0,32281	0,03352	0,94534

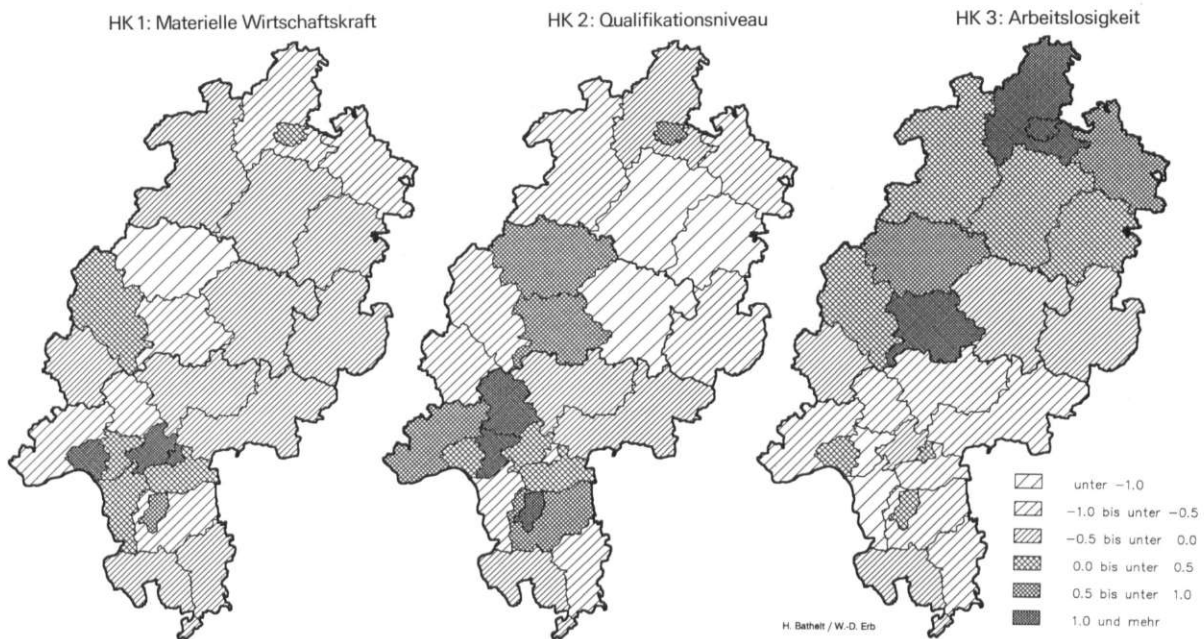


Abb. 2: Räumliche Darstellung der Hauptkomponentenwerte
Spatial distribution of principal component scores

gebnissen der durchgeführten Hauptkomponentenanalyse also nicht aufrecht erhalten werden.

4 Interpretationsprobleme und deren Ursachen

Soweit die typische Vorgehensweise einer regionalwirtschaftlichen Anwendung faktorenanalytischer Verfahren! Tatsächlich zeigt sich auf der Basis der Hauptkomponenten- bzw. Faktorenanalyse eine zumindest auf den ersten Blick plausible wirtschaftsstrukturelle räumliche Differenzierung von Hessen. Unternimmt man jedoch eine genauere Analyse der Hauptkomponentenwerte, die über die Interpretation der berechneten Maximal- und Minimalwerte hinausgeht, so stößt man auf zahlreiche unverständliche Wertekonstellationen. Bei einem Vergleich ausgewählter Kreise lassen sich für jede Merkmalsdimension Unstimmigkeiten in den Größenverhältnissen von Hauptkomponentenwerten finden, die anhand der Ausgangsvariablen nicht nachvollziehbar sind, so zum Beispiel (vgl. Abb. 2):

1. Es ist nicht plausibel, warum der Kreis Gießen anhand der Werte der Hauptkomponente „materielle Wirtschaftskraft“ deutlich schlechter abschneidet als der Lahn-Dill-Kreis. Mit Ausnahme der Lohn- und Gehaltssumme besitzt Gießen bei allen Ausgangsvariablen höhere Ausprägungen als der Lahn-Dill-Kreis, müsste also im Vergleich

zum Lahn-Dill-Kreis einen größeren Hauptkomponentenwert erhalten.

2. Unverständlich bleibt auch, warum die Kreise Hochtaunus und Odenwald durch eine ungünstigere Arbeitsmarktsituation (Hauptkomponente HK 3) gekennzeichnet sein sollen als Groß-Gerau, obwohl die beiden hoch ladenden Variablen Arbeitslosen- und Erwerbslosenquote in den Kreisen Hochtaunus und Odenwald geringere Werte aufweisen als in Groß-Gerau.

Analog dazu lassen sich in den Ergebnissen der ML-Faktorenanalyse fast beliebig viele unplausible Größenrelationen der kreisspezifischen Faktorenwerte feststellen. Entsprechende Probleme tauchen ebenso in den Lösungen für zwei und für vier extrahierte Merkmalsdimensionen auf. Das bedeutet: Bei den ermittelten Unstimmigkeiten handelt es sich weder um ein spezifisches Problem der Hauptkomponentenanalyse noch um ein von einer speziellen Lösung abhängiges Phänomen. Um das enorme Ausmaß der scheinbar inkonsistenten Wertezuordnungen zu verdeutlichen, beschränken sich die folgenden Ausführungen auf die Bewertung der materiellen Wirtschaftskraft (HK 1) des Kreises Gießen im Vergleich zu allen anderen hessischen Kreisen.

Nach den Ergebnissen der Hauptkomponentenanalyse mit drei extrahierten Merkmalsdimensionen korreliert die Hauptkomponente HK 1 hoch positiv mit den Variablen Beschäftigtenquote, Bruttowert-

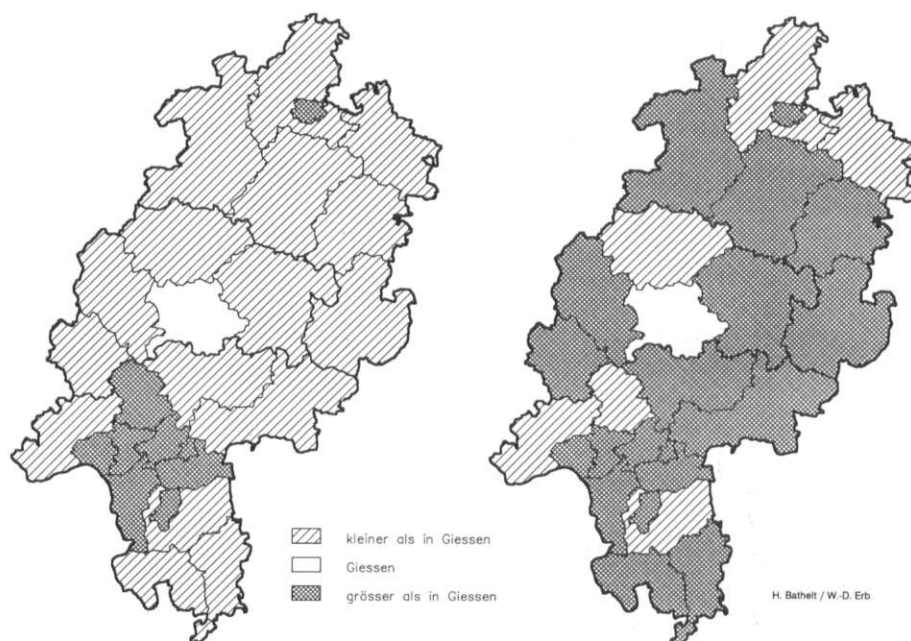


Abb. 3: Räumliche Darstellung der Bruttowertschöpfung BWS (links) und der Hauptkomponente HK 1 „materielle Wirtschaftskraft“ (rechts) mit dem Landkreis Giessen als Bezugskreis

Spatial distribution of indicator scores of economic wellbeing (BWS and HK 1 respectively) – Giessen related to other Hessian administrative districts

schöpfung, Lohn- und Gehaltssumme, Autobahndichte und Straßendichte (vgl. Tab. 4). Anhand dieser Variablen läßt sich leicht eine Klassifizierung in Kreise mit höheren Werten als Giessen und Kreise mit geringeren Werten als Giessen vornehmen. Die resultierenden Gruppen sind in Abb. 3 exemplarisch für die Bruttowertschöpfung dargestellt. Bei allen anderen Variablen, die mit der Hauptkomponente HK 1 hoch korrelieren, ergibt sich eine fast identische Verteilung. In Abb. 3 wird deutlich, daß der Kreis Giessen zu denjenigen hessischen Kreisen mit vergleichsweise hoher materieller Wirtschaftskraft zählt.

Wählt man für die Hauptkomponente HK 1, die nach der üblichen Interpretation einen Indikator der materiellen Wirtschaftskraft darstellt, eine analoge Darstellung wie für die Bruttowertschöpfung, so müßte ein ähnliches Verteilungsmuster resultieren. Die Werte der Hauptkomponente HK 1 liefern jedoch ein stark verzerrtes, ja geradezu inverses Bild über die materielle Wirtschaftskraft von Giessen im Vergleich zu den anderen hessischen Kreisen. Die wirtschaftliche Stärke des Kreises Giessen wird anhand der Ergebnisse der Hauptkomponentenanalyse völlig unterbewertet, obwohl die Ausgangsvariablen ein eindeutig anderes Verteilungsmuster aufweisen (vgl. Abb. 3).

Tab. 5 zeigt analog dazu für ausgewählte hessische Kreise die Werte der Hauptkomponente HK 1 und die standardisierten Werte derjenigen fünf Ausgangsvariablen, die mit der Hauptkomponente HK 1 hoch korrelieren. Aufgrund dieser Gegenüberstellung lassen sich insgesamt elf hessische Kreise identifizieren, die anhand aller fünf hoch ladenden Ausgangsvariablen zwar eine geringere (höchstens gleich große) materielle Wirtschaftskraft besitzen, aber trotzdem einen höheren Hauptkomponentenwert erhalten (vgl. auch Abb. 3). Obwohl die Größenrelationen dieser Kreise zu Giessen durch die Ausgangsvariablen eindeutig bestimmt sind und außerdem die Hauptkomponente HK 1 keine bedeutsamen negativen Ladungen besitzt (vgl. Tab. 4), die eventuell einen Kompensationseffekt hervorrufen könnten, liefert die Hauptkomponente HK 1 völlig entgegengesetzte, der tatsächlichen materiellen Wirtschaftskraft widersprechende Werte.

GIESE (1978) weist in einem kritischen Beitrag über die Anwendung von Faktorenanalysen auf das gleiche Problem hin. Er führt die festgestellten Inkonsistenzen überwiegend auf verletzte Anwendungsvoraussetzungen (speziell Normalitätsbedingungen) zurück und deutet auf die Problematik des Auftretens von Ausreißern hin. Diese Einschränkungen treffen zwar

Tabelle 5: Standardisierte Werte der Hauptkomponente HK 1 („materielle Wirtschaftskraft“) und ausgewählter Variablen (Z-Werte), die HK 1 hoch laden – Landkreis Gießen im Vergleich zu anderen Kreisen

Standardized scores of HK 1 and related high-loading variables – Giessen related to other Hessian administrative districts

Stadt-/Landkreis	Z-BQ (1)	Z-BWS (2)	Z-LUG (3)	Z-AD (4)	Z-SD (5)	Durchschnitt (1) bis (5)	HK 1 (7)
Giessen	0,10088	-0,16591	-0,28635	0,29671	0,07058	0,00318	-0,54495
Lahn-Dill	0,09092	-0,22170	-0,24015	-0,30033	-0,30032	-0,19432	0,03934
Main-Kinzig	-0,31083	-0,27266	0,06113	-0,36318	-0,62730	-0,30257	-0,11470
Limburg-Weilburg	-0,38719	-0,65664	-0,79449	-0,47316	0,63181	-0,33593	-0,16307
Fulda	0,02784	-0,35112	-0,82114	-0,53601	-0,01239	-0,33856	-0,03544
Hersfeld-Rotenburg	-0,19130	-0,37632	-0,42189	0,04532	-0,89083	-0,36700	-0,03296
Bergstraße	-0,63953	-0,59218	-0,65337	-0,14322	-0,54922	-0,51550	-0,49342
Wetterau	-0,82214	-0,63491	-0,45996	-0,20606	-0,49065	-0,52274	-0,43957
Schwalm-Eder	-0,65281	-0,64066	-1,04958	-0,47316	-0,11975	-0,58719	-0,34826
Waldeck-Frankenberg	0,02784	-0,33898	-0,90110	-0,97593	-0,97380	-0,63239	-0,48007
Odenwald	-0,28426	-0,50530	-0,80693	-1,08591	-0,87131	-0,71074	-0,40355
Vogelsberg	-0,51668	-0,63522	-1,15339	-0,64599	-0,76883	-0,74402	-0,35180

auch im vorliegenden Fall zu, reichen allerdings für eine zufriedenstellende Erklärung nicht aus.⁶⁾ Unseres Erachtens sind die scheinbaren Inkonsistenzen vielmehr direkt mit der Rotation der Hintergrundvariablen (Faktoren bzw. Hauptkomponenten) verbunden und sehr viel einfacher, als man zunächst annehmen könnte. Um dies zu verdeutlichen, soll im folgenden der Einfluß der Rotation auf die Berechnung der Hauptkomponenten- bzw. Faktorenwerte untersucht werden. Die Ermittlung der objektspezifischen Werte der Hintergrundvariablen erfolgt durch Multiplikation der standardisierten Datenmatrix mit einer Koeffizientenmatrix: $F = Z C$. Bei einer Hauptkomponentenanalyse läßt sich die Matrix der Factor Score Coefficients analog zu Gleichung [5] durch algebraische Umformung exakt berechnen als $C = L (L'L)^{-1}$. Die Matrix der Factor Score Coefficients resultiert also aus einer Multiplikation der Ladungsmatrix mit der Inversen der Produktmatrix $L'L$. In welchem Umfang die zur Interpretation der Hauptkomponenten verwendete Ladungsmatrix L in die Berechnung der Hauptkomponentenwerte eingeht, hängt ganz entscheidend von der inversen Produktmatrix $(L'L)^{-1}$ ab.

1. Im Fall einer *unrotierten Lösung* nach der Hauptachsenmethode repräsentieren die Ladungen einer Hauptkomponente Eigenvektoren der Korrelationsmatrix R . Die Länge der Eigenvektoren

entspricht den zugehörigen Eigenwerten. Außerdem stehen die Eigenvektoren definitionsgemäß paarweise senkrecht zueinander. Die Produktmatrix $L'L$ sowie ihre Inverse $(L'L)^{-1}$ besitzen somit die Form einer Diagonalmatrix, d. h. außer den Hauptdiagonalelementen haben alle anderen Matrixelemente den Wert Null (vgl. ÜBERLA 1971, S. 238). Insgesamt ist die Koeffizientenmatrix C einer unrotierten Lösung nach der Hauptachsenmethode bis auf hauptkomponentenspezifische Skalierungsfaktoren (den zu den Spalten zugehörigen Eigenwerten) mit der Ladungsmatrix L identisch. Die Ladungen geben sowohl an, in welchem Maß die einzelnen Hauptkomponenten die Verteilungen der Ausgangsvariablen erklären (ursprüngliche Fragestellung nach Gleichung [2]), als auch das Gewicht, mit dem die Ausgangsvariablen in die Berechnung der Hauptkomponentenwerte eingehen (Fragestellung bei der Benennung der Hauptkomponenten nach Gleichung [5]). In diesem Fall ist die Interpretation der Hauptkomponenten anhand der höchsten Ladungen völlig korrekt.

2. Das Bild ändert sich, wenn man eine *orthogonale Rotation* der Hauptkomponenten durchführt. Die resultierenden Ladungsvektoren stellen nach der Rotation keine Eigenvektoren mehr dar und stehen nicht mehr paarweise senkrecht zueinander. Weder die Produktmatrix $L'L$ noch ihre Inverse $(L'L)^{-1}$ haben demnach die Form einer Diagonalmatrix (vgl. HARMAN 1967, S. 348f.). Das wiederum hat zur Folge, daß die Matrix der Factor Score Coefficients spaltenweise nicht mehr pro-

⁶⁾ Auch nach Rangskalierung der Ausgangsvariablen (d. h. dem Ausschalten sog. Ausreißer) traten ähnliche Unstimmigkeiten in den Hauptkomponentenwerten auf.

Tabelle 6: Matrix der Factor Score Coefficients $C_{9,3}$ einer Hauptkomponenten-(HK)/ML-Faktorenanalyse (ML)

Matrix of factor score coefficients $C_{9,3}$ of principal component ML factor analysis

Variable	HK 1	HK 2	HK 3	ML 1	ML 2	ML 3
Z-BQ	0,18064	-0,01993	0,10970	0,00725	-0,00223	-0,00050
Z-BWS	0,30919	-0,10303	-0,06088	1,23664	-0,39311	-0,40842
Z-LUG	0,20185	0,07537	-0,12430	0,00314	0,00111	-0,00149
Z-AD	0,45003	-0,28333	-0,19718	0,00378	-0,00135	-0,00141
Z-SD	0,22398	-0,06375	0,06091	0,00084	0,00051	0,00080
Z-ABQ	-0,22249	0,56639	0,10302	-0,00525	0,02136	0,00281
Z-AKQ	-0,22621	0,57612	0,04246	-0,38492	1,14152	0,08917
Z-ALQ	-0,19160	0,09453	0,55987	-0,00165	0,00053	0,00382
Z-ELQ	-0,03287	0,02584	0,46366	-0,22813	0,19134	1,12347

portional zur Ladungsmatrix ist. Die einzelnen Ladungen entsprechen immer noch den Korrelationskoeffizienten zwischen den Hauptkomponenten und den Ausgangsvariablen, geben allerdings nicht mehr das Gewicht an, mit dem die Ausgangsvariablen in die Berechnung der Hauptkomponentenwerte eingehen. Die vom Betrag her höchsten Ladungen deuten zwar immer noch auf diejenigen Ausgangsvariablen hin, die im Durchschnitt am besten durch die einzelnen Hauptkomponenten repräsentiert werden, geben jedoch keine Auskunft darüber, wie die Hauptkomponentenwerte aus den Ausgangsvariablen hervorgehen. Um beide Aspekte zu erfassen, ist es unseres Erachtens entgegen der gängigen Lehrbuchmeinung notwendig, außer der Ladungsmatrix L auch die Matrix der Factor Score Coefficients C bei der inhaltlichen Deutung der Hauptkomponenten zu verwenden.

Zur Verdeutlichung der angesprochenen Zusammenhänge sind in Tab. 6 die Koeffizientenmatrizen aus der Hauptkomponenten- und der ML-Faktorenanalyse für das Beispiel der regionalwirtschaftlichen Strukturanalyse von Hessen dargestellt. Die Verteilungen der Factor Score Coefficients zeigen, daß infolge der Rotation nicht nur die hoch ladenden Variablen einer bestimmten Hauptkomponente bzw. eines Faktors in die Berechnung der zugehörigen kreisspezifischen Werte eingehen, sondern auch andere Variablen, deren Ladungen auf dieser Hintergrundvariable nahe Null sind. Das Phänomen tritt sowohl im Fall der Hauptkomponenten- als auch im Fall der ML-Faktorenanalyse bei jeder Merkmalsdimension auf und ist Ausdruck eines Kompensationseffekts. So gehen in die Berechnung der Hauptkomponentenwerte von HK 1 wie erwartet die hoch ladenden Variablen Beschäftigtenquote, Bruttowert-

schöpfung, Lohn- und Gehaltssumme, Autobahndichte und Straßendichte in standardisierter Form mit einem Gewicht von rund 0.20 (oder mehr) ein. Zugleich wirken aber auch die standardisierten Variablen Abiturientenquote, Akademikerquote und Arbeitslosenquote, die auf den anderen Hauptkomponenten HK 2 bzw. HK 3 hoch laden, mit etwa gleichem Gewicht, aber umgekehrtem Vorzeichen auf die Hauptkomponentenwerte von HK 1 (vgl. Tab. 6). Bei der Ermittlung der Hauptkomponentenwerte findet also eine Art Kompensation zwischen verschiedenen Ausgangsvariablen statt. Solche Variablen, die mit einer bestimmten Hauptkomponente hoch korrelieren, werden quasi in die Hauptkomponentenwerte „hineingerechnet“, andere Variablen, die mit den anderen Hauptkomponenten hoch korrelieren, werden entsprechend „herausgerechnet“. Dieser Zusammenhang zeigt sich noch deutlicher anhand der Factor Score Coefficients der ML-Faktorenanalyse (vgl. Tab. 6).

Geometrisch interpretiert entspricht eine orthogonale Rotation nichts anderem als einer Drehung des rechtwinkligen Hauptkomponenten-bzw. Faktorenrums. Die Tatsache, daß durch diese Drehung algebraisch Kompensationseffekte zwischen Variablen auftreten, ist nicht ungewöhnlich, sondern direkt auf das Rotationsziel zurückzuführen. Nach Abschluß einer Hauptkomponentenanalyse mit Rotation soll jede Ausgangsvariable möglichst mit einer einzelnen Hauptkomponente hoch korrelieren, die Ladungen auf allen anderen Hauptkomponenten sollen möglichst nahe Null sein. Auf den definitionsgemäß unkorrelierten Hauptkomponenten sollen also völlig verschiedene Variablengruppen hoch laden, die in der Ausgangssituation durchaus miteinander korrelieren können. Damit die extrahierten Hauptkomponenten tatsächlich dieser Bedingung genügen, muß

zwangsläufig eine rechnerische Kompensation stattfinden.

Die auftretenden Kompensationsvorgänge können allerdings nur aufgedeckt werden, wenn außer der Ladungsmatrix auch die Matrix der Factor Score Coefficients bei der Benennung und Interpretation der Hauptkomponenten verwendet wird. Die Hauptkomponente HK 1 ist demnach kein „reiner“ Wirtschaftskraftindikator, wie aufgrund der hoch ladenden Variablen zunächst vermutet werden kann. Korrekterweise müßte man sagen: HK 1 repräsentiert die *materielle Wirtschaftskraft* hessischer Kreise *bereinigt um Arbeitslosigkeits- und Qualifikationseffekte*. Bei einer solchen Interpretation wird auch klar, warum Gießen im Vergleich zu vielen anderen hessischen Kreisen durch die Hauptkomponentenwerte von HK 1 so stark unterbewertet wird (vgl. Abb. 3 und Tab. 5). Der Kreis Gießen besitzt überdurchschnittlich hohe Akademiker-, Abiturienten-, Arbeitslosen- und Erwerbslosenquoten. Dadurch ist der Kompensationseffekt auf der mit den Wirtschaftskraftvariablen korrelierenden Hauptkomponente HK 1 für Gießen größer als für andere hessische Kreise.

5 *Schlußfolgerungen*

Die vorliegende Untersuchung knüpft an die kritischen Beiträge von BAHRENBURG u. GIESE (1975b) und GIESE (1978) zur Anwendung faktorenanalytischer Verfahren in der Geographie an, versucht allerdings gegenüber dem Problem nicht erfüllter Anwendungsvoraussetzungen (z. B. fehlende Normalität oder das Auftreten von Ausreißern) inhaltlich-interpretatorische Aspekte stärker hervorzuheben. Den Ausgangspunkt der Überlegungen bildeten scheinbare Inkonsistenzen in den Verteilungen der Hauptkomponentenwerte im Rahmen einer regionalwirtschaftlichen Strukturanalyse der hessischen Stadt- und Landkreise. Bestimmte Kreise erhielten durch die Hauptkomponentenanalyse Wertezuordnungen, die in ihren Größenverhältnissen zueinander unverständlich waren, da anhand der Ausgangsvariablen in Verbindung mit den Hauptkomponentenladungen ein umgekehrtes Bild zu erwarten war. Entsprechende Interpretationsprobleme tauchten bei jeder extrahierten Hauptkomponente auf. Hätte man die Ergebnisse der Hauptkomponentenanalyse unreflektiert so interpretiert und weiterbehandelt, wie es in der gängigen Fachliteratur empfohlen wird (angeprochen ist insbesondere die inhaltliche Füllung der Hauptkomponenten), dann hätte sich für Teilergebnisse der Analyse eine Konstellation ergeben,

die wegen der Eindeutigkeit der Struktur der Ausgangsvariablen nicht erklärbar gewesen wäre. Zum Beispiel hätte man die materielle Wirtschaftskraft des Kreises Gießen in Relation zu den Umlandkreisen deutlich zu gering eingeschätzt. Selbstverständlich sind die extrahierten Hintergrundvariablen einer Hauptkomponenten- oder Faktorenanalyse nicht mit den Ausgangsvariablen identisch, sondern stellen rein hypothetische Konstrukte dar. Eine ausgleichende bzw. kompensierende Auswirkung ist deshalb zu erwarten und sogar sinnvoll. Als Grundlage für konkrete planerische Maßnahmen oder das Entwerfen einer regionalen Entwicklungsstrategie wären die Resultate der durchgeführten Hauptkomponentenanalyse nicht tauglich gewesen.

Mehrere Gründe sprechen dafür, daß die scheinbaren Inkonsistenzen nicht auf eine ungünstige Ausgangssituation zurückzuführen sind und deshalb auch für andere Fachwissenschaften wie Soziologie oder Psychologie, in denen faktorenanalytische Verfahren zum methodischen Standardrepertoire gehören (vgl. ARMINGER 1979, ÜBERLA 1971), Relevanz besitzen:

1. Der zugrundeliegende Datensatz wurde aus inhaltlichen Erwägungen abgeleitet und nicht speziell konstruiert.
 2. Das dargestellte Problem trat sowohl bei einer Hauptkomponenten- als auch bei einer ML-Faktorenanalyse auf, ist also nicht ausschließlich verfahrensspezifisch.
 3. Weiterhin sind die scheinbar inkonsistenten Wertezuordnungen nicht von einer speziellen Lösung abhängig, wie man leicht vermuten könnte. Vergleichbare Schwierigkeiten traten jeweils bei den Lösungen für zwei, für drei und für vier Merkmalsdimensionen auf.
 4. Ebenso ist das Auftreten von Ausreißern (die z. B. aus der formalen Gleichbehandlung von Stadtkreisen und Landkreisen resultieren könnten) als Hauptfehlerquelle auszuschalten, denn die geschilderten Interpretationsprobleme traten auch bei Rechenvarianten mit rangskalierten Variablen auf.
- Scheinbar inkonsistente Hauptkomponenten- bzw. Faktorenwerte sind bei genauerem Durchdenken der einzelnen Verfahrensschritte einer Hauptkomponenten- bzw. Faktorenanalyse keineswegs völlig überraschend, sondern resultieren direkt aus der Rotation. Durch die Rotation findet auf den einzelnen Hauptkomponenten bzw. Faktoren unter Wahrung der Orthogonalität quasi eine Kompensation von Variablenwerten statt, so daß ursprünglich mittelstark miteinander korrelierende Variablen (in Tab. 2 z. B. die Variablen Bruttowertschöpfung und Abitu-

rientenquote) nach der Rotation mit unterschiedlichen Hauptkomponenten bzw. Faktoren stark und mit jeweils allen anderen Hintergrundvariablen schwach korrelieren. Die dabei auftretenden Kompensationen sind anhand der Ladungsmatrizen im Nachhinein nicht mehr zu erkennen, sondern werden erst durch Hinzuziehen der Factor Score Coefficients deutlich. Bei der Benennung und Interpretation der extrahierten Merkmalsdimensionen wird dieser Kompensationseffekt jedoch nicht mehr berücksichtigt. Im Sinn einer problemorientierten und praxisrelevanten Anwendung von faktorenanalytischen Verfahren können deshalb folgende Empfehlungen (schwerpunktmäßig auf die Hauptkomponentenanalyse ausgerichtet) gegeben werden:

1. Große Vorsicht ist bei Untersuchungen mit einer Vielzahl von Objekten geboten. In einer solchen Situation ist es nicht mehr möglich, durch Einzelfallprüfungen alle scheinbaren Unstimmigkeiten in der Struktur der Hauptkomponenten- bzw. Faktorenwerte zu identifizieren. Für die Interpretation der Ergebnisse ist deshalb äußerste Sorgfalt ratsam. Dies gilt auch für die Auswahl der einzubeziehenden Variablen. Faktorenanalytische Verfahren sollten nicht als Ersatz für eine sorgfältige, theoriegeleitete und dem Untersuchungsgegenstand angemessene Variablenauswahl herangezogen werden.

2. Relativ unproblematisch erscheint der Einsatz einer Hauptkomponentenanalyse ohne anschließende Rotation. Die Hauptkomponentenladungen beschreiben in diesem Fall (abgesehen von einem Skalierungsfaktor) sowohl den Einfluß der Hauptkomponenten auf die Variablen (ursprüngliche Fragestellung der Hauptkomponentenanalyse) als auch den Einfluß der Variablen auf die Hauptkomponenten (Fragestellung bei der Interpretation der Hauptkomponenten und der Berechnung der Hauptkomponentenwerte). D. h., die Hauptkomponentenladungen geben exakt wieder, mit welchem Gewicht die verschiedenen Variablen in die Berechnung der Hauptkomponentenwerte eingehen. Die Hauptkomponentenanalyse ohne Rotation bildet damit ein wichtiges Verfahren, um voneinander unabhängige Variablen zu erzeugen, die anschließend in einer Clusteranalyse, Diskriminanzanalyse oder multiplen Regressionsanalyse weiterverwendbar sind. Oft wird man dabei eine vollständige Lösung (Extraktion aller Hauptkomponenten: $r = m$) bevorzugen. Diese Vorgehensweise besitzt jedoch in vielen Anwendungen den Nachteil, daß die erzeugten Hintergrundvariablen wegen der Konstellation der Hauptkomponentenladungen nicht sinnvoll interpretiert bzw. inhaltlich gegeneinander abgegrenzt werden können.

3. Sofern eine Hauptkomponentenanalyse mit anschließender orthogonaler Rotation durchgeführt wird, reicht der in den Standardlehrbüchern vorgeschlagene Weg einer Benennung der Hauptkomponenten anhand der höchsten Ladungen nicht aus. Die Ladungen allein geben keinen Aufschluß darüber, wie stark die einzelnen Variablen in die Berechnung der Hauptkomponentenwerte eingehen, sondern sind Ausdruck der nach der Rotation (mit entsprechenden Kompensationseffekten) vorliegenden Korrelationen zwischen Hauptkomponenten und Ausgangsvariablen. Eine fehlerfreie Behandlung der extrahierten Hauptkomponenten ist nur unter Berücksichtigung der Factor Score Coefficients möglich. Folgende Optionen scheinen unseres Erachtens diesbezüglich vertretbar:

a) Falls die geschilderten Probleme im Rahmen einer Hauptkomponentenanalyse nicht auftreten, können die Ergebnisse in der üblichen Weise interpretiert und weiterverarbeitet werden. Eine solche Konstellation ist vor allem dann zu erwarten, wenn die rotierte und die nicht-rotierte Lösung einander ähnlich sind, die Ladungen also angeben, wie stark die einzelnen Variablen in die Berechnung der Hauptkomponentenwerte eingehen.

b) Falls wie im Beispiel der Strukturanalyse von Hessen scheinbar inkonsistente Hauptkomponentenwerte auftreten, ist die Hauptkomponentenanalyse mit anschließender Rotation trotzdem ein sinnvolles Verfahren, um die Anzahl der notwendigen Merkmalsdimensionen und die mit den Merkmalsdimensionen hoch korrelierenden Variablen zu identifizieren. Sofern zusätzlich ein Rückschluß von den Hauptkomponenten auf die Einzelobjekte angestrebt wird, bietet es sich an, Stellvertretervariablen für jede Hauptkomponente zu verwenden. Die Auswahl der Stellvertretervariablen sollte sich an den Hauptkomponentenladungen, an den a posteriori-Kommunalitäten, an den Factor Score Coefficients und an sachinhaltlichen Erwägungen orientieren. Anstatt einer einzelnen kann unter Umständen eine Kombination von mehreren Stellvertretervariablen je Hauptkomponente (z. B. das arithmetische Mittel mehrerer standardisierter Ausgangsvariablen) sinnvoll sein (vgl. GIESE 1978, 1985). Die Methode der Stellvertretervariablen (d. h. das Festhalten an den Ausgangsvariablen) empfiehlt sich auch dann, wenn die extrahierten Dimensionen einer Hauptkomponentenanalyse als erklärende Variablen in eine nachgeschaltete multiple Regressionsanalyse eingehen sollen. Dabei ist jedoch zu beachten, daß die Stellvertretervariablen anders als die Hauptkomponenten keine voneinander unabhängigen Größen darstellen,

das Problem der Multikollinearität damit also noch nicht bewältigt ist.

c) Anstatt der Verwendung von Stellvertretervariablen ist von vornherein auch eine schiefwinklige Rotation der extrahierten Hauptkomponenten möglich (vgl. dazu ÜBERLA 1971, S. 211ff.; HARMAN 1967, S. 273ff.; KIM u. MUELLER 1978, S. 37ff.). Ob man prinzipiell bereit ist, eine schiefwinklige Rotation zu akzeptieren, hängt in starkem Maß von der subjektiven Einstellung gegenüber einer Hauptkomponentenlösung ab, deren Merkmalsdimensionen nicht mehr (wie ursprünglich angestrebt) voneinander unabhängig sind. In einigen Fällen könnten die angesprochenen Probleme der Hauptkomponentenwerte durch eine schiefwinklige Rotation zwar verringert oder vermieden werden, aus sachinhaltlichen Überlegungen ist die Methode der Stellvertretervariablen jedoch vorzuziehen.

d) Schließlich ist sogar vorstellbar, daß man die Hauptkomponentenwerte nach einer Rotation wie gewohnt objektspezifisch darstellt und weiterverarbeitet. In diesem Fall muß jedoch peinlichst darauf geachtet werden, daß man unter Verwendung der Matrix der Factor Score Coefficients die Komplexität der einzelnen Hauptkomponenten (insbesondere die Kompensationseffekte auf den Hauptkomponenten) stärker hervorhebt, als dies allein aufgrund der Hauptkomponentenladungen möglich wäre.

4. Verwendet man die ermittelten Hauptkomponenten als erklärende Variablen in einem nachgeschalteten Verfahren, so sollte man bei einer Interpretation der endgültigen Ergebnisse (etwa bei der Charakterisierung der Gruppen aus einer Clusterlösung) stets auf die Merkmalsausprägungen der Ausgangsvariablen und nicht auf die Hauptkomponentenwerte zurückgreifen, um eine nicht abschätzbare Verkettung scheinbarer Inkonsistenzen zu vermeiden.

Literatur

- ARMINGER, G.: Faktorenanalyse. Statistik für Soziologen 3. Stuttgart 1979.
- BACKHAUS, K., ERICHSON, B., PLINKE, W., SCHUCHARD-FICHER, C. u. WEIBER, R.: Multivariate Analysemethoden. Eine anwendungsorientierte Einführung. 4. Auflage. Berlin, Heidelberg, New York 1987.
- BAHRENBERG, G. u. GIESE, E.: Statistische Methoden und ihre Anwendung in der Geographie. Teubner Studienbücher. Stuttgart 1975 a.
- : Zum Problem der Normalität und Transformation bei der Faktorenanalyse bzw. Hauptkomponentenanalyse. In: GIESE, E. (Hrsg.): Symposium „Quantitative Geographie“. Gießener Geographische Schriften 32. Gießen 1975 b, S. 9-29.
- BATHELT, H. u. ERB, W.-D.: Anwendungs- und Interpretationsprobleme faktorenanalytischer Verfahren. Eine kritische Untersuchung. In: Studien zur Wirtschaftsgeographie (April 1991). Gießen 1991.
- COOLEY, W. W. u. LOHNES, P. R.: Multivariate Data Analysis. New York, London, Sydney, Toronto 1971.
- DAULTREY, S.: Principal Components Analysis. Catmog 8. Norwich 1976.
- DILLON, W. R. u. GOLDSTEIN, M.: Multivariate Analysis. Methods and Applications. New York 1984.
- FAHRMEIR, L. u. HAMERLE, A. (Hrsg.): Multivariate statistische Verfahren. Berlin, New York 1984.
- GIESE, E.: Kritische Anmerkungen zur Anwendung faktorenanalytischer Verfahren in der Geographie. In: Geographische Zeitschrift 66, 1978, S. 161-182.
- : The Development and Present State of Research into „Quantitative Geography“ in the German-Speaking Countries. In: BAHRENBERG, G. u. STREIT, U. (Hrsg.): German Quantitative Geography. Münstersche Geographische Arbeiten 11. Paderborn 1981, S. 9-26.
- : Klassifikation der Länder der Erde nach ihrem Entwicklungsstand. In: Geographische Rundschau 37, 1985, S. 164-175.
- GIESE, E., BENKE, E. u. TOWARA, M.: Zum Problem der Festlegung des kommunalrechtlichen Status von Städten. Schriften des Zentrums für regionale Entwicklungsforschung der Justus-Liebig-Universität Gießen 24. Saarbrücken 1982.
- GODDARD, J. u. KIRBY, A.: An Introduction to Factor Analysis. Catmog 7. Norwich 1976.
- HARMAN, H. H.: Modern Factor Analysis. 2nd Edition. Chicago, London 1967.
- KEMPER, F.-J.: Die Anwendung faktorenanalytischer Rotationsverfahren in der Geographie des Menschen. In: GIESE, E. (Hrsg.): Symposium „Quantitative Geographie“. Gießener Geographische Schriften 32. Gießen 1975, S. 34-47.
- KEMPER, F.-J. u. SCHMIEDECKEN, W.: Faktorenanalyse zum Klima Mitteleuropas. In: Erdkunde 31, 1977, S. 255-272.
- KILCHENMANN, A.: Untersuchung mit quantitativen Methoden über die fremdenverkehrs- und wirtschaftsgeographische Struktur der Gemeinden im Kanton Graubünden (Schweiz). Zürich 1968.
- KIM, J.-O. u. MUELLER, C. W.: Factor Analysis. Statistical Methods and Practical Issues. Beverly Hills, London 1978.
- LANGE, N. DE: Städtetypisierung in Nordrhein-Westfalen im raum-zeitlichen Vergleich 1961 und 1970 mit Hilfe multivariater Methoden – eine empirische Städtesystemanalyse. Münstersche Geographische Arbeiten 8. Paderborn 1980.
- : Zum Problem interkorrelierter Variablen bei clusteranalytischen Verfahren. In: Berichte des Symposiums des Arbeitskreises: „Theorie und quantitative Methodik in der Geographie“. Klagenfurter Geographische Schriften 2. Klagenfurt 1981, S. 167-184.

- NIPPER, J.: Mobilität der Bevölkerung im Engeren Informationsfeld einer Solitärstadt. Gießener Geographische Schriften 33. Gießen 1975.
- RAO, C. R.: Linear Statistical Inference and its Application. New York 1965.
- RASE, W. D.: Gemeindetypen im Saarland. Analyse räumlicher Strukturen mit multivariaten statistischen Methoden. In: Geographische Rundschau 26, 1974, S. 391-399.
- SCHWARZ, R.: Die Leistung mathematischer Ansätze zur Verarbeitung geographischer Daten. In: Geographische Rundschau 26, 1974, S. 381-390.
- SPSS Inc. (Ed.): SPSS/PC + Advanced Statistics™ V2.0. Chicago 1988.
- STEINER, D.: Die Faktorenanalyse. Ein modernes statistisches Hilfsmittel des Geographen für die objektive Raumgliederung und Typenbildung. In: Geographica Helvetica 26, 1965, S. 20-34.
- : Geographische Raumgliederung und Mustererkennung. In: Publikationen des Geographischen Instituts der ETH-Zürich 55. Zürich 1975, S. 19-45.
- ÜBERLA, K.: Faktorenanalyse. Eine systematische Einführung für Psychologen, Mediziner, Wirtschafts- und Sozialwissenschaftler. 2. Auflage. Berlin, Heidelberg, New York 1971.
- WRIGLEY, N. u. BENNETT, R. J. (Eds.): Quantitative Geography: A British View. London, Boston, Henley 1981.