

Faktorenanalytische Untersuchungen zur Wirtschaftsstruktur der deutschen Großstädte nach der Berufszählung 1907

Blotevogel, Hans Heinrich

Veröffentlichungsversion / Published Version

Sammelwerksbeitrag / collection article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Blotevogel, H. H. (1979). Faktorenanalytische Untersuchungen zur Wirtschaftsstruktur der deutschen Großstädte nach der Berufszählung 1907. In W. H. Schröder (Hrsg.), *Moderne Stadtgeschichte* (S. 74-111). Stuttgart: Klett-Cotta. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-327834>

Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Faktorenanalytische Untersuchungen zur Wirtschaftsstruktur der deutschen Großstädte nach der Berufszählung 1907

Wenn man den wissenschaftsgeschichtlichen Ausbreitungsprozeß der Faktorenanalyse verfolgt, so spricht einiges dafür, daß in absehbarer Zukunft auch die deutschsprachige Stadtgeschichtsforschung diesem mathematisch-statistischen Analyseinstrument wachsende Aufmerksamkeit schenken wird. Entwickelt wurde das Verfahren, das innerhalb der multivariaten Statistik eine zentrale Rolle einnimmt und dessen Anwendung auch im Mittelpunkt dieses Beitrages steht, um die Wende zum 20. Jahrhundert innerhalb der Psychologie, in der es auch bis zu den vierziger Jahren in den Grundfragen weitgehend geklärt und zur standardmäßigen Anwendung weiterentwickelt wurde¹. Der praktischen Anwendung waren allerdings durch den enorm hohen Rechenaufwand enge Grenzen gesetzt, und erst durch die Entwicklung leistungsfähiger Großrechner in den fünfziger Jahren wurde die Verarbeitung größerer Datenmengen sowie die Anwendung anspruchsvollerer Varianten der Faktorenanalyse möglich. In diesen Jahren setzte auch die Diffusion in andere Wissenschaften ein: Die Biologie, die Soziologie, die Wirtschaftswissenschaften, die Medizin, die Pädagogik und recht bald – jedenfalls in den USA und Großbritannien – auch die Geographie melden seit den fünfziger Jahren in laufend steigendem Umfang faktorenanalytische Untersuchungen. Heute zählt die Faktorenanalyse bzw. ihre Varianten wie die Hauptkomponentenanalyse zu den meistgebrauchten Standardverfahren nahezu aller empirischen Wissenschaften, und auch im Kanon mathematisch-statistischer Analyseinstrumente der Geographie bzw. Raumwissenschaften hat sie einen festen Platz erobert².

Indessen pflegt die Ausbreitung von Neuerungen dem Gesetz zu folgen, daß sie nach anfänglich rascher Diffusion allmählich an Anziehungskraft verlieren. In der Psychologie wurde die Faktorenanalyse schon seit den sechziger Jahren zunehmend kritisch beurteilt, und seit neuestem mehrten sich auch in den Sozialwissenschaften skeptische Stimmen³. Wie die Entwicklung weitergeht, ist noch kaum absehbar. Die auch in der Geographie anfänglich etwas euphorischen Erwartungen, die in der Faktorenanalyse *das* universell anwendbare Analyseinstrument sahen, sind inzwischen wohl überall einer kritischen und differenzierteren Einstellung gewichen, aber auch

¹ Überla, Karl, *Faktorenanalyse*, 2nd. ed., Berlin 1971, S. 7 ff.

² King, Leslie, *Statistical Analysis in Geography*, Englewood Cliffs, N.J. 1969. – Yeates, Maurice, *An Introduction to Quantitative Analysis in Human Geography*, New York 1974. – Bahrenberg, Gerhard und Ernst Giese, *Statistische Methoden und ihre Anwendung in der Geographie*, Stuttgart 1975. – Johnston, R.J., *Multivariate Statistical Analysis in Geography*, London 1978.

³ Beispielsweise: Kriz, Jürgen, *Statistik in den Sozialwissenschaften*, Reinbek 1973.

die Skeptiker wie z. B. Schlosser⁴, die die Faktorenanalyse im ganzen für obsolet halten, gehen m. E. zu weit. Es ist allerdings richtig, daß die Faktorenanalyse häufig zu unreflektiert und unkritisch angewandt worden ist. Dabei mag die hohe Attraktivität des Verfahrens neben wissenschaftsgeschichtlichen Gründen paradoxerweise auch in der mathematischen Kompliziertheit, aber auch faszinierenden Logik liegen, auch wenn damit noch keine fachliche Relevanz garantiert ist.

Aus der historischen Städteforschung liegen bisher noch kaum faktorenanalytische Erfahrungen vor, so daß im folgenden ein Beitrag zur Positionsbestimmung der Faktorenanalyse im fachlichen Kontext geleistet werden soll. Dabei wird anhand eines konkreten Problems aus der historisch-geographischen Stadtforschung der Grundansatz der Faktorenanalyse – bzw. exakter: der Hauptkomponentenanalyse – vorgestellt und kritisch beleuchtet. Hierbei wird der üblicherweise begangene Weg der Faktorenanalyse, der von den Korrelationen zwischen den Ausgangsvariablen ausgeht, mit anderen, m. E. bisher zu wenig diskutierten Alternativen verglichen.

Die hier gestellte Aufgabe ist mit erheblichen Darstellungsproblemen verbunden. In der Tat muß man sich für ein volles Verständnis der Faktorenanalyse durch umfangreiche, mit Formeln gespickte Lehrbücher arbeiten, wobei die Vertrautheit mit den Grundregeln der Matrixalgebra noch eines der geringsten Probleme bildet. Aber ein solch tiefes Eindringen in die mathematischen Grundlagen und Hintergründe ist m. E. entbehrlich, sofern nur das Interesse besteht, das Anliegen und die wichtigsten Analyseschritte im Grundansatz zu verstehen, um faktorenanalytische Untersuchungen nachvollziehen und beurteilen zu können. Für diese Ebene des Verständnisses ist eine mathematische Formulierung nicht unbedingt nötig, so daß hier versucht werden soll, die Hauptansätze graphisch, d. h. durch Diagramme, sowie durch verbale Erläuterungen zu verdeutlichen. Dafür müssen Darstellungslücken und Vereinfachungen bewußt in Kauf genommen und Interessenten an einer vollständigen und exakten Darstellung auf die umfangreiche Literatur verwiesen werden⁵.

Die Hauptleistung der Faktorenanalyse besteht in der Datenreduktion, in der Verdichtung von zuvor häufig unübersichtlich großen Datenmengen, um inhärente Strukturen herauszupräparieren. Sie wird zumeist für überwiegend induktive Zwecke eingesetzt, d. h. in Fällen, wo zunächst keine oder nur vage Hypothesen über Zusammenhänge und kausale Verknüpfungen bestehen; ihre Hauptleistung besteht weniger in der Hypothesenprüfung, sondern als Hilfsmittel zur Hypothesenbildung.

In der sozialwissenschaftlichen Stadtforschung lassen sich zwei Hauptanwendungsgebiete unterscheiden:

1. Auf der Mikroebene, auf die hier nicht weiter eingegangen werden kann, wird die Faktorenanalyse hauptsächlich bei der Auswertung von Befragungen eingesetzt⁶.

⁴ Schlosser, Otto, *Einführung in die sozialwissenschaftliche Zusammenhangsanalyse*, Reinbek 1976.

⁵ Harman, H.H., *Modern Factor Analysis*, 2nd. ed., Chicago 1967. – Überla, *Faktorenanalyse*. – Revenstorf, Dirk, *Lehrbuch der Faktorenanalyse*, Stuttgart 1976.

⁶ Dazu beispielsweise: Holm, Kurt, *Die Befragung*, Band 3: *Die Faktorenanalyse*, München 1976.

2. Auf der Meso- und Makroebene, d. h. bei der Analyse von räumlichen Aggregaten wie Stadtteilen, ganzen Städten und Städtegruppen, findet die Faktorenanalyse in erster Linie in der sog. *Sozialraumanalyse*⁷ bzw. der mit ihr eng verwandten sog. *Faktorialökologie*⁸ Verwendung. Ziel dieser Forschungsrichtungen ist die Untersuchung innerstädtischer Strukturen auf allgemeine latente, voneinander unabhängige Grunddimensionen. Teilweise unabhängig, teilweise angeregt hiervon entwickelte sich die faktorenanalytische Untersuchung von *Städtegruppen*. Hier lassen sich wiederum mindestens drei Richtungen unterscheiden: Zum einen versuchte man, das Problem der funktionalen ökonomischen Städteklassifizierung mit Hilfe der Faktorenanalyse besser und objektiver zu lösen⁹; zum zweiten wurde der Ansatz der Faktorialökologie auf Städtesysteme übertragen¹⁰, und schließlich bediente man sich der Faktorenanalyse zur Konstruktion komplexer mehrdimensionaler Indizes, beispielsweise zur Zentralitätsmessung¹¹.

Die meisten dieser Forschungsansätze, die hier nur sehr unvollständig angedeutet werden können, beziehen sich ausschließlich auf Gegenwartsverhältnisse, und erst aus jüngster Zeit liegen einige Studien vor, in denen das Konzept der innerstädtischen Faktorialökologie auf Städte im 19. und beginnenden 20. Jahrhundert angewandt wurde, um die Veränderungen innerstädtischer Grunddimensionen unter dem Einfluß von Industrialisierung bzw. Modernisierung zu erfassen¹². Demgegen-

⁷ Dieser Ansatz der *Social Area Analysis* bildet eine Weiterentwicklung der älteren sozial-ökologischen Tradition der sog. Chicagoer Schule der Soziologie (Park u. a.). — Shevky, Eshref und Wendell Bell, *Social Area Analysis*, Stanford 1956. — Shevky, Eshref und Wendell Bell, *Sozialraumanalyse*, in: Atteslander, Peter und Bernd Hamm (eds.), *Materialien zur Siedlungssoziologie*, Köln 1974, S. 125–139.

⁸ Robson, Brian, *Urban Analysis*, Cambridge 1969. — Rees, Philip H., *Concepts of Social Space: Toward an Urban Social Geography*, in: Berry, Brian und Frank Horton (eds.), *Geographic Perspectives on Urban Systems*, Englewood Cliffs, N. J. 1970, S. 306–394. — Rees, Philip H., *Problems of Classifying Subareas Within Cities*, in: Berry, Brian und Katherine Smith (eds.), *City Classification Handbook*, New York 1972, S. 265–330. — Friedrichs, Jürgen, *Stadtanalyse*, Reinbek 1977. — Hamm, Bernd, *Die Organisation der städtischen Umwelt*, Frauenfeld 1977.

⁹ Price, Daniel O., *Factor Analysis in the Study of Metropolitan Centers*, in: *Social Forces*, 20 (1942), S. 449–455. — Moser, C.A. und Wolf Scott, *British Towns: a Statistical Study of Their Social and Economic Differences*, Edinburgh 1961. — Ahmad, Qazi, *Indian Cities: Characteristics and Correlates*, Chicago 1965. — Schmidt, Gerhard, et al., *Anwendung der Faktorenanalyse bei der Gemeindetypisierung*, in: *Petermanns Mitteilungen*, 118 (1974), S. 189–194. — Schmidt, Gerhard und Otti Margraf, *Die Klassifikation von Zentren mittels der Faktorenanalyse und Dendrogrammen*, in: *Petermanns Mitteilungen* 120 (1976), S. 108–115.

¹⁰ Hadden, Jeffrey und Edgar Borgatta, *American Cities: Their Social Characteristics*, Chicago 1965. — Berry, Brian und Katherine Smith (eds.), *City Classification Handbook: Methods and Applications*, New York 1972.

¹¹ Klemmer, Paul, *Der Metropolisierungsgrad der Stadtregionen*, Hannover 1971. — Hellberg, Hans, *Zentrale Orte als Entwicklungsschwerpunkte in ländlichen Gebieten*, Göttingen 1972. — Köck, Helmut, *Das zentralörtliche System von Rheinland-Pfalz*, Bonn 1975.

¹² Goheen, Peter G., *Victorian Toronto, 1850 to 1900*, Chicago 1970. — Shaw, Mark, *The*

über soll hier an die faktorenanalytische Untersuchung von Städtegruppen, insbesondere an den Ansatz der ökonomischen Funktionstypisierung, angeknüpft werden.

Die Erarbeitung funktionaler Städtetypologien und ihre Verwendung zur Charakterisierung von Städtegruppen hat insbesondere in der Stadtgeographie eine lange Tradition¹³. Dabei lassen sich mindestens zwei unterschiedliche Konzepte unterscheiden: Zum einen werden städtische Funktionstypen als theoretische Konstrukte verwendet, etwa im Sinne von *Idealtypen* nach Max Weber (z. B. *Konsumentenstadt*¹⁴) oder im Sinne von *theoretischen Modellen*¹⁵ (z. B. *Zentraler Ort*). Zum anderen werden städtische Funktionstypen als Klassen innerhalb von empirischen Gliederungssystemen verstanden, deren Klassifikationsmerkmale und Schwellenwerte eindeutig definiert sind. Die erste Variante, deren Zweckmäßigkeit allein von der Qualität und Fruchtbarkeit der dahinter stehenden Theorie abhängt, soll hier im weiteren ausgeklammert werden, da sich solche theoretischen Typologien im allgemeinen einer unmittelbaren empirischen Überprüfung entziehen. Sehr viel häufiger sind dagegen klassifikatorische Typologien¹⁶, die zumeist auf Beschäftigungsdaten basieren und auf die Ausgliederung von Typen wie Bergbau-, Industrie-, Handels-, Verwaltungs- und Fremdenverkehrsstädten zielen. Solche Klassifikations-

Ecology of Social Change: Wolverhampton 1851–1871, in: Institute of British Geographers, Transactions N.S. 2 (1977), S. 332–348. – Schwippe, Heinrich und Clemens Wischermann, *Zur Untersuchung städtischer Raumgefüge und Wohnungsstrukturen im 19. Jahrhundert nach Quellen des Katasters und der Statistik*, Vortrag im Rahmen einer Arbeitstagung im Zentralinstitut für sozialwissenschaftliche Forschung der Freien Universität Berlin am 12.11.1977.

¹³ Beispielhaft seien genannt: Bobek, Hans, *Über einige funktionelle Stadttypen und ihre Beziehungen zum Lande*, in: Comptes rendus du Congrès international de Géographie Amsterdam 1938, Tome II, Sect. IIIa, Leiden 1938, S. 88–102. – Harris, Chauncy, *A Functional Classification of Cities in the United States*, in: Geographical Review 33 (1943), S. 86–99. – Huttenlocher, Friedrich, *Funktionale Siedlungstypen*, in: Berichte der Deutschen Landeskunde, 7 (1950), S. 76–86. – Huttenlocher, Friedrich, *Städtetypen und ihre Gesellschaften anhand südwestdeutscher Beispiele*, in: Geographische Zeitschrift 51 (1963), S. 161–182. – Smith, Robert H.T., *Method and Purpose in Functional Town Classification*, in: Annals of the Association of American Geographers 55 (1965), S. 539–548. – Schwarz, Gabriele, *Allgemeine Siedlungsgeographie*, 3rd ed., Berlin 1966. – Schöller, Peter, *Die deutschen Städte*, Wiesbaden 1967. – Windelband, Ursula, *Typologisierung städtischer Siedlungen*, Gotha 1973.

¹⁴ Weber, Max, *Die Stadt*, in: Archiv f. Sozialwiss. u. Sozialpol., 47 (1921), S. 621 ff.

¹⁵ Hempel, Carl G., *Typologische Methoden in den Sozialwissenschaften* (1952), in: Topitsch, Ernst (ed.), *Logik der Sozialwissenschaften*, Köln 1965, S. 85–103.

¹⁶ Beispielhaft seien genannt: Harris, *Functional Classification*. – Pownall, L.L., *The Functions of New Zealand Towns*, in: Annals of the Association of American Geographers, 43 (1953), S. 332–350. – Steigenga, William, *A Comparative Analysis and Classification of Netherland's Towns*, in: Tijdschrift voor Economische en Sociale Geografie, 46 (1955), S. 105–119. – Duncan, Otis und Albert Reiss, *Social Characteristics of Urban and Rural Communities*, 1950, New York 1956. – Nelson, Howard, J., *A Service Classification of American Cities*, in: Economic Geography, 31 (1955), S. 189–210. – Wilkinson, Thomas O., *A Functional Classification of Japanese Cities, 1920–1955*, in: Demography, 1 (1964), S. 177–185.

systeme sind nicht ohne Kritik geblieben¹⁷: Zum einen wird bemängelt, solchen Typologien fehle zumeist ein Theoriebezug, eine klare Fragestellung, so daß die Typisierung häufig um ihrer selbst vorgenommen werde und keine Ansatzpunkte für weitergehende Erkenntnismöglichkeiten biete. Zum anderen sei die Bildung von Funktionstypen und die Zuordnung von Städten weitgehend willkürlich, da das Ergebnis von a priori zu treffenden Entscheidungen über die zugrunde gelegten Klassifikationsmerkmale und Schwellenwerte abhängt. Beide Einwände sind nicht unbezweifelbar. Rein deskriptive Klassifizierungen um ihrer selbst willen haben sicherlich nur einen begrenzten wissenschaftlichen Wert; sie werden erst dann fruchtbar, wenn sie Ausgangspunkte für weitergehende Untersuchungen, etwa über Zusammenhänge mit räumlichen Verteilungsmustern, Größenkategorien, Wachstumsraten usw., bilden. Auf einige solcher weiterführenden Analysemöglichkeiten soll am Ende kurz eingegangen werden.

Mit dem zweiten Einwand ist genau die Problematik angesprochen, in der durch die Anwendung der Faktorenanalyse möglicherweise ein Fortschritt erzielt werden kann. Denn wenn die bei *konventionellen* funktionalen Städteklassifizierungen erforderliche Auswahl von einigen wenigen als wesentlich erachteten Klassifikationsmerkmalen aus einer prinzipiell unbegrenzten Zahl möglicher Klassifikationsmerkmale als willkürlich kritisiert wird, so liegt die Frage nahe, ob dieser Auswahlprozeß nicht objektiviert werden kann. Anstelle mehr oder weniger gut begründbarer Vorentscheidungen über die Relevanz von Klassifikationsmerkmalen sollte es möglich sein, aus den verfügbaren Daten selbst, gleichsam induktiv, die *wesentlichen* Kriterien zu ermitteln. Genau dies leistet die Faktorenanalyse, allerdings aufgrund eines ganz bestimmten Konzeptes von „Wesentlichkeit“: Sie *bündelt* diejenigen Merkmale, die untereinander jeweils korrelieren, zu neuen künstlichen Merkmalen, die untereinander unkorreliert sind. Dabei liegt die Vorstellung zugrunde, daß mehrere Einzelmerkmale, die untereinander hoch korrelieren, also ein gleiches oder ähnliches Verteilungsmuster in der Städtegruppe aufweisen, als Indikatoren für *eine* Funktion angesehen werden können, daß die empirisch meßbaren, also die *manifesten* Merkmale als Indikatoren angesehen werden können, aus deren Interkorrelationen auf die *dahinter stehenden, latenten* Grunddimensionen geschlossen werden kann.

Damit ist die Fragestellung der folgenden empirischen Analyse umrissen: eine weitgehend induktive Untersuchung der Wirtschaftsstruktur von Städten des Deutschen Reichs auf inhärente Grunddimensionen, die sich aufgrund der räumlichen Verteilungsmuster der einzelnen ökonomischen Aktivitäten innerhalb des Stadtsystems ergeben. Als Datengrundlage für eine solche Teststudie wurde die Berufstatistik von 1907 ausgewählt. Dieser Zeitschnitt erscheint besonders aufschlußreich, weil er bereits gegen Ende der stürmischen Entwicklung der Hochindustrialisierungsperiode liegt und die tiefgreifenden Strukturwandlungen noch unüberformt erkennen lassen müßte. Ergänzungen dieses statischen Querschnittsansatzes durch

¹⁷ Hadden und Borgatta, *American Cities*, S. 17 f., S. 70 f.; Berry und Smith, *Handbook*, S. 13, S. 47; Schöller, *deutsche Städte*, S. 17; Smith, *Method*.

komparativ-statische oder gar Längsschnittanalysen wären sicherlich wünschenswert, müssen jedoch weiterführenden Untersuchungen vorbehalten bleiben.

Gegenüber der im gleichen Jahr stattgefundenen gewerblichen Betriebszählung, der Vorläuferin der heutigen Arbeitsstättenzählung, hat die Berufszählung den Vorteil, daß allein sie einen annähernd lückenlosen Überblick über das Erwerbsleben bietet, während in der Betriebszählung damals nicht nur die Landwirtschaft und die nicht Erwerbstätigen wie Studenten und Rentner, sondern auch weite Teile des tertiären Sektors, insbesondere der öffentliche Dienst, unberücksichtigt blieben. Andererseits erfaßt die Berufszählung die Person am Wohnort, die Betriebszählung dagegen – für unsere Fragestellung an sich zweckmäßiger – am Arbeitsort, so daß einige Verzerrungen im Fall starker Pendlerverflechtungen zu beachten sind. Allerdings hat die Berufszählung wiederum den Vorteil, daß die Zuordnung der Erwerbstätigen zu den Kategorien der Wirtschaftssystematik nicht wie bei der Betriebszählung nach der überwiegenden Art des Betriebes erfolgt, wodurch bekanntlich erhebliche Verzerrungen in der Beschäftigungsstatistik resultieren. Deshalb wurde nach Abwägung der Vor- und Nachteile schließlich der Berufszählung der Vorzug gegeben.

Da die wirtschaftssystematische Aufgliederung der Berufe in der benutzten Quelle¹⁸ mit 218 Kategorien überaus differenziert ist und manche in Städten kaum vorkommenden Berufsarten enthält, wurde zunächst eine Zusammenfassung von sehr gering besetzten Berufsarten innerhalb der übergeordneten Berufsgruppen vorgenommen und damit die Zahl der in die Analyse eingehenden Berufsarten von 218 auf 105 reduziert. Wegen der verschwindend geringen Erwerbstätigenzahlen in den aggregierten Berufsarten sind kaum Verzerrungen zu befürchten; es wurde dadurch jedoch erreicht, daß sämtliche Erwerbstätige der Städte bei der Analyse berücksichtigt werden können, so daß die Notwendigkeit zur Variablenauswahl entfällt.

Gewichtiger erscheint demgegenüber die Problematik der wirtschaftssystematischen Gliederung. Beispielsweise ist die Abteilung „Handel“ nur verhältnismäßig gering unterteilt und enthält die heterogene und stark besetzte Berufsart *Waren- und Produktenhandel*, während andere Berufsgruppen wie z. B. die meisten handwerklichen Zweige, verhältnismäßig tief aufgegliedert sind. Dadurch entstehen Repräsentations- und Gewichtungsprobleme, auf die noch einzugehen sein wird. Im übrigen dürfte die Aufgliederung jedoch fein genug sein, um bei der *Bündelung* der einzelnen Berufsarten nicht zu sehr von Zufälligkeiten und Verzerrungen der Gliederungssystematik beeinflußt zu werden.

Die Auswahl der zu untersuchenden Städte konnte aus technischen Gründen zunächst nur die 42 Großstädte des Deutschen Reiches in ihrer damaligen administrativen Abgrenzung umfassen. Hiervon sind fünf (Altona, Barmen, Charlottenburg, Rixdorf (Neukölln) und Schöneberg) als agglomerationsangehörige Gemeinden so eng mit der Kernstadt ihrer jeweiligen Agglomeration (Hamburg, Elberfeld, Berlin) verflochten, daß sie pro Agglomeration zusammengeschlossen wurden. Untersu-

¹⁸ *Berufs- und Betriebszählung vom 12. Juni 1907*, Berufsstatistik, Abt. VI: *Großstädte*, Berlin 1910, (= Statistik des Deutschen Reichs, Band 207).

chungen, in denen solche Aggregierungen unterlassen wurden¹⁹, sowie eigene Vergleichsanalysen zeigen, daß sonst die sehr viel stärker ausgeprägten funktionalen Differenzierungen *innerhalb* von Agglomerationen das Bild überlagern würden. Eine Einbeziehung von Mittel- und evtl. von Kleinstädten sowie der Vergleich unter den Analyseergebnissen verschiedener Stadtgrößenklassen wäre wünschenswert, doch steht hier die Erprobung des methodischen Instrumentariums im Vordergrund, während ergänzende Anlaysen einer gesonderten Arbeit vorbehalten bleiben müssen.

Aufgrund dieser Überlegungen ergibt sich die in *Abb. 1* auszugsweise dargestellten Datenmatrix, die allen folgenden Analysen zugrunde liegt. Ihre 37 Zeilen repräsentieren die Städte, ihre 105 Spalten die Berufsarten, im folgenden auch Variablen genannt. Um nun die untereinander korrelierenden Variablenbündel durch unkorrelierte *Kunstvariablen*, die sog. Faktoren zu ersetzen, sind mehrere Analyseschritte erforderlich, die im folgenden knapp graphisch erläutert werden.

Zunächst erfolgt eine Umrechnung der Absolutwerte in Prozentwerte, um die Größenunterschiede zwischen den Städten herauszufiltern, die sonst die unterschiedlichen Ausprägungen der Berufsarten in den Städten überlagern würden. Denn wie in *Abb. 1* ersichtlich, sind beispielsweise in Berlin viel Landwirte, Hüttenwerker, Goldschmiede und Klempner aufgeführt, doch deutet dies natürlich nicht auf ein gemeinsames Standortmuster dieser Berufsarten hin, sondern allein auf die Größe Berlins. Würde man nicht mit Relativ-, sondern mit Absolutwerten weiterarbeiten, ergäben sich zahlreiche triviale Korrelationen allein aufgrund der Stadtgrößen, so daß die Funktionsunterschiede durch einen allgemeinen, wenig aussagekräftigen Größenfaktor überlagert und größtenteils verdeckt würden.

Im nächsten Schritt wird die sog. *z-Standardisierung* vorgenommen, durch die die ursprünglichen Prozentwerte, die im folgenden auch Rohwerte oder *x*-Werte genannt werden, in normierte *z*-Werte umgewandelt werden. Dieser Schritt soll anhand von zwei ausgewählten Variablen (Berufsarten *C 1–10*: Gruppe Handelsgewerbe; Berufsart *E 2*: öffentliche Verwaltung, Rechtspflege) graphisch verdeutlicht werden. Die Rohwerte lassen sich zunächst in Form eines einfachen Streuungsdiagramms darstellen, das durch die beiden Variablen als Achsen aufgespannt wird und in dem die Städte als Punkte repräsentiert sind. Hier lassen sich an den Achsenskalen die Prozentanteile der hier berücksichtigten Berufsarten unmittelbar ablesen²⁰.

Im ersten Teilschritt der Standardisierung wird von allen Rohwerten der jeweilige Variablenmittelwert subtrahiert. Dies hat graphisch zur Folge, daß das Achsenkreuz des Streuungsdiagramms in die Mitte, exakter: in das arithmetische Mittelzentrum des Punkteschwarmes, verschoben wird, ohne daß dadurch jedoch die Konfiguration der Punkte verändert wird (*Abb. 2b*). Die auf den Achsen nach dieser Translation ablesbaren Werte bezeichnet man als *Abweichungswerte*, d. h. als Ab-

¹⁹ Beispielsweise bei Hadden und Borgatta, *American Cities*, sowie bei Berry und Smith, *Handbook*.

²⁰ Um zugleich die besondere Struktur der fünf genannten agglomerationsangehörigen Städte zu verdeutlichen, werden in dem Zwei-Variablen-Beispiel im Unterschied zur Hauptanalyse die 42 Städte in unaggregierter Form dargestellt.

Abb.1 Ausschnitte aus den Matrizen der absoluten und relativen Rohwerte

	1 Landwirtschaft, Gärtnerei, Tierzucht, Forstwirtschaft, Fischerei	2 Hüttenbetrieb, auch Frisch- und Streckwerke	3 Gewinnung von Stein- und Braunkohlen, Briquetfabri- kation, Koks	4 Sonstiger Bergbau, Salinen- wesen, Torfgräberei	5 Industrie der Steine und Erden	6 Gold- und Silberschmiede, Juweliere, sonstige Verar- beitung edler Metalle, Münzstätten	7 Eisengießerei, Emaillierung von Eisen	8 Klempner	9 Grob- (Huf-) Schmiede	10 Schlosserei, einschl. Ver- fertigung von feuerfesten Geldschranken usw.	11 Sonstige Gewerbe der Metallverarbeitung	12 Verfertigung von Maschinen, Werkzeugen, Apparaten	...	105 Von Unterstützung Lebende, Insassen von Wohltätigkeits- anstalten, Armenhäusern usw.
1 Aachen	1320	2263	84	61	886	90	229	122	121	1225	3330	1689	...	1418
2 Berlin	6692	409	149	255	8104	2799	4976	6913	6042	32548	25610	33260	...	18607
3 Bochum	538	830	14879	0	884	7	2563	245	533	1470	253	1263	...	177
4 Braunschweig	1334	2	3	35	639	135	330	452	488	1078	842	4479	...	806
5 Bremen	3808	2	3	16	646	596	468	777	582	1294	381	2929	...	2875
6 Breslau	2453	135	150	45	2008	373	659	941	1614	4623	2324	5818	...	4898
7 Chemnitz	895	126	27	6	1521	43	2367	833	1032	1370	2729	19031	...	1055
8 Danzig	1405	91	1	0	329	67	94	304	1059	1983	340	824	...	1290
9 Dortmund	607	9248	5336	28	1312	15	327	259	686	3221	363	5337	...	299
...
37 Wiesbaden	1127	4	7	13	539	82	46	230	115	1006	294	468	...	496

In Prozent aller am Ort hauptberuflich Erwerbstätigen:														
1 Aachen	1.86	3.19	0.12	0.09	1.25	0.13	0.32	0.17	0.17	1.73	4.70	2.38	...	2.00
2 Berlin	0.51	0.03	0.01	0.02	0.62	0.21	0.38	0.52	0.46	2.47	1.94	2.53	...	1.41
3 Bochum	1.06	1.64	29.34	0.00	1.74	0.01	5.05	0.48	1.05	2.90	0.50	2.49	...	0.35
4 Braunschweig	2.11	0.00	0.00	0.06	1.01	0.21	0.52	0.71	0.77	1.70	1.33	7.07	...	1.27
5 Bremen	3.63	0.00	0.00	0.02	0.62	0.57	0.45	0.74	0.55	1.23	0.36	2.79	...	2.74
6 Breslau	1.07	0.06	0.07	0.02	0.88	0.16	0.29	0.41	0.70	2.02	1.01	2.54	...	2.14
7 Chemnitz	0.70	0.10	0.02	0.00	1.18	0.03	1.84	0.65	0.80	1.07	2.13	14.82	...	0.82
8 Danzig	1.87	0.12	0.00	0.00	0.44	0.09	0.13	0.41	1.41	2.65	0.45	1.10	...	1.72
9 Dortmund	0.77	11.74	6.77	0.04	1.67	0.02	0.42	0.33	0.87	4.09	0.46	6.77	...	0.38
...
37 Wiesbaden	2.19	0.01	0.01	0.03	1.05	0.16	0.09	0.45	0.22	1.95	0.57	0.91	...	0.96

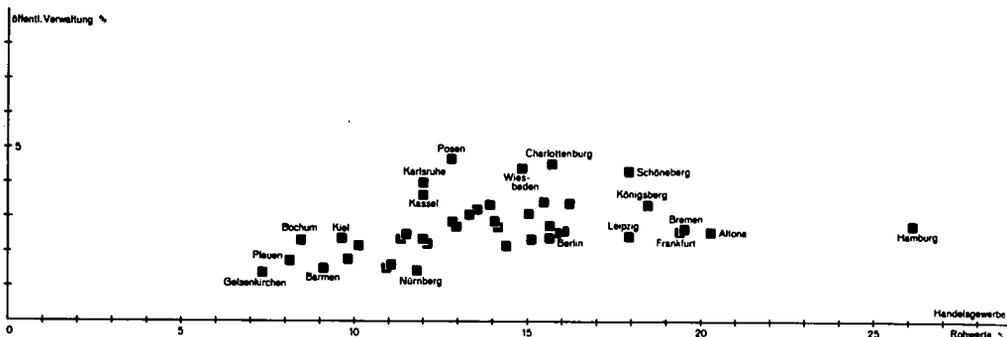
Quelle: Berufs- und Betriebszählung vom 12. Juni 1907. Berufsstatistik. Abt. VI: Großstädte. Berlin 1910. = Statistik des Deutschen Reichs, Band 207.

weichungen vom jeweiligen Variablenmittelwert; sie werden im folgenden auch y -Werte genannt.

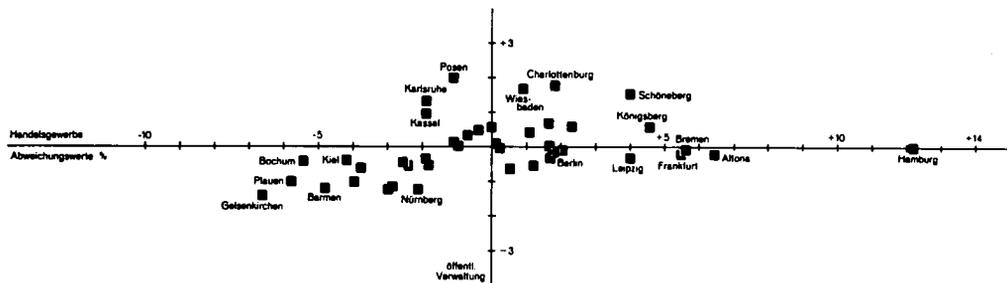
In einem weiteren Schritt werden nun auch die Streuungen der beiden Variablen vereinheitlicht, indem jeder Abweichungswert durch die Standardabweichung s der jeweiligen Variablen dividiert wird. Graphisch hat diese Transformation zur Folge, daß der zuvor mehr in der Richtung der Abszisse gelegene, schmale Punkteschwarm

Abb. 2 Z-Standardisierung am Zwei-Variablen-Beispiel

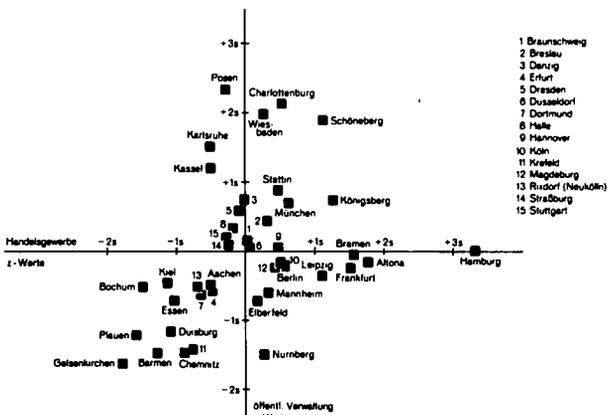
2a Streuungsdiagramm der Rohwerte



2b Streuungsdiagramm der Abweichungswerte



2c Streuungsdiagramm der z-Werte



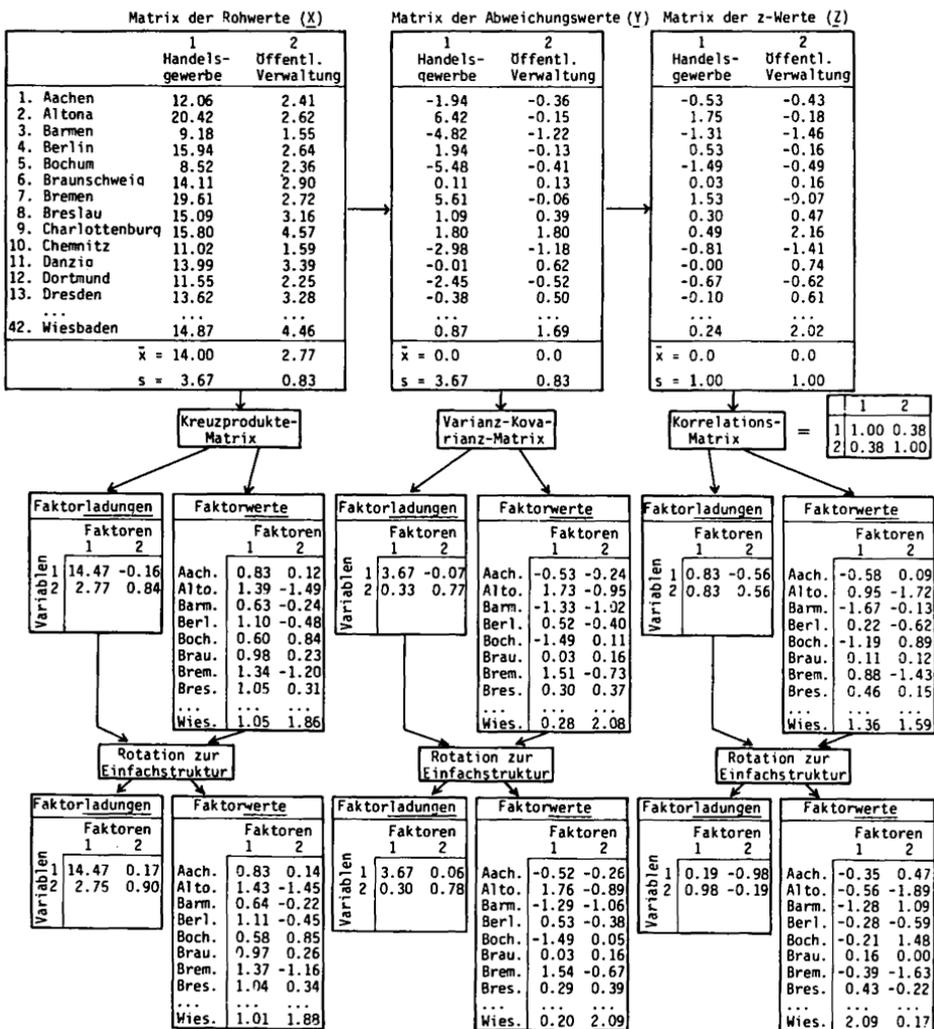
in der Richtung der Ordinate *auseinandergezogen* wird, so daß beide Achsen gewissermaßen das gleiche Gewicht erhalten. Die auf diese Weise entstehenden z -Werte haben die Eigenschaft, daß der Mittelwert aller Variablenwerte gleich 0 und die Standardabweichung sowie die Varianz gleich 1 sind. Durch diese Standardisierung sind sämtliche Größen- und Streuungsunterschiede *zwischen* den Variablen beseitigt, und lediglich die relativen Verteilungsunterschiede und -ähnlichkeiten auf die Städte gehen in die weiteren Analysen ein.

Diese Standardisierung wird, wenn auch implizit, vollzogen, wenn nun im nächsten Schritt die Korrelationen zwischen den Variablen berechnet werden. Um die interkorrelierenden Variablenbündel herauszufinden, müssen sämtliche möglichen Paare von Variablen gebildet und deren Korrelationskoeffizienten berechnet werden. Das Ergebnis läßt sich am zweckmäßigsten in Form einer Korrelationsmatrix darstellen, deren Zeilen und Spalten die Variablen repräsentieren. In unserem Zwei-Variablen-Beispiel erhalten wir somit eine kleine 2×2 -Tabelle (*vgl. Abb. 3*), deren Diagonalfelder 1.0 aufweisen (Korrelation einer Variablen mit sich selbst), während in den beiden übrigen Feldern der Korrelationskoeffizient r *zwischen* den Variablen ($= 0.38$) steht. Bei dem entsprechenden Analyseschritt des großen Datensatzes entsteht hier selbstverständlich eine sehr große Korrelationsmatrix (105×105), die jedoch im Prinzip gleich aufgebaut ist; d. h. sie enthält in der Diagonale soviel Einsen wie Variablen vorliegen (105), und außerhalb der Diagonalen in symmetrischer Anordnung die Korrelationen von sämtlichen möglichen Variablenpaaren.

Aus dieser Korrelationsmatrix werden nun durch ein mathematisches Theorem, auf das hier nicht näher eingegangen werden kann, die sog. Faktoren *extrahiert*. Das Prinzip läßt sich jedoch graphisch anhand des Zwei-Variablen-Beispiels verdeutlichen. Wenn wir davon ausgehen, daß die Faktorenanalyse versucht, mehrere korrelierende Variablen durch möglichst wenige Kunstvariablen zu ersetzen, so stellt sich in unserem Beispiel die Aufgabe, die beiden Variablen durch eine einzige neue Variable so zu ersetzen, daß die in den beiden Einzelvariablen enthaltene Information möglichst weitgehend durch die neue Variable erhalten bleibt. Diese Aufgabe wird optimal erfüllt, wenn die neue Achse, die den gesuchten Faktor repräsentiert, so gelegt wird, daß sie sich möglichst eng an den Punkteschwarm anlehnt (*Abb. 4a*). Der größte Teil der Information, die zuvor auf die beiden Einzelvariablen aufgeteilt war, ist nun in der durch die neue Achse repräsentierten *einen* Dimension vereinigt. Die Position der Städte auf dieser neuen Achse kann nun als Maß für eine kombinierte Handels- und Verwaltungsfunktion interpretiert werden. Den durch eine relativ hohe Handels- und Verwaltungsbedeutung geprägten Städten wie Hamburg, Schöneberg und Charlottenburg stehen am anderen Ende der neuen Skala die industriell geprägten Städte Gelsenkirchen, Plauen und Barmen gegenüber.

Durch diese neue Achse wird zwar die Haupttendenz des Punkteschwarmes im Streuungsdiagramm der z -Werte erfaßt, doch ist bei weitem noch nicht eine vollständige oder gar nur befriedigende Informationsverdichtung gelungen. Der verhältnismäßig geringen Korrelation von 0.38 entspricht eine sehr lockere Streuung des Punkteschwarmes um die Hauptachse, so daß die verbleibenden Abstände der Städte von dieser Achse noch beträchtlich sind. Diese Abweichungen, die durch die Haupt-

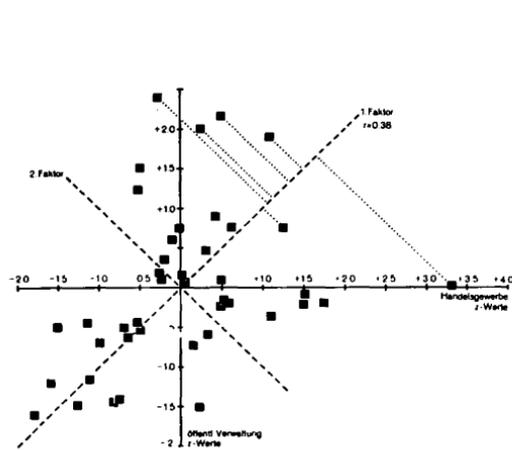
Abb. 3 Schema der verschiedenen Faktorisierungsvariationen am Zwei-Variablen-Beispiel



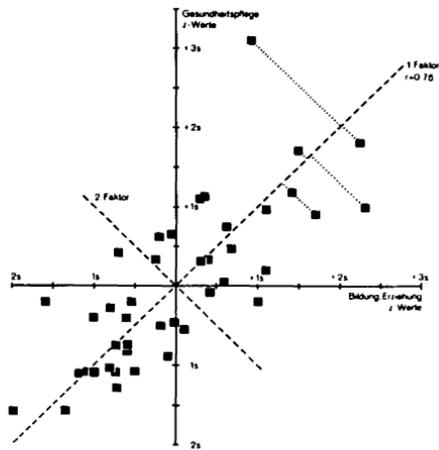
dimension nicht repräsentiert sind, können nun durch eine zweite Achse erfaßt werden, die senkrecht auf der ersten steht. Die Position der Städte auf dieser zweiten Achse gibt an, inwieweit ihre Struktur von der Hauptdimension abweicht und typologisch entweder mehr vom Handel oder von der öffentlichen Verwaltung geprägt wird.

Abb. 4 Faktorisierung von zwei z-standardisierten Variablen

4a Faktorisierung von zwei schwach korrelierenden Variablen



4b Faktorisierung von zwei hoch korrelierenden Variablen



Im Grunde ist mit der Bestimmung dieser neuen Achsen, durch die die gesuchten Faktoren veranschaulicht werden können, lediglich eine Drehung des Achsen-systems vorgenommen worden, und zwar so, daß mit *einer* Achse die Position des Punkteschwarmes möglichst gut erfaßt wird. In der Tat kann man die Faktoren-bzw. Hauptkomponentenanalyse im wesentlichen als ein Verfahren zur Achsen-transformation auffassen. Der Sinn der Transformation besteht darin, durch neue künstliche Achsen gewissermaßen länglich geformte Punkteschwärme aufzuspüren und abzubilden, da diese ellipsenförmigen Punkteschwärme hohe Korrelationen symbolisieren, durch die wiederum die gesuchten Variablenbündel angezeigt werden.

An diesem Zwei-Variablen-Beispiel kann bereits ein Problem der Faktorenanalyse aufgezeigt werden. Während den ursprünglichen Variablenachsen eine konkrete empirische Bedeutung zukommt, da sie ja bestimmte Berufsarten repräsentieren, bilden die neuen Achsen künstliche Variablen, denen nicht ohne weiteres eine klare inhaltliche Bedeutung zukommt. Sie bedürfen einer inhaltlichen Interpretation, die zweckmäßigerweise von den Beziehungen zu den Ausgangsvariablen ausgeht. Auf diese Weise wurde die sich ergebende Hauptachse als komplexe Handels- und Verwaltungsdimension angesprochen.

Gegenüber diesem Problem kommt der entscheidende Vorteil der Faktorenanalyse erst zum Tragen, wenn durch die Konstruktion der neuen Achsen die Zahl der Variablen erheblich vermindert werden kann, indem mehrere, untereinander hoch korrelierende Variablen durch jeweils eine einzige Kunstvariable ersetzt werden können. In dem bisher angesprochenen Zwei-Variablen-Beispiel gelingt diese Zusammenfassung nur unzureichend, aber wenn in einem neuen Beispiel zwei höher korre-

lierende Variablen ausgewählt werden, so wird die Berechtigung deutlich. Bei den in *Abb. 4b* dargestellten beiden Variablen 98 (Bildung, Erziehung und Unterricht) und 99 (Gesundheitspflege und Krankendienst) ergibt sich eine Korrelation von $r = 0.76$, der eine verhältnismäßig dichte Schärung des Punkteschwarmes um die den ersten Faktor repräsentierende Hauptachse entspricht. In diesem Fall lassen sich die beiden Ausgangsvariablen bereits erheblich besser durch die neuen Faktorenachsen ersetzen, denn die verbleibenden Abweichungen sind verhältnismäßig gering.

Nachdem die Faktoren extrahiert sind, benötigen wir Angaben über die Beziehungen zwischen den neuen Achsen, also den gesuchten Faktoren, und den Ausgangsvariablen. Für eine inhaltliche Interpretation der Faktoren muß beachtet werden, für welche Einzelvariablen die neuen Kunstvariablen stehen. Diese Angaben kann man der Tabelle der sog. Faktorladungen entnehmen; sie gibt an, welche Variablen die neuen Faktoren hoch *laden*, d. h. welche Variablen für die Bildung welcher Faktoren verantwortlich sind. Die Faktorladungen sind – jedenfalls im *Normalfall* der Faktorenanalyse²¹ – sowohl als Korrelationskoeffizienten zwischen den Einzelvariablen und den Faktoren wie auch als Regressionskoeffizienten der Einzelvariablen auf die Faktoren zu interpretieren.

Zum anderen benötigen wir Angaben über die Werte der untersuchten Städte auf den neuen Faktorenachsen. Diese sog. Faktorwerte können in dem Zwei-Variablen-Beispiel unmittelbar als Projektionen der Städte auf die Skalen der neuen Achsen abgelesen werden.

Diese beiden Tabellen, die Matrix der Faktorladungen und die Matrix der Faktorwerte, sind im wesentlichen das Ergebnis der Faktorenanalyse. Die Matrix der z -standardisierten Rohwerte wird gleichsam zerlegt in die Matrizen der Faktorladungen und Faktorwerte (vgl. *Abb. 3*); das Produkt dieser beiden Matrizen reproduziert die Ausgangsdaten. Lösbar wird dieses Problem, indem zunächst aus den z -Werten die Korrelationsmatrix errechnet wird und aus dieser mit Hilfe des sog. Fundamentalsatzes der Faktorenanalyse (Thurstone) die Matrix der Faktorladungen ermittelt wird.

Wenn wir in der Darstellungsform der Matrixalgebra die Tabelle der standardisierten Ausgangswerte mit Z bezeichnen, die quadratische Tabelle der Korrelationen mit R , die Tabelle der Faktorladungen mit A sowie schließlich die Tabelle der Faktorwerte mit P , so lassen sich die algebraischen Zusammenhänge nach dem Modell der Hauptkomponentenanalyse durch die folgenden Grundgleichungen darstellen:

$$\begin{aligned} Z^t &= A \cdot P^t \\ R &= Z^t \cdot Z \cdot \frac{1}{n} \\ R &= A \cdot A^t \end{aligned}$$

Dabei bedeutet ein hochgestelltes t *transponiert*, d. h. bei der betreffenden Matrix werden Zeilen und Spalten miteinander vertauscht. n gibt die Zahl der Untersuchungsobjekte an und entspricht damit der Zahl der Zeilen in den Matrizen Z und

²¹ Hier verstanden als orthogonale Lösung der Faktorisierung einer Korrelationsmatrix.

P. Die dritte Gleichung bezeichnet man als *Fundamentaltheorem* der Faktorenanalyse nach Thurstone²².

Während sich die Analyseschritte graphisch nur im zwei- bis dreidimensionalen Fall veranschaulichen lassen, sind rechnerisch auch Lösungen mehrdimensionaler, also multivariater Problemstellungen möglich. Erst wenn die Analyse auf größere Datensätze angewandt wird, gewinnt der Effekt der Variablenbündelung und Informationsverdichtung an Bedeutung. Prinzipiell ebenso wie in unserem Zwei-Variablen-Beispiel verfahren wurde, kann nun die Faktorisierung des großen Datensatzes vorgenommen werden, indem aus der 105×105 -Korrelationsmatrix die Faktoren extrahiert und die gesuchten Matrizen der Faktorladungen und Faktorwerte berechnet werden.

Bei diesem Schritt ist eine Entscheidung darüber erforderlich, wieviel Faktoren im folgenden berücksichtigt werden sollen. Anhand des Zwei-Variablen-Beispiels wurde deutlich, daß zwei Variablen vollständig durch zwei Faktoren ersetzt werden können, d. h. daß sich die Punktekonfiguration vollständig im Raum von zwei Faktoren beschreiben ließ. Das gleiche gilt im Fall der 105 Variablen, dem ein Punkteschwarm in einem – selbstverständlich nicht mehr anschaulich vorstellbaren – 105-dimensionalen Achsensystem entspricht. Unser Ziel ist jedoch, diesen Punkteschwarm durch wesentlich weniger Dimensionen zu beschreiben und dabei nur diejenigen Dimensionen zu berücksichtigen, die im Sinne des Zwei-Variablen-Beispiels als *Hauptachsen* angesehen werden können, während all diejenigen Dimensionen, die lediglich die restlichen Abweichungen auf sich vereinigen, zu vernachlässigen sind.

Dieses Problem der Bestimmung der Anzahl der zu berücksichtigenden Faktoren ist im allgemeinen nicht eindeutig lösbar. Zumeist wird geprüft, inwieweit die entstehenden Faktoren inhaltlich interpretierbar sind, und nach diesem Gesichtspunkt wurden in diesem Fall 7 Faktoren berücksichtigt, während auf die Information, die auf die restlichen Dimensionen entfällt, verzichtet wird. Diese 7 Faktoren können nun als neue Achsen aufgefaßt werden; der 105dimensionale sehr unübersichtliche Variablenraum wird somit durch einen wesentlich übersichtlicheren 7dimensionalen Faktorenraum angenähert und ersetzt. Durch die teilweise recht hohen Korrelationen unter den 105 Variablen wird es möglich, zahlreiche Einzelvariablen so zusammenzufassen, daß der größte Teil der auf die Einzelvariablen entfallenden Information in dem neuen Faktorensystem erhalten werden kann. Allerdings geht ein Teil der Information verloren, und vor allem diejenigen Variablen, die nur wenig mit anderen korrelieren, tragen kaum zur Faktorenbildung bei. Insgesamt reproduzieren die 7 Faktoren 57,6 % der ursprünglich auf die 105 Variablen entfallenden (durch die z-Standardisierung normierten) Varianz.

Zwar ist nun das erste Ziel der Hauptkomponentenanalyse, die Reduktion der zahlreichen Variablen auf wenige Faktoren, erreicht; doch um das Ergebnis auch inhaltlich fruchtbar zu machen, bedarf es im allgemeinen eines weiteren Analyseschrittes, der sog. Rotation zur Einfachstruktur. Bisher war die Faktorenbildung darauf abgestellt, mit dem ersten Faktor die allgemeine Hauptrichtung des viel-

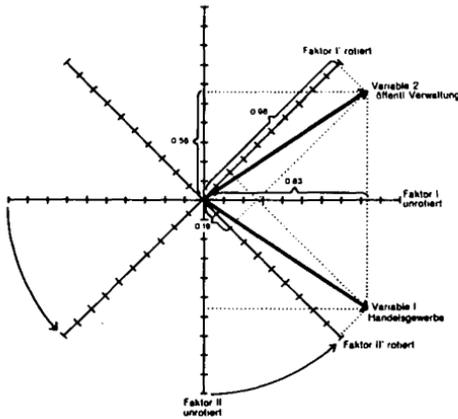
²² Überla, *Faktorenanalyse*, S. 52 f.

dimensionalen Punkteschwarmes, also einen möglichst großen Komplex des Interkorrelationsgeflechtes abzubilden, während der zweite Faktor die danach übrig bleibende Haupttendenz, der dritte Faktor die danach noch vorhandene Tendenz usw. erfaßte. Dieser Ansatz führt nicht unbedingt dazu, die gesuchten Variablen**bü**ndel aufzuspüren; hierzu ist eine neue Drehung des Faktorenachsensystems erforderlich, um die Achsen gezielt auf hoch korrelierende Variablengruppen auszurichten.

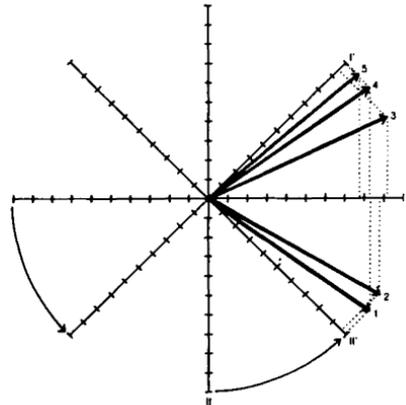
Zur Verdeutlichung des Prinzips der Rotation zur Einfachstruktur soll noch einmal an unser Zwei-Variablen-Beispiel angeknüpft werden. Es lassen sich nämlich nicht nur die neu gebildeten Faktoren im Raum, d. h. im Achsensystem der Ausgangsvariablen veranschaulichen, sondern umgekehrt auch die Variablen im Raum der Faktoren, indem durch die Faktorladungen die Position der beiden Variablen als Vektoren im Faktorenraum angegeben wird (Abb. 5). Dabei ist übrigens der Cosinus des Winkels zwischen beiden Vektoren gleich dem Korrelationskoeffizienten zwischen beiden Variablen, so daß sich in dieser Darstellungsweise untereinander hoch korrelierende Variablen durch eng beieinander liegende Vektoren**bü**ndel symbolisieren lassen, während andererseits die Vektoren von zwei unkorrelierten Variablen senkrecht, also orthogonal zueinander stehen ($\cos 90^\circ = r = 0.0$).

Abb. 5 Rotation zur Einfachstruktur

5a Rotation zur Einfachstruktur am Zwei-Variablen-Beispiel



5b Rotation zur Einfachstruktur bei zwei Faktoren und fünf Variablen (fiktives Beispiel)



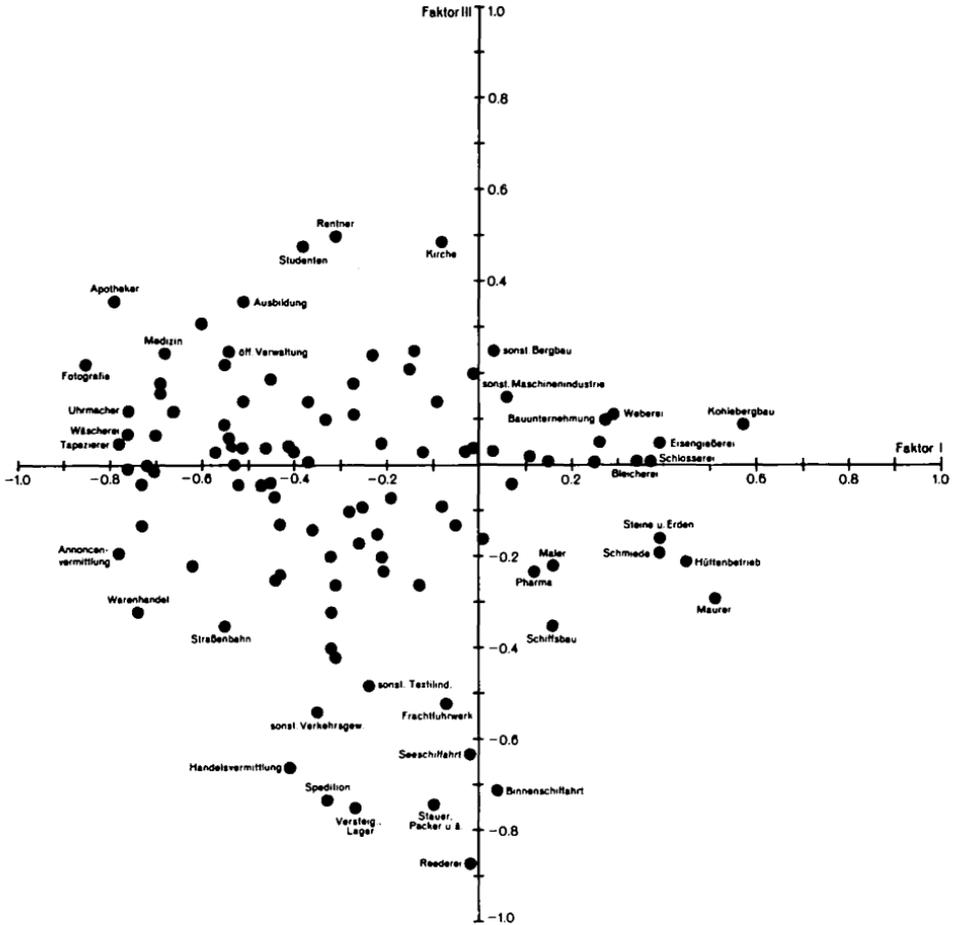
In unserem einführenden Beispiel gab der erste Faktor die Haupttendenz des zweidimensionalen Punkteschwarmes wieder, und dementsprechend liegt die Achse genau in der Hauptrichtung der beiden Variablenvektoren, die beide mit einem Winkel von 33.9° ($\cos 33.9^\circ = \text{Ladung} = r = 0.83$) von der Faktorennachse abweichen. Dementsprechend entfallen auf den zweiten Faktor nur noch Ladungen von 0.56, die einem Winkel von 55.9° zu den Variablen entsprechen.

Sofern man nun die Verdichtung der Information beider Variablen zu *einer* Dimension anstrebt und den zweiten Faktor vernachlässigen möchte, kann hier bereits abgebrochen werden. Eine Rotation zur Einfachstruktur dagegen versucht, die Position der Faktorenachsen möglichst eng an den Variablenvektoren auszurichten, so daß die Variablen entweder möglichst hohe oder niedrige, jedoch möglichst wenige mittelhohe Ladungen aufweisen, um die Interpretation der Faktoren zu erleichtern. In unserem Beispiel würde eine Rotation die Faktorenachsen möglichst eng an die beiden Variablenvektoren anlehnen (vgl. *Abb. 5a*), obwohl dies nur annäherungsweise möglich ist, solange die Faktoren im Unterschied zu den Variablen unkorreliert, also orthogonal sind²³. Dieser Schritt erscheint jedoch in unserem Zwei-Variablen-Beispiel wenig sinnvoll, da ja gerade eine Reduktion der Dimensionen zur Informationsverdichtung das wichtigste Ziel unserer Analyse ist, und so soll ein weiteres, hypothetisches Beispiel das Prinzip der Rotation zur Einfachstruktur verdeutlichen. Nehmen wir an, 5 Variablen ließen sich vollständig im Raum zweier Faktoren darstellen – dies ist möglich bei einer bestimmten Interkorrelationsstruktur –, so würde in unrotierter Form der erste Faktor wiederum die Haupttendenz der Variablenrichtung erfassen und dadurch eine Position genau *zwischen* den beiden Variablenbündeln einnehmen. Wie in *Abb. 5b* ersichtlich, kann erst durch eine Rotation zur Einfachstruktur eine gezielte Ausrichtung an den Bündeln der Variablenvektoren erreicht werden. Erst jetzt werden die Faktoren inhaltlich gut interpretierbar, indem einerseits hohe Ladungen anzeigen, daß der Sachverhalt der betreffenden Variablen gut durch den Faktor wiedergegeben wird, während Ladungen um 0 anzeigen, daß zwischen den betreffenden Faktoren und Variablen keine oder kaum Zusammenhänge bestehen.

Um die Übertragung dieses Ansatzes auf die Analyse des großen Datensatzes zu verdeutlichen, sei zunächst das unrotierte Faktorenmuster in zwei Diagrammen dargestellt. Aus Platzgründen können für die graphische Veranschaulichung allerdings nur die ersten drei Faktoren, die zusammen aber immerhin 36,7 % der gesamten Variablenvarianz repräsentieren, berücksichtigt werden (*Abb. 6*). Es zeigt sich, daß der erste Faktor bereits in unrotierter Form recht gut interpretiert werden kann, da er offenbar eine allgemeine Dimension *Industrie – tertiärer Sektor* anzeigt. Die Konzentration des Variablenschwarmes im linken Bereich des Diagramms (hohe Ladungen auf dem ersten Faktor) kann als Indikator für eine komplexe Zentralitätsfunktion der Städte angesprochen werden, die in erster Linie aus den Berufsarten des tertiären Sektors, aber auch aus einer Reihe von Gewerbesparten des sekundären Sektors mit marktorientiertem Standortverteilungsmuster im Städtesystem gebildet wird. Dabei sei zum Verständnis des Diagramms angemerkt, daß das negative Vorzeichen der Ladungen insofern bedeutungslos ist, als sich mathematisch und interpretatorisch völlig gleichwertige Lösungen ergeben, wenn die Vorzeichen sämtlicher Ladungen eines oder auch aller Faktoren umgekehrt werden.

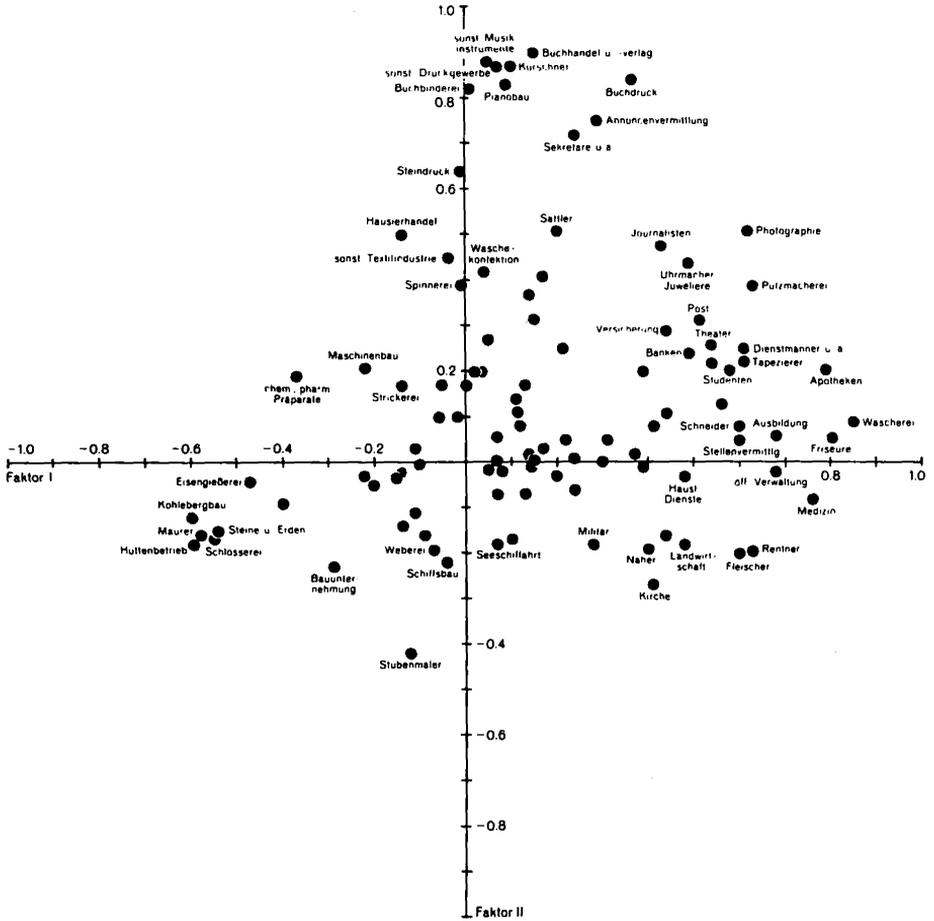
Der zweite (unrotierte) Faktor ist demgegenüber weniger klar deutbar. Die höchsten Ladungen zeigen die Variablengruppen Druckerei, Verlag und Musikinstrumen-

²³ Eine schiefwinklige Rotation würde die Variablen exakt reproduzieren.



sten Blick ersichtlich, verändert sich das Bild nur geringfügig. Tendenziell können durch die gezielte Ausrichtung der Achsen an den jeweils besonders hoch ladenden Variablenbündeln einige Ladungen erhöht und dafür andere, schwer deutbare mittel-hohe Ladungen verringert werden, so daß die Beziehungen zwischen Faktoren und Variablen etwas klarer werden. Bei der hier vorliegenden Variablenstruktur und hohen Variablenzahl war von vornherein nicht mit einer sehr klaren Einfachstruktur zu rechnen.

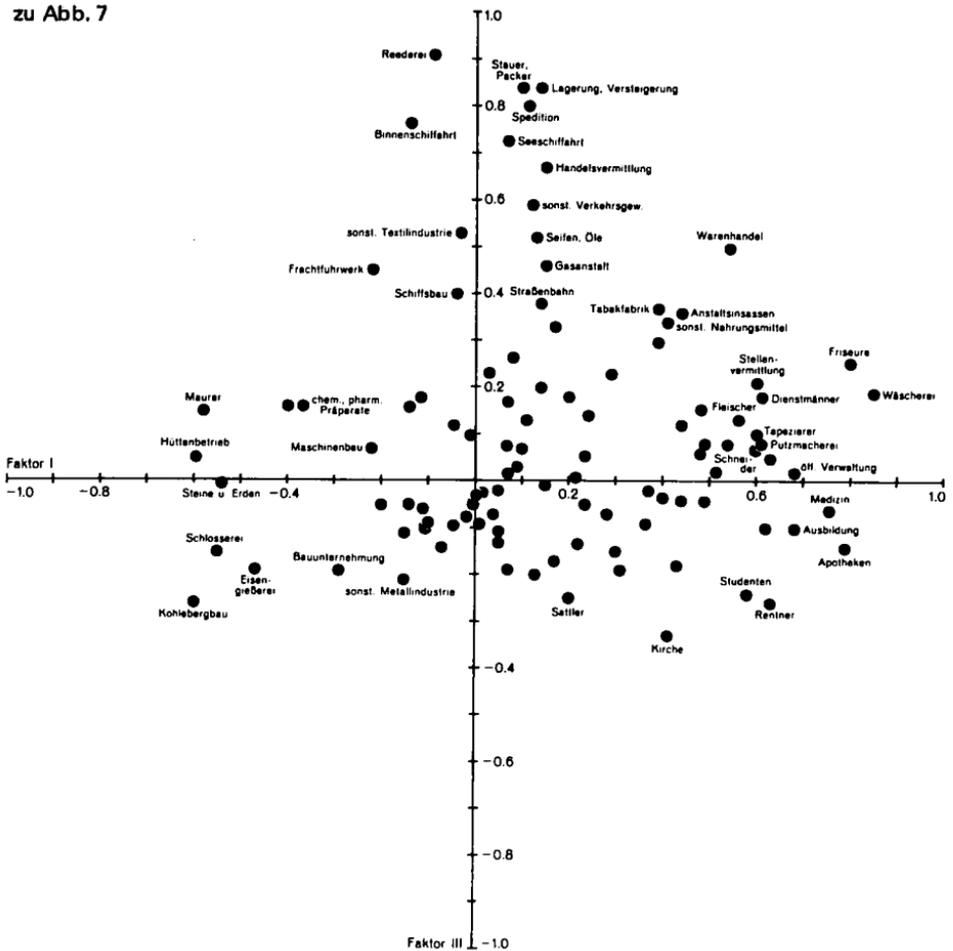
Abb.7 Die Variablen im Raum der drei ersten orthogonal rotierten Faktoren



Weitere Ergebnisse dieser Analyse, insbesondere auch über die übrigen Faktoren, werden in tabellarischer Form in *Abb. 8* mitgeteilt, auf die anstelle einer näheren verbalen Umschreibung der Einzelergebnisse verwiesen wird. Da jeder dieser Faktoren ein ganzes Bündel von Einzelvariablen repräsentiert, können diese Faktoren als untereinander unkorrelierte Hauptdimensionen der ökonomischen Funktionsstruktur im System der deutschen Großstädte um 1907 angesprochen werden.

Darüber hinaus interessieren die Ausprägungen dieser Faktoren in den untersuchten Städten. Diese Informationen enthält die Matrix der Faktorwerte, die Auf-

zu Abb. 7



schlüsse über die Positionen der Städte im 7dimensionalen Raum der Faktoren gibt. Da die Faktoren unkorrelierte Hauptdimensionen repräsentieren, erscheint es gerechtfertigt, hohe Faktorwerte als Funktionsschwerpunkte in den Städten zu interpretieren. An Stelle einer weiteren Tabelle wurden die Faktorwerte zur besseren Veranschaulichung kartographisch dargestellt (Abb. 9), wobei allerdings nur der jeweils höchste Faktorwert einer Stadt berücksichtigt werden konnte. Damit ist zwar noch keine Typisierung im strengen klassifikatorischen Sinne erreicht, sondern lediglich eine Darstellung überproportionaler Funktionsausprägungen; aber der sonst häufig

Abb. 8 Ergebnisse einer Hauptkomponentenanalyse der z-Werte (Korrelationsfaktorisierung)

Methode: Hauptkomponentenanalyse der Korrelationsmatrix, orthogonale Rotation nach Varimax mit 7 Faktoren.

Faktor I (bipolar): Tertiärer Sektor, insb. <u>private Dienstleistungen u. öffentl. Dienst;</u> <u>invers: Bergbau und Schwerindustrie</u> Varianzanteil: 13.6 % Höchste Faktorladungen der Variablen:		Faktor IV: Diversifizierte Leichtindustrie I, <u>insb. Metall-, Holz- und Elektroindustrie</u> Varianzanteil: 8.02 %	
58 Wäscherei, Wäscherinnen	+0.85	6 Gold- u. Silberschmiede, Juweliere .	+0.84
57 Friseurerei	+0.80	14 Automobil- u. Fahrräderfabrikation .	+0.79
24 Apotheker	+0.79	25 Sonstige chemische Industrie	+0.79
99 Gesundheitspflege, Krankendienst .	+0.76	21 Elektrotechniker, elektr. Maschinen	+0.77
96 Öffentl. Verwaltung, Rechtspflege .	+0.68	44 Sonstige Holzindustrie	+0.73
98 Bildung, Erziehung, Unterricht . .	+0.68	11 Sonstige Metallverarbeitung	+0.72
53 Putzmacherinnen	+0.63	8 Klempner	+0.71
103 Rentner, Pensionäre	+0.63	45 Bäckerei, Konditorei, Lebküchler .	+0.71
66 Fotografen	+0.62	19 Mathemat., physik., chem. Instrumente	+0.71
91 Dienstmänner, Wach- u. SchließBges.	+0.61	65 Stein- und Zinkdruckerei	+0.57
40 Tapezierer	+0.61	43 Tischlerei	+0.55
79 Stellenvermittlung	+0.60	...	
51 Schneider, Kleiderkonfektion	+0.60	Faktor V (bipolar): Tertiärer Sektor, insb.	
46 Fleischerei	+0.60	<u>Presse, Banken, Kultur;</u>	
104 Studierende, Schüler	+0.58	<u>invers: Diversifizierte Leichtindustrie II,</u>	
93 Gast- und Schankwirtschaft	+0.56	<u>insb. Maschinen-, Leder-, Nahrungsm.industrie</u>	
71 Waren- und Produktenhandel	+0.54	Varianzanteil: 5.97 %	
102 Musik, Theater, Schaustellungen . .	+0.54	70 Arbeiter, Gehilfen ohne nähere Zuordn.	+0.59
82 Post- und Telegraphenbetrieb	+0.52	84 Posthalterei, Personenfurwerk	+0.58
...		74 Zeitungsverlag u. -spedition	+0.57
5 Industrie der Steine und Erden	-0.54	100 Journalisten, Schriftsteller	+0.53
61 Maurer	-0.58	72 Geld- und Kredithandel	+0.51
2 Hüttenbetrieb	-0.60	...	
3 Kohlebergbau, Brikett- u. Koksfabrik.	-0.60	20 Verfertigung von Lampen	-0.43
Faktor II: Druck, Verlag, Musikinstrumente,		55 Schuhmacherei	-0.49
<u>Pelze</u>		22 Sonstige Maschinenindustrie	-0.50
Varianzanteil: 10.42 %		Faktor VI (bipolar): Textilindustrie I, insb.	
73 Buchhandel, Verlag	+0.90	<u>Weberei, Färberei, Posamentenfabrikation;</u>	
18 Sonstige Musikinstrumente	+0.88	<u>invers: Militär, Schiffsbau</u>	
67 Sonstiges Vervielfältigungsgewerbe .	+0.87	Varianzanteil: 5.73 %	
54 Kürschnerei, Pelzwarenzurichtung . .	+0.87	30 Weberei	+0.82
64 Buchdruckerei	+0.84	33 Färberei	+0.74
17 Pianofortefabrikation, Orgelbau . . .	+0.83	56 Sonstiges Bekleidungsgewerbe	+0.53
38 Buchbinderei, Kartonagenfabrikation .	+0.82	35 Posamentenfabrikation	+0.51
80 Annoncervermittlung, Auskunftbüros .	+0.75	29 Spinnerei, Hasperei, Zwirnerei usw.	+0.46
101 Stenographen, Privatsekretäre u.ä. .	+0.72	...	
65 Stein- und Zinkdruckerei	+0.64	15 Schiffsbau	-0.41
39 Riemerei und Sattlerei	+0.51	61 Maurer	-0.43
66 Fotografen	+0.51	16 Uhrmacher	-0.43
...		95 Militär	-0.45
Faktor III: Hafenfunktionen		9 Grobschmiede	-0.48
Varianzanteil: 8.15 %		63 Sonstiges Baugewerbe	-0.65
88 Reederei, Schiffsbefrachtung	+0.91	Faktor VII: Textilindustrie II, insb. Häkelei,	
77 Stauer, Schauerleute, Packer u.ä. . .	+0.84	<u>Stickerei, Spitzenfabr., Musterzeichner</u>	
78 Versteigerung, Verleihung, Aufbewahr.	+0.84	Varianzanteil: 5.70 %	
86 Spedition und Güterbeförderung . . .	+0.80	34 Bleicherei, Appretur	+0.79
90 Binnenschifffahrt	+0.76	69 Musterzeichner, Ziseleure u.ä.	+0.77
89 See- und Küstenschifffahrt	+0.72	32 Häkelei, Stickerei, Spitzenfabrikation	+0.75
76 Handelsvermittlung	+0.67	52 Wäschekonfektion	+0.52
92 Sonstiges Verkehrsgewerbe	+0.59	...	
36 Sonstige Textilindustrie	+0.53	10 Schlosserei	-0.55
28 Sonstige Industrie der Fette, Öle u.ä.	+0.52	83 Eisenbahnbetrieb	-0.63
71 Waren- und Produktenhandel	+0.50		
...			

begangene Weg einer sog. Distanzgruppierung im Faktorenumraum²⁴, wodurch eine Klassifizierung der Städte nach ihrer Ähnlichkeit hinsichtlich der sieben Faktoren vorgenommen werden könnte, wird hier nicht besprochen, da durch die eintretenden Faktorenumkombinationen die klare Faktorenumstruktur wieder weitgehend verwischt würde und erhebliche Interpretationsprobleme zu erwarten sind.

An dieser Stelle ist in einem kurzen Exkurs auf einige methodische Probleme der bisherigen Analyse einzugehen, da einige Fragen in der Literatur recht kontrovers behandelt werden.

1. Zur Frage der Normalverteilung. Da die Hauptkomponentenanalyse als analytisches Achsentransformationsverfahren in deskriptivem Sinne angewandt wurde, entfällt die Forderung nach normalverteilten Variablen²⁵. Aus diesem Grunde wird auf Transformationen der Variablen zur Beseitigung etwaiger Abweichungen verzichtet, zumal hierdurch kaum lösbare Interpretationsprobleme entstünden. Auch die Ableitung bzw. Anwendung des Korrelationskoeffizienten r erfordert keineswegs binormal verteilte Variablen, sofern dieser Wert nicht als Stichprobenschätzung, sondern als deskriptives Maß verwendet wird²⁶. Bei extremer Schiefe wird die Höhe des Korrelationskoeffizienten jedoch sehr durch Extremwerte beeinflusst, so daß er an Aussagekraft verliert. Zudem treten bei nicht normalverteilten Daten häufig nichtlineare Beziehungen auf, während die Faktorenumanalyse wegen ihrer Verankerung in der Korrelationsstatistik ausschließlich lineare Zusammenhänge zwischen Variablen erfassen kann. Dadurch muß mit einer tendenziellen Unterschätzung vorhandener Zusammenhänge gerechnet werden. Leider sind solche Verzerrungen kaum abschätzbar, da die hohe Variablenzahl keine lückenlose Überprüfung der bivariaten und multivariaten Verteilungsformen erlaubt. Stichprobenhafte Überprüfungen ergaben allerdings nicht in größerem Ausmaß extrem nichtlineare Beziehungen sowie krasse Abweichungen von der Normalverteilung. Wegen der hohen Variablenzahl ist auch kaum damit zu rechnen, daß etwaige Verzerrungen dieser Art das Ergebnis nennenswert beeinflussen.

2. Zur Signifikanz der Ergebnisse. Da es sich im vorliegenden Fall nicht um Stichprobenwerte, sondern um die Untersuchung einer Grundgesamtheit handelt, ent-

²⁴ Beispielhaft seien genannt: Ahmad, *Indian Cities*. — Kilchenmann, André und Ernst Gächter, *Neuere Anwendungsbeispiele von quantitativen Methoden, Computer und Plotter in der Geographie und Kartographie*, in: *Geographica Helvetica*, (1969), S. 68–81. — Bähr, Jürgen, *Gemeindetypisierung mit Hilfe quantitativer statistischer Verfahren*, in: *Erdkunde*, 25 (1971), S. 249–264. — Bartels, Dietrich und Wolf Gaebe, *Abgrenzung der Agglomerationen Rhein-Main, Rhein-Neckar und Karlsruhe*, in: *Vorschläge zur Neugliederung des Bundesgebiets, Materialien*, o. O. 1972, S. 263–322. — Schmidt und Margraf, *Klassifikation von Zentren*.

²⁵ Vgl. auch: Janson, Carl-Gunnar, *Some Problems of Ecological Factor Analysis*. in: Dogan, Mattei und Stein Rokkan (eds.), *Quantitative Ecological Analysis in the Social Sciences*, Cambridge, Mass. 1969, S. 332; Daultrey, S., *Principal Components Analysis*, Norwich 1976, S. 41.

²⁶ Vgl. Nefzger, M.G. und James Drasgow, *The Needless Assumption of Normality of Pearson's r* , in: *American Psychologist*, 12 (1957), S. 623–625. — Goldfried, Marvin und James Drasgow, *A Normal Distribution for What?*, in: *Journal of General Psychology*, 70 (1964), S. 21–28.

fällt jegliche inferenzstatistische Fragestellung. Weder die Städteauswahl noch die Untersuchungsvariablen noch der gewählte Zeitschnitt können als Zufallsauswahl angesehen werden, so daß der Anwendung von Signifikanztests die Grundlage fehlt.

3. Zur Problematik der Relation Zahl der Variablen: Zahl der Untersuchungseinheiten. Teilweise wird die Meinung vertreten, die Zahl der Untersuchungsobjekte müsse erheblich größer, zumindest jedoch um 1 größer sein als die Zahl der Variablen. Abgesehen von inferenzstatistischen Begründungen wird hierfür argumentiert, das allgemeine Modell der Faktorenanalyse, nämlich die Beschreibung der Meßwerte durch Linearkombinationen der Faktoren mit Hilfe der Ladungen und Faktorwerte, verlange diese Voraussetzung aus mathematischen Gründen. Indessen gilt dies nur insofern, als nicht mehr Faktoren extrahiert werden können als Untersuchungsobjekte vorhanden sind. Im vorliegenden Fall können die Variablen-Meßwerte bereits durch Linearkombinationen von 36 Faktoren vollständig reproduziert werden, da zusätzlich noch eine Dimension durch die Standardisierung (Subtraktion der Variablenmittelwerte) verloren geht. Diese Reduzierung der maximal extrahierbaren Faktorenzahl ist allerdings ohne praktische Bedeutung, da ohnehin eine Reduktion der Dimensionen auf eine erheblich geringere Zahl als 36 angestrebt wird.

4. Zur Problematik Hauptkomponentenanalyse versus Faktorenanalyse im engeren Sinne. In dem vorliegenden Beitrag werden die Begriffe Hauptkomponentenanalyse und Faktorenanalyse weithin synonym verwandt, ungeachtet der Tatsache, daß zahlreiche Autoren wesentliche konzeptionelle Unterschiede zwischen der hier vorgestellten Hauptkomponentenanalyse und der Faktorenanalyse im engeren Sinne, d. h. der Faktorisierung einer Korrelations- oder Streuungsmatrix unter Berücksichtigung der Eigenvarianz der Variablen (häufig durch eine Schätzung der sog. *Kommunalitäten*)²⁷ sehen. Indessen sind diese Unterschiede nur in Bezug auf inferenzstatistische Aspekte sowie bei kleinen Datensätzen von praktischer Bedeutung. Vergleichsrechnungen ergaben infolge der hohen Variablenzahl weitgehend identische Ergebnisse mit für die Interpretation irrelevanten Unterschieden. Aus diesem Grunde wurde auch eine Rotation der Hauptkomponenten vorgenommen, die dadurch zwar ihre Eigenschaft der Varianzmaximierung verlieren, aber eine für die inhaltliche Fragestellung sinnvollere Position erhalten. Dies mag von der „reinen“ Theorie der Modelle her angreifbar sein²⁸, doch lassen sich auch gewichtige Gegenargumente bringen²⁹.

²⁷ Beispielsweise Hauptachsen-Faktorenanalyse, Maximum-Likelihood-Faktorenanalyse.

²⁸ Überla, *Faktorenanalyse*, S. 88. — Mather, P.M., *Varimax and Generality: A Comment*, in: *Area*, 3 (1971), S. 252–254. — Mather, P.M., *Varimax and Generality: A Second Comment*, in: *Area*, 4 (1972), S. 27–30. — Daultrey, *Principal Components Analysis*, S. 44.

²⁹ Davies, W.K.D., *Varimax and the Destruction of Generality: A Methodological Note*, in: *Area*, 3 (1971), S. 112–118. — Davies, W.K.D., *Varimax and Generality: A Reply*, in: *Area*, 3 (1971), S. 254–259. — Davies, W.K.D., *Varimax and Generality: A Second Reply*, in: *Area*, 4 (1972), S. 207–208. — Gaensslen, Hermann und Werner Schubö, *Einfache und komplexe statistische Analyse*, 2nd. ed., München 1976. — Revenstorf, *Lehrbuch*, S. 190 f., S. 319.

Diese wenigen methodisch-technischen Anmerkungen zur Begründung der hier vorgenommenen Analyseschritte müssen genügen. Sie geben bereits einen Einblick in die Problematik der Anwendung der Faktorenanalyse, denn diese ist nicht als *Methode* im strengen Sinn zu bezeichnen, die durch eine feste Folge von Arbeitsschritten zu einem eindeutigen Ergebnis führt, sondern sie umfaßt zum einen als Sammelbegriff ein ganzes Bündel verschiedener Modellansätze, deren nahezu einziges gemeinsames Merkmal die Achsentransformation ist, und zum andern besteht jede dieser faktorenanalytischen Varianten aus einer Folge von Analyseschritten, die wiederum in unterschiedlichster Form kombiniert werden können.

So ist auch hier zu fragen, ob mit dem bisher aufgezeigten *üblichen Weg* einer Hauptkomponentenanalyse ein den Daten und der Fragestellung adäquates Ergebnis erreicht wurde. Sind tatsächlich empirische Strukturen der historischen Wirklichkeit aufgedeckt und abgebildet worden, oder handelt es sich weitgehend um Methodenartefakte, die entweder auf die mangelhafte Datenqualität oder auf die angewandte Analysemethoden zurückzuführen sind?

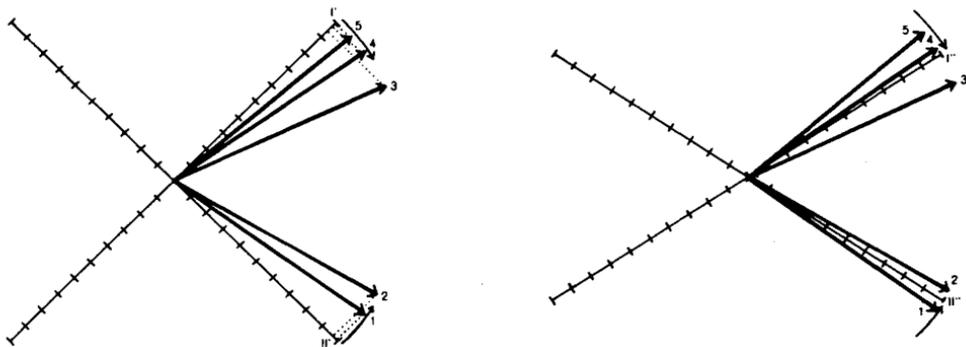
Im Interesse der übergeordneten methodologischen Fragestellung soll hier die Frage der Datenbasis ausgeklammert werden, obwohl hier in der historischen Stadtforschung häufig ein Hauptproblem liegen dürfte, und der Analysegang an einigen Stellen auf seine Folgerichtigkeit hin abgeklopft werden. Dabei wird hier nicht auf Einzelprobleme wie z. B. die Faktorenzahl, das Rotationskriterium oder die Kommunalitätsschätzung eingegangen, da hierdurch das Ergebnis nur marginal beeinflusst wird, sondern auf einige kritische Punkte, deren Berücksichtigung im Analyseablauf zu teilweise wesentlich anderen Ergebnissen führt.

Als erster Punkt sei das Orthogonalitätsproblem herausgegriffen. Wie oben gezeigt wurde, zielt die Faktorenanalyse auf eine Transformation zu unkorrelierten Faktoren, die graphisch durch senkrecht aufeinander stehende, also orthogonale Achsen symbolisiert werden können. Mit diesem Modell wurde vor allem in der Frühzeit der Faktorenanalyse die Vorstellung verbunden, die Faktoren könnten als Indikatoren für kausale Einflüsse angesehen werden (daher die Bezeichnung *Faktoren*), die unabhängig voneinander wirksam seien und deshalb als nicht korrelativ miteinander verknüpft angenommen werden könnten. Diese Vorstellung von wechselseitig unabhängigen, unkorrelierten Faktoren erscheint jedoch für die hier vorliegende Fragestellung ökonomischer Funktionen im Städtesystem höchst problematisch, denn wenn die hier ermittelten Hauptkomponenten als mögliche Indikatoren für städtische Wachstumsfaktoren interpretiert werden können, so ist von vornherein nicht damit zu rechnen, daß diese Faktoren voneinander unabhängig wirksam sind. Städtisches Wachstum ist zweckmäßiger als zeitlicher Prozeß zu verstehen, in dem zahlreiche Wechselwirkungen die Wachstumsfaktoren miteinander verknüpfen, so daß auch in der Standortverteilung der Funktionen auf das Städtesystem korrelative Zusammenhänge zu erwarten sind.

Diesem Einwand kann durch eine schiefwinklige Rotation zumindest teilweise Rechnung getragen werden. Eine solche Rotation des Faktorennachsensystems versucht ebenso wie eine orthogonale Rotation zur Einfachstruktur, das Achsensystem auf untereinander hoch korrelierte Variablenbündel auszurichten, doch wird dabei

die Orthogonalitätsbedingung aufgeben. Das Prinzip einer solchen schiefwinkligen Rotation kann wiederum an dem einfachen Beispiel, in dem fünf Variablen im Raum zweier Faktoren dargestellt werden, verdeutlicht werden (*Abb. 10*). Hier wird deutlich, daß durch die orthogonale Rotation zur Einfachstruktur nur eine grobe Annäherung der Faktorenachsen an die Variablenbündel erreicht werden kann. Erst ein Verzicht auf die Orthogonalität der Faktoren erlaubt, die Faktorenachsen mit großer Genauigkeit an den Variablenbündeln auszurichten, da diese selbst untereinander korrelieren.

Abb. 10 Schiefwinklige Rotation zur Einfachstruktur
Fiktives Beispiel mit zwei Faktoren und fünf Variablen (vgl. Abb. 5b)



Eine solche schiefwinklige Rotation wurde auch mit der bereits oben vorgestellten orthogonalen 7-Faktoren-Lösung vorgenommen. Da die wichtigsten Ergebnisse in *Abb. 11* tabellarisch zusammengefaßt sind, kann darauf verwiesen werden. U. a. zeigt sich, daß das entstehende Faktorenmuster nicht völlig von der orthogonalen Lösung abweicht, sondern daß das inhaltliche Profil der Faktoren nur leicht variiert wird. Dies spricht für die Stabilität der Lösung gegenüber Veränderungen des methodischen Instrumentariums. Der weitgehenden Übereinstimmung entsprechen verhältnismäßig geringe Korrelationen zwischen den Faktoren (*Abb. 11*).

Ein zweiter Kritikpunkt bezieht sich auf die anfangs vorgenommene Standardisierung der Prozentwerte zu normierten z -Werten, die der Berechnung der Korrelationsmatrix vorausging. Durch diese Standardisierung werden – wie an dem einführenden Zwei-Variablen-Beispiel aufgezeigt wurde – die Variablen *gleichgeschaltet*, indem die Mittelwert- und Streuungsunterschiede zwischen den Variablen durch Subtraktion bzw. Division ausgefiltert werden. In dem hier vorliegenden Datensatz erscheint jedoch eine solche Normierung der Variablen problematisch, da die Skalierung aller Variablen vergleichbar ist und deshalb Mittelwert- und Streuungsunterschiede wie zwischen *Handelsgewerbe* und *öffentlicher Verwaltung* durchaus empirisch bedeutsam sind. Eine Lösungsmöglichkeit dieses Problems besteht darin, auf

Abb. 11 Ergebnisse einer Hauptkomponentenanalyse der z-Werte nach schiefwinkliger Rotation

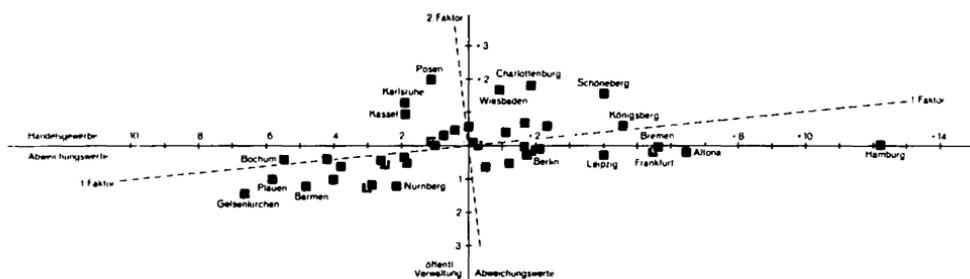
Methode: Hauptkomponentenanalyse der Korrelationsmatrix, (schiefwinklige) Direct-Quartimin-Rotation mit 7 Faktoren

Faktor I (bipolar): Tertiärer Sektor, insb. <u>private Dienstleistungen u. öffentl. Dienst;</u> <u>invers: Bergbau und Schwerindustrie</u> Varianzanteil: 10.41 % Höchste Faktorladungen der Variablen:		8 Klempner +0.71 45 Bäckerei, Konditorei, Lebküchler . . . +0.71 19 Mathemat., physik., chem. Instrumente +0.69 43 Tischlerei +0.55 ...																																																								
58 Wäscherei, Wäscherinnen +0.77 57 Friseure +0.71 24 Apotheker +0.62 99 Gesundheitspflege, Krankendienst . . +0.61 98 Bildung, Erziehung, Unterricht . . . +0.58 103 Rentner, Pensionäre +0.58 ... 5 Industrie der Steine und Erden . . . -0.57 61 Maurer -0.58 3 Kohlebergbau -0.61 2 Hüttenbetrieb -0.66 10 Schlosserei -0.70		Faktor V (bipolar): Tertiärer Sektor, insb. <u>Presse, Banken, Kultur;</u> <u>invers: Diversifizierte Leichtindustrie II,</u> <u>insb. Maschinen- u. Lederindustrie</u> Varianzanteil: 6.62 % 84 Posthaltere, Personenfuhrwerk . . . +0.61 70 Arbeiter, Gehilfen ohne nähere Zuordng. +0.59 100 Journalisten, Schriftsteller +0.59 74 Zeitungsverlag u. -spedition +0.58 93 Gast- und Schankwirtschaft +0.56 72 Geld- und Kredithandel +0.56 68 Maler und Bildhauer (Künstler) . . . +0.51 ... 55 Schuhmacherei -0.46 22 Sonstige Maschinenindustrie -0.50																																																								
Faktor II: Druck, Verlag, Musikinstrumente, <u>Pelze</u> Varianzanteil: 10.16 % 73 Buchhandel, Verlag +0.93 18 Sonstige Musikinstrumente +0.92 54 Kürschnerei, Pelzwarenzurichtung . . . +0.90 67 Sonstiges Vervielfältigungsgewerbe . . +0.87 64 Buchdruckerei +0.85 17 Pianoortefabrikation, Orgelbau +0.83 38 Buchbinderei, Kartonagenfabrikation . +0.79 80 Annoncervermittlung, Auskunftbüros . . +0.73 101 Stenographen, Privatsekretäre u.ä. . . +0.70 65 Stein- und Zinkdruckerei +0.58 ...		Faktor VI: Textilindustrie II, insb. Häkelei, <u>Stickerei, Spitzenfabrik., Musterzeichner</u> Varianzanteil: 6.00 % 34 Bleicherei, Appretur +0.80 69 Musterzeichner, Ziseleure u.ä. +0.78 32 Häkelei, Stickerei, Spitzenfabrikation +0.76 52 Wäschekonfektion +0.52 ... 83 Eisenbahnbetrieb -0.68																																																								
Faktor III: Hafenfunktionen Varianzanteil: 8.09 % 88 Reederei, Schiffsbefrachtung +0.93 77 Stauer, Schauerleute, Packer u.ä. . . . +0.85 78 Versteigerung, Verleihung, Aufbewahrg. +0.84 86 Spedition und Güterbeförderung . . . +0.79 90 Binnenschiffahrt +0.77 89 See- und Küstenschiffahrt +0.74 76 Handelsvermittlung +0.66 92 Sonstiges Verkehrsgewerbe +0.58 ...		Faktor VII (bipolar): Textilindustrie I, insb. <u>Weberei, Färberei, Posamentenfabrikation;</u> <u>invers: Militär, Schiffsbau</u> Varianzanteil: 5.69 % 30 Weberei +0.82 33 Färberei +0.73 56 Sonstiges Bekleidungsgewerbe +0.52 35 Posamentenfabrikation +0.51 ... 15 Schiffsbau -0.41 16 Uhrmacher -0.45 9 Grobschmiede -0.45 95 Militär -0.47 63 Sonstiges Baugewerbe -0.65																																																								
Faktor IV: Diversifizierte Leichtindustrie I, <u>insb. Metall-, Holz- und Elektroindustrie</u> Varianzanteil: 7.81 % 6 Gold- und Silberschmiede, Juweliere . +0.85 25 Sonstige chemische Industrie +0.81 14 Automobil- u. Fahrräderfabrikation . +0.80 21 Elektrotechniker, elektr. Maschinen . +0.77 44 Sonstige Holzindustrie +0.74 11 Sonstige Metallverarbeitung +0.74		Korrelationen zwischen den Faktoren: <table border="1"> <thead> <tr> <th></th> <th>I</th> <th>II</th> <th>III</th> <th>IV</th> <th>V</th> <th>VI</th> <th>VII</th> </tr> </thead> <tbody> <tr> <th>II</th> <td></td> <td>0.17</td> <td>1.00</td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <th>III</th> <td></td> <td>0.05</td> <td>0.05</td> <td>1.00</td> <td></td> <td></td> <td></td> </tr> <tr> <th>IV</th> <td></td> <td>0.07</td> <td>0.20</td> <td>0.03</td> <td>1.00</td> <td></td> <td></td> </tr> <tr> <th>V</th> <td></td> <td>0.12</td> <td>0.08</td> <td>0.09</td> <td>0.06</td> <td>1.00</td> <td></td> </tr> <tr> <th>VI</th> <td></td> <td>-0.16</td> <td>0.00</td> <td>-0.05</td> <td>-0.02</td> <td>-0.15</td> <td>1.00</td> </tr> <tr> <th>VII</th> <td></td> <td>-0.07</td> <td>0.05</td> <td>-0.01</td> <td>0.05</td> <td>-0.04</td> <td>0.07</td> </tr> </tbody> </table>		I	II	III	IV	V	VI	VII	II		0.17	1.00					III		0.05	0.05	1.00				IV		0.07	0.20	0.03	1.00			V		0.12	0.08	0.09	0.06	1.00		VI		-0.16	0.00	-0.05	-0.02	-0.15	1.00	VII		-0.07	0.05	-0.01	0.05	-0.04	0.07
	I	II	III	IV	V	VI	VII																																																			
II		0.17	1.00																																																							
III		0.05	0.05	1.00																																																						
IV		0.07	0.20	0.03	1.00																																																					
V		0.12	0.08	0.09	0.06	1.00																																																				
VI		-0.16	0.00	-0.05	-0.02	-0.15	1.00																																																			
VII		-0.07	0.05	-0.01	0.05	-0.04	0.07																																																			

die z -Standardisierung zu verzichten und von den (mittelwertzentrierten) Abweichungswerten auszugehen und damit anstelle der Korrelationsmatrix die Streuungsmatrix der Abweichungswerte, die sog. *Varianz-Kovarianz-Matrix* zu faktorisieren. Diese Matrix enthält in der Diagonale die Varianzen der Variablen und in den übrigen Feldern die Kovarianzen sämtlicher möglichen Variablenpaare. Die Faktorisierung dieser Matrix ist ebenso möglich wie die einer Korrelationsmatrix und dürfte prinzipiell angezeigt sein, wenn die Streuungsunterschiede der Variablen nicht meßtechnische Artefakte sind, sondern eine empirische Bedeutung haben, die im Falle einer z -Standardisierung verschenkt würde.

Betrachten wir das Ergebnis einer solchen Varianz-Kovarianz-Faktorisierung an unserem Zwei-Variablen-Beispiel (Abb. 12). Der erste Faktor verläuft nun nicht mehr in der Diagonale zwischen beiden Variablenachsen, wird also nicht mehr von

Abb. 12 Faktorisierung von zwei mittelwertzentrierten Variablen



beiden Variablen gleichmäßig *geladen*, sondern zeichnet den Punkteschwarm entsprechend der empirischen Streuung der Variablen ab: Da die höchste Streuung auf die Variable *Handelsgewerbe* entfällt, liegt der erste Faktor näher an der Abszisse, während die *kleinere Variable öffentliche Verwaltung* einen wesentlich geringeren Einfluß auf die Position der Hauptachse besitzt. Auch hier kann eine Varimax-Rotation angeschlossen werden, die jedoch nach eigenen Erfahrungen kaum das Faktorenmuster verändert, da die Faktoren zumeist an den am stärksten streuenden Variablen bzw. Variablenbündeln ausgerichtet bleiben³⁰. Dieser Effekt läßt sich andeutungsweise auch an unserem Zwei-Variablen-Beispiel erkennen; auf die Wiedergabe der entsprechenden Ergebnisse der Varianz-Kovarianz-Faktorisierung des großen Datensatzes muß jedoch aus Platzgründen verzichtet werden.

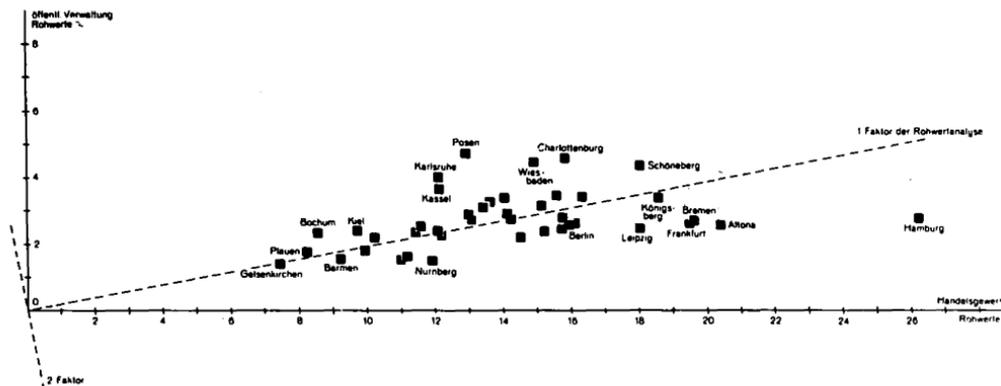
Statt dessen soll noch auf eine weitere Variante eingegangen werden. Die Varianz-Kovarianz-Matrix war anhand der mittelwertzentrierten Abweichungswerte gebildet

³⁰ Auch die von Holm, *Befragung 3*, S. 143 f. vorgeschlagene Reduzierung der Variablenvarianz in der Diagonale entsprechend einer Kommunalitätsschätzung hat bei größeren Variablenzahlen kaum Einfluß.

worden, doch es ist möglich, auch die Subtraktion der Variablenmittelwerte zu unterlassen und von den Rohwerten auszugehen. In diesem Fall wird analog zur Korrelations- bzw. Varianz-Kovarianz-Matrix die sog. *Kreuzproduktmatrix* faktorisiert, und die Faktorenanalyse wird zur sog. *Rohwertanalyse*³¹.

Das entsprechende Ergebnis für unser Zwei-Variablen-Beispiel ist wiederum in *Abb. 3* in numerischer sowie in *Abb. 13* in graphischer Form dargestellt: In der Rohwertanalyse wird vom Koordinatenursprung aus durch den ersten Faktor der

Abb. 13 Faktorisierung von zwei Rohwert-Variablen



entfernt liegende Punkteschwarm erfaßt, so daß alle Städte weitgehend unabhängig von ihrer typologischen Ausrichtung hohe Faktorwerte auf dieser ersten Hauptachse aufweisen. Wie sich auch bei der Rohwertanalyse des großen Datensatzes zeigte, auf deren Einzelergebnisse hier ebenfalls verzichtet werden muß, ist dieser erste Faktor als allgemeiner Generalfaktor zu interpretieren, der von den meisten Variablen hoch geladen wird, insbesondere von denen, die eine verhältnismäßig gleichmäßige Verteilung auf das Städtesystem aufweisen. Damit sind in erster Linie die Berufsarten des tertiären Sektors sowie einige marktorientierte Zweige des sekundären Sektors wie Bäcker, Metzger usw. angesprochen, während vor allem Bergbau und Schwerindustrie sowie die Textilindustrie durch ihre teilweise extremen räumlichen Konzentrationen auf wenige Städte hiervon abweichende Rohwertfaktoren bilden.

Ein solcher Faktor, der nahezu dem Spearman'schen Konzept eines Generalfaktors gleichkommt, hat Vor- und Nachteile. Seine inhaltliche Interpretation erscheint durchaus sinnvoll, da er einen großen Variablenkomplex anzeigt, der wohl am besten durch *Multifunktionalität* gekennzeichnet werden kann. Denn es ist mit Recht her-

³¹ Pawlik, Kurt, *Dimensionen des Verhaltens*, Bern 1968, S. 48 f. – Holm, *Befragung 3*, S. 138 ff.

vorgehoben worden³², daß die meisten Städte, vor allem die größeren, gerade durch eine solche Multifunktionalität charakterisiert werden, während demgegenüber den speziellen Stadtfunktionen zumeist nur eine akzidentielle Bedeutung zukommt. Dennoch erscheint ein solcher allgemeiner Faktor problematisch, denn er verdeckt durch seine zahlreichen hohen Ladungen die differenzierteren Funktionsunterschiede, da der größte Kreuzproduktanteil vieler Variablen bereits durch den ersten Faktor *gebunden* wird. Dieser Effekt ist im Ansatz der Rohwertanalyse angelegt, sofern wie hier ausschließlich Rohwerte mit positivem Vorzeichen in die Analyse eingehen.

Auf der Rohwertanalyse basiert jedoch ein weiterer, m. E. vorteilhafter Ansatz, der zudem einen Einwand berücksichtigt, den jüngst Schlosser gegen die Faktorenanalyse vorgebracht hat³³. Er weist mit Recht darauf hin, daß die bei der Faktorenanalyse üblicherweise vorgenommene *z*-Standardisierung stichprobenabhängig ist, so daß in unserem Fall lediglich über- und unterdurchschnittliche (und zwar nur in Bezug auf die untersuchten Großstädte) Besetzungen der Berufsarten in die Analyse eingehen. Ein m. E. besserer Weg wäre demgegenüber, nicht die Abweichungen vom Durchschnitt der Großstädte, sondern die Abweichungen von den Anteilen der Berufsarten im gesamten Deutschen Reich zugrunde zu legen. Diese unterschiedlichen Ansätze hängen mit dem Verständnis bzw. der Definition von *Stadtfunktion* zusammen, mit der Antwort auf die Frage: Funktion für was? Der bei der *y*- und *z*-Standardisierung vorgenommene Bezug auf die Mittelwerte der untersuchten Städte erscheint im Grunde nur dann sinnvoll, wenn als das *System*, innerhalb dessen die Funktion ausgeübt wird, die Gruppe der untersuchten Städte verstanden werden kann. Demgegenüber erscheint die Alternative sinnvoller, den Begriff *Funktion* auf die Gesamtbevölkerung des Deutschen Reichs in seiner damaligen politischen Abgrenzung zu beziehen³⁴. Einem solchen Verständnis wird m. E. nicht eine schematische *y*- oder *z*-Standardisierung, sondern eher die problemorientierte Standardisierung der Rohdaten durch Subtraktion der jeweiligen Berufsartenanteile im gesamten Deutschen Reich gerecht.

Mit diesem Problem ist zugleich eine Inkonsistenz in der bisherigen Literatur über Stadtfunktionen und funktionale Städtetypen angesprochen: Zumeist gehen solche Untersuchungen von über- bzw. unterproportionalen Besetzungen der Wirtschaftszweige nach Beschäftigungs- oder Erwerbstätigenzahlen aus und erfassen damit lediglich die (stichprobenabhängigen) über- bzw. unterdurchschnittlichen Ausprägungen von Stadtfunktionen. Schon vom Ansatz her neigen solche Untersuchungen – wie prinzipiell auch die oben wiedergegebenen Ergebnisse der von den Korrelationen ausgehenden Hauptkomponentenanalysen – zu einer Überschätzung und Überbetonung von lediglich akzidentiellen Funktionsunterschieden, während die möglicherweise für die meisten Städte gerade typische Funktionskomplexität durch die stichprobenbezogenen Mittelwertbereinigungen ausgefiltert wird.

32 Beispielsweise: Schöller, *deutsche Städte*, S. 15 f.

33 Schlosser, *Zusammenhangsanalyse*, S. 61 ff.

34 Dabei müssen allerdings grenzüberschreitende Verflechtungen ausgeklammert werden.

Als Konsequenz aus diesen Überlegungen wurde von allen Rohwerten (in %) der Anteil der jeweiligen Berufsart im Deutschen Reich subtrahiert, dann die Kreuzproduktematrix gebildet, diese faktorisiert sowie anschließend eine orthogonale Varimax-Rotation durchgeführt. Da mit 8 Faktoren bereits 98,3 % der (reichsmittelwertbezogenen) Abweichungsquadratsumme repräsentiert ist, wurde hier eine nach der Rotation inhaltlich recht klar profilierte 8-Faktoren-Lösung ausgewählt. Bei den in *Abb. 14* zusammengefaßten Ergebnissen ist zu beachten, daß die Faktorenladungen hier nur als Koeffizienten der Regression der Variablen auf die Faktoren zu verstehen sind und deshalb entsprechend dem *Gewicht* der Variablen unterschiedlich skaliert sind. Besonders stark kommt nun die Variable *Land- und Forstwirtschaft* zum Tragen, da diese Berufsart natürlich in allen Städten sehr stark *unterrepräsentiert* ist, während andererseits die Variablen *Warenhandel, Maschinenbau* und *Militär* in den Großstädten im Durchschnitt am stärksten *überrepräsentiert* sind.

Das Ergebnis zeigt 8 Faktoren, die alle als bipolare Land-Stadt-Dimensionen anzusprechen sind, aber jeweils unterschiedliche Komplexe ökonomischer Stadtfunktionen anzeigen. Wegen des hohen *Varianzanteils* verdient der erste Faktor besondere Aufmerksamkeit. Er repräsentiert eine großstädtische Grunddimension, einen ganzen Komplex städtischer Funktionen, die in den meisten Großstädten annähernd gleichmäßig ausgeprägt und gegenüber dem Reichsdurchschnitt deutlich *überrepräsentiert* sind. Dieser Faktor ist inhaltlich durchaus vergleichbar mit dem *Generalfaktor* der bereits oben erwähnten Rohwertanalyse, er bekommt jedoch erst durch die Zentrierung der Rohwerte auf die Reichsmittelwerte ein eindeutigeres Profil als großstädtische Grunddimension. Ubiquitär vorhandene Berufsarten wie z. B. zahlreiche, auch ländliche Handwerks- und Industriezweige (Schmiede, Wagner, Spinner, Maurer, Papier-, Leder-, Holz- und Tabakindustrie usw.) sind nun gegenüber diesem Hauptfaktor indifferent oder tragen gar – sofern sie in den Großstädten im Durchschnitt unterrepräsentiert sind – zusätzlich zur Landwirtschaft zur entgegengesetzten Ausrichtung des ersten Faktors bei. Die höchsten Ladungen in der anderen Richtung dieser Dimension finden wir vor allem im tertiären Sektor, sowohl in den Bereichen Handel und private Dienstleistungen wie auch im öffentlichen Dienst. Hier werden im großen und ganzen die zentralörtlich wirksamen Berufsarten angesprochen, und es zeigt sich, daß die Zentralität im Mittelpunkt der allgemeinen großstädtischen Grunddimension steht. Dennoch kann von einer reinen Zentralitätsdimension nicht gesprochen werden, denn offensichtlich zeigen auch einige Zweige des sekundären Sektors ähnliche Standortmuster im Städtesystem: an erster Stelle das Bekleidungs-gewerbe, auch Maschinenbau, Buchdruckerei sowie Teile des Baugewerbes und einige Handwerkszweige für den spezifischen Bedarf der Großstadtbevölkerung. Die enge Standortverknüpfung mit den zentralörtlich wirksamen Berufsarten kann prinzipiell auf zweierlei Weise erklärt werden. Zum einen kann die Nähe zum Absatzmarkt wirksam gewesen sein, zum andern bieten Städte mit hohem Anteil des tertiären Sektors und häufig überdurchschnittlicher Infrastrukturausstattung besonders günstige Standortbedingungen auch für die Ansiedlung und das Wachstum von Betrieben des sekundären Sektors, sofern andere Standortbindungen (Rohstoffe, Arbeitsmarkt mit niedrigem Lohnniveau) zurücktreten.

Abb. 14 Ergebnisse einer Faktorisierung der auf die Reichsmittelwerte zentrierten Rohwerte

Methode: Hauptkomponentenanalyse der Kreuzprodukte-Matrix der auf die Reichsmittelwerte zentrierten Rohwerte, orthogonale Rotation nach Varimax mit 8 Faktoren.

<p>Faktor I (bipolar): Großstädtische Grunddimension <u>(Zentralität u. standortverwandte Industrie);</u> <u>invers: Landwirtschaft</u> Anteil der Abweichungsquadratsumme: 65.63 % Höchste Faktorladungen der Variablen:</p> <p>71 Waren- und Produktehandel +5.01 51 Schneider, Kleiderkonfektion +1.97 12 Maschinen, Werkzeuge, Apparate +1.88 103 Rentner, Pensionäre +1.57 95 Militär +1.49 93 Gast- und Schankwirtschaft +1.47 94 Häusliche Dienste +1.45 96 Öffentliche Verwaltung, Rechtspflege +1.13 10 Schlosserei +0.79 82 Post- und Telegraphenbetrieb +0.74 83 Eisenbahnbetrieb +0.66 60 Bauunternehmung +0.65 64 Buchdruckerei +0.65 104 Studierende, Schüler +0.64 43 Tischlerei +0.58 58 Wäscherei, Wäscherinnen +0.57 81 Versicherungsgewerbe +0.51 30 Weberei -0.70 5 Industrie der Steine und Erden -1.27 3 Kohlebergbau -1.33 1 Landwirtschaft, Gärtnerei, Fischerei -27.20</p>	<p>Faktor IV (bipolar): Textilindustrie I, insb. <u>Häkelei, Stickerei, Spitzen, Bleicherei;</u> <u>invers: Landwirtschaft u. Kohlebergbau</u> Anteil der Abweichungsquadratsumme: 3.27 %</p> <p>32 Häkelei, Stickerei, Spitzenfabrikation +4.08 34 Bleicherei, Appretur +2.23 103 Rentner, Pensionäre -0.56 3 Kohlebergbau -1.67 1 Landwirtschaft, Gärtnerei, Fischerei -3.70</p>
<p>Faktor II (bipolar): Kohlebergbau; <u>invers: Landwirtschaft</u> Anteil der Abweichungsquadratsumme: 16.97 %</p> <p>3 Kohlebergbau +5.66 7 Eisengießerei u. Emaillierung v. Eisen +1.83 2 Hüttenbetrieb +1.62 95 Militär +0.94 10 Schlosserei +0.80 32 Häkelei, Stickerei, Spitzenfabrikation +0.72 71 Waren- und Produktehandel +0.69 60 Bauunternehmung +0.57 1 Landwirtschaft, Gärtnerei, Fischerei -12.80</p>	<p>Faktor V (bipolar): Militär, Schiffsbau; <u>invers: Kohlebergbau und Landwirtschaft</u> Anteil der Abweichungsquadratsumme: 3.03 %</p> <p>95 Militär +4.54 15 Schiffsbau +0.77 104 Studierende, Schüler +0.53 7 Eisengießerei -0.80 1 Landwirtschaft, Gärtnerei, Fischerei . -1.91 3 Kohlebergbau -3.18</p>
<p>Faktor III (bipolar): Maschinenindustrie; <u>invers: Landwirtschaft</u> Anteil der Abweichungsquadratsumme: 3.40 %</p> <p>12 Maschinen, Werkzeuge, Apparate +3.08 11 Sonstige Metallverarbeitung +0.54 31 Strickerei und Wirkerei +0.53 71 Waren- und Produktehandel +0.52 103 Rentner, Pensionäre -0.95 1 Landwirtschaft, Gärtnerei, Fischerei -5.33</p>	<p>Faktor VI (bipolar): Textilindustrie II, insb. <u>Weberei und Färberei;</u> <u>invers: Landwirtschaft und Kohlebergbau</u> Anteil der Abweichungsquadratsumme: 2.56 %</p> <p>30 Weberei +4.13 33 Färberei +0.66 95 Militär -0.61 3 Kohlebergbau -2.05 1 Landwirtschaft, Gärtnerei, Fischerei . -2.74</p>
	<p>Faktor VII (bipolar): Schwerindustrie I, insb. <u>Hüttenbetrieb; invers: Landwirtschaft</u> Anteil der Abweichungsquadratsumme: 2.15 %</p> <p>2 Hüttenbetrieb +3.09 71 Waren- und Produktehandel +0.64 12 Maschinen, Werkzeuge, Apparate +0.63 60 Bauunternehmung +0.52 95 Militär -0.73 3 Kohlebergbau -1.48 1 Landwirtschaft, Gärtnerei, Fischerei . -3.44</p>
	<p>Faktor VIII (bipolar): Schwerindustrie II, insb. <u>Eisengießerei; invers: Bergbau u. Landwirtschaft</u> Anteil der Abweichungsquadratsumme: 1.29 %</p> <p>7 Eisengießerei +3.14 1 Landwirtschaft, Gärtnerei, Fischerei . -1.55 3 Kohlebergbau -1.65</p>

Häufig werden auch diese beiden Faktoren der Marktorientierung und Agglomerationsvorteile gemeinsam wirksam gewesen sein, etwa wenn der erste Faktor auslösend und der zweite Faktor wachstumsverstärkend gewirkt hat, wie am Beispiel der Berliner Bekleidungsindustrie aufgezeigt werden könnte.

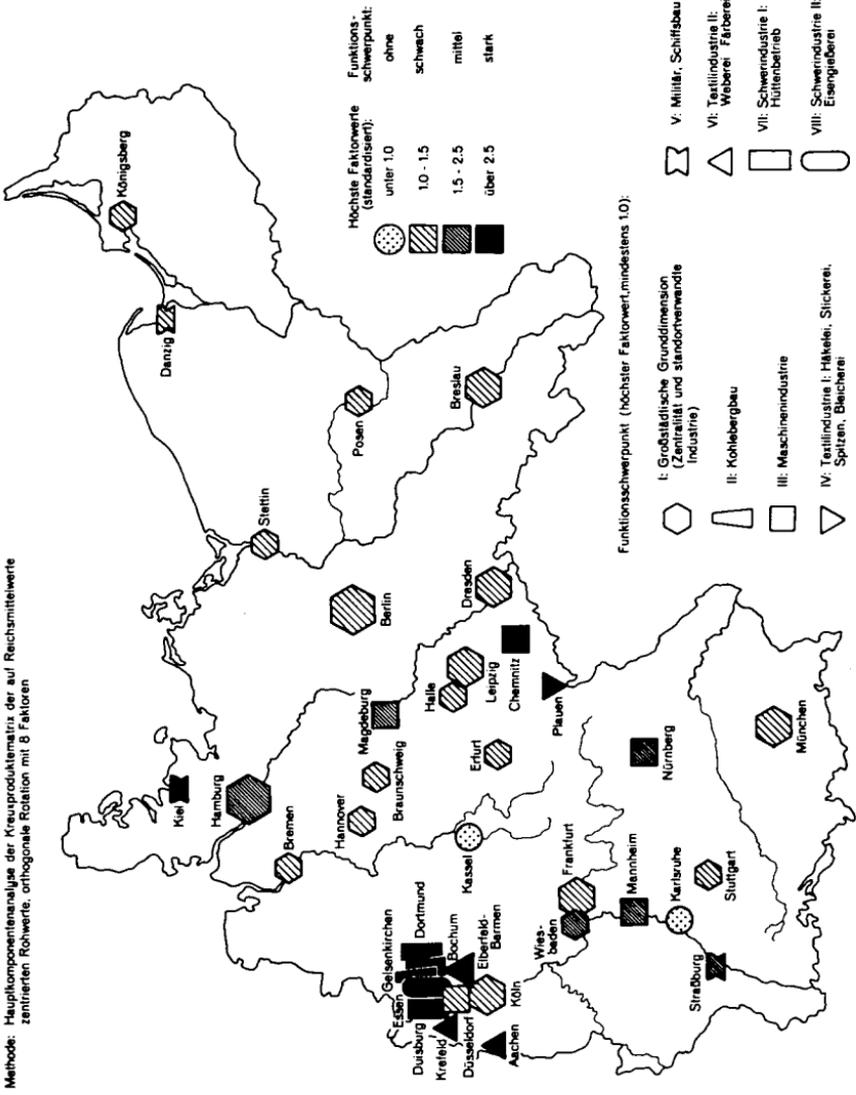
Auf die weiteren Faktoren braucht hier kaum näher eingegangen zu werden. Sie bezeichnen jeweils – zumeist in recht klarer inhaltlicher Profilierung – städtische Sonderfunktionen, die nach ihrem Standortmuster stark abweichen und die die Funktionsspezialisierung entweder einer oder einiger weniger Städte widerspiegeln. Dies wird auch deutlich an der Tabelle der Faktorwerte, auf deren vollständige Wiedergabe verzichtet werden muß, deren Aussagen zum Teil jedoch in *Abb. 15* kartographisch zusammengefaßt sind. Auch hier wurde aus den oben genannten Gründen auf eine Distanzgruppierung verzichtet und statt dessen eine Darstellung der jeweils höchsten Faktorwerte, die als Funktionsschwerpunkte interpretiert werden können, gewählt. Dadurch ergibt sich für die meisten Städte eine Zuordnung zum ersten Faktor, so daß die multifunktionale, durch Zentralität und standortverwandte Industriefunktionen geprägte Funktionsstruktur der meisten Großstädte aufgezeigt wird. Von diesen sind deutlich die Städte mit speziellen Funktionsschwerpunkten abgehoben.

Die kritische Überprüfung einiger Arbeitsschritte bei der Durchführung einer *üblichen* Hauptkomponentenanalyse ergab einige Punkte, deren Berücksichtigung zu Alternativlösungen und -ergebnissen führte, wobei hier nur ein kleiner Teil möglicher Analysevarianten angesprochen werden konnte. Dadurch sollte zugleich die hohe Flexibilität der unter dem Sammelbegriff *Faktorenanalyse* gefaßten Verfahren verdeutlicht werden, die im allgemeinen nicht zu einem eindeutigen Ergebnis führen, sondern an mehreren Stellen verfahrenstechnische Entscheidungen mit weitreichenden Auswirkungen auf das Ergebnis erfordern. Diese Flexibilität sollte nicht als Nachteil ausgelegt werden, sie erfordert jedoch vom Bearbeiter eine gründliche Beschäftigung mit dem umfangreichen Komplex faktorenanalytischer Verfahren, damit die den Daten und der Fragestellung adäquateste Analysevariante bestimmt werden kann. Vor einer routinemäßigen, unreflektierten Anwendung muß nachdrücklich gewarnt werden.

Wie sind nun die hier ermittelten Ergebnisse zu bewerten? Sowohl die *übliche* Faktorenanalyse wie auch die Faktorisierung der Abweichungswerte vom Reichsdurchschnitt erbrachten Ergebnisse, die – in Abhängigkeit von ihren jeweils unterschiedlichen Prämissen und Zielsetzungen – durchaus sinnvolle Aufschlüsse über die ökonomische Funktionsstruktur im System der deutschen Großstädte zur Hochindustrialisierungsperiode leisten. Während die traditionellen Klassifikationsmethoden a priori Entscheidungen über die als wesentlich erachteten Gliederungskriterien erfordern, kann deren Auswahl mit Hilfe der Faktorenanalyse wesentlich objektiviert werden. Auf der Basis der faktorenanalytisch gewonnenen funktionalen Grunddimensionen kann anschließend eine Typisierung der Städte nach ihrer Strukturähnlichkeit vorgenommen werden, beispielsweise mit Hilfe einer Distanzgruppierung oder eines anderen clusteranalytischen Verfahrens, doch kann darauf hier nicht näher eingegangen werden.

Abb. 15 Funktionsschwerpunkte der deutschen Großstädte anhand der Faktorwerte einer Rohwertanalyse

Methode: Hauptkomponentenanalyse der Kreuzproduktmatrix der auf Rechmittellwerte zentrierten Rohwerte, orthogonale Rotation mit 8 Faktoren



Sekundäre Funktionsschwerpunkte (hohe Faktorwerte auf einem zweiten Faktor) sind nicht dargestellt

Der Wert der Faktorenanalyse wäre jedoch recht begrenzt, wenn sich ihre Ergebnisse in der Erarbeitung von Klassifikationsgrundlagen erschöpften, zumal für Problemstellungen dieser Art clusteranalytische Verfahren zunehmend bevorzugt werden³⁵. Klassifikationssysteme sind zwar – sofern sie eine optimale Informationsverdichtung leisten – sehr nützlich und im Stadium eines nur gering entwickelten Theoriefundamentes sicherlich auch sehr wertvoll; doch sollten solche deskriptiven Schemata nicht als Selbstzweck angesehen werden. Wirklich fruchtbar werden die faktorenanalytische Ermittlung von *Grunddimensionen* und darauf aufbauenden Klassifizierungen erst, wenn sie Anknüpfungspunkte für weiterführende Arbeiten in Richtung auf eine Theorie der Stadtentwicklung bieten.

Mögliche Richtungen dieser Weiterarbeit können hier nur angedeutet und als potentielle Fragestellungen formuliert werden:

- Zeigen die verschiedenen Hauptfunktionen bzw. Städtetypen unterschiedliche räumliche Verteilungsmuster, etwa hinsichtlich regelhafter Streuung oder regionaler Konzentration?
- Weicht die Funktionsstruktur von Mittel- und Kleinstädten von der hier untersuchten der Großstädte ab?
- Wie verhalten sich die hier ermittelten Städtetypen zu den Variablen städtischen Bevölkerungs- und Wirtschaftswachstums?
- Bestehen – auch räumliche – Zusammenhänge zwischen städtischen Funktionen und wachstumsauslösenden Innovationen?
- Auf welche Weise korrespondiert die ökonomische Funktion einer Stadt mit der Sozialstruktur ihrer Bevölkerung und mit ihrer kommunalpolitischen Entwicklung?

Zweifellos sind noch weitere Aspekte der städtischen Sozial- und Wirtschaftsentwicklung denkbar, deren Zusammenhänge mit den hier ermittelten oder anderweitig faktorenanalytisch bestimmaren Kategorien Gegenstand weiterführender Hypothesenbildung sein könnten. Auch eine Erweiterung zu komparativ-statischen oder gar Längsschnittuntersuchungen erscheint wünschenswert, selbst wenn nur in deskriptiver Form die Veränderungen im System der Funktionsdimensionen über eine längere Zeitspanne erfaßt werden könnten.

In dieser Forschungsperspektive vermag, wie die Erfahrungen anderer Wissenschaften gezeigt haben, die Faktorenanalyse neben anderen Instrumenten der multi-

³⁵ Forst, Hans Theo, *Zur Klassifizierung von Städten nach wirtschafts- und sozialstatistischen Strukturmerkmalen*, Würzburg 1974. – Steinhausen, Detlef und Klaus Langer, *Clusteranalyse*, Berlin 1977. – Sitterberg, Georg, *Multivariate Analyse der Struktur und Entwicklung von Städten*, Münster 1977. – Für eine Anwendung der Clusteranalyse auf den hier vorliegenden Datensatz sowie einen Vergleich der Ergebnisse mit traditionellen Typisierungsmethoden vgl.: Blotvogel, Hans Heinrich, *Methodische Probleme der Erfassung städtischer Funktionen und funktionaler Städtetypen anhand quantitativer Analysen der Berufsstatistik 1907*, in: Ehbrecht, Wilfried (ed.), *Voraussetzungen und Methoden geschichtlicher Städteforschung*, Köln 1979.

variante Statistik wertvolle Hilfen zu geben. Der häufig erhobene Einwand, die spezifische Quellen- bzw. Datensituation des Historikers erlaube kaum die Anwendung solch aufwendiger Analyseverfahren, erscheint nur teilweise berechtigt. In Anbetracht der unübersehbaren, bisher kaum ausgewerteten statistischen Quellenbestände des 19. und 20. Jahrhunderts kann er ohnehin kaum für die letzten zwei Jahrhunderte gelten, und darüber hinaus ist darauf aufmerksam zu machen, daß der Einsatz mathematisch-statistischer Verfahren keineswegs auf die Verarbeitung *quantitativer*, d. h. auf metrischem Skalenniveau gegebener Daten beschränkt ist, wie aus Unkenntnis häufig angenommen wird. Die hier vorgestellte Faktorenanalyse gehört sicherlich zu den komplizierteren und anspruchsvolleren Analyseinstrumenten, deren Anwendungsfeld wohl immer auf eng begrenzte Bereiche der Stadt- und Stadtgeschichtsforschung begrenzt bleiben wird. Der hier unternommene Versuch diente dazu, auf einige Anwendungsmöglichkeiten und -probleme dieses in der deutschsprachigen Stadtgeschichtsforschung noch wenig bekannten Analyseinstruments aufmerksam zu machen³⁶.

Summary: The Economic Structure of German Cities in 1907: An Analysis of Census of Occupation Data by Factor Analysis

Today factor analysis belongs to the mostly used mathematical-statistical instruments of analysis in numerous sciences as for example in psychology, sociology, economy, and geography. Within the research of urban history it has been applied relatively seldom, although factor analytical problems as especially the reduction and concentration of large sets of data as well as the formulation of hypotheses about correlative connections often appear, particularly in the urban history of the 19th and 20th century. The aim of this contribution is a – necessarily simplified – representation of the basic steps of factor analysis excluding the mathematical foundation as well as a comprehensive discussion of different variants of factor analysis.

The subject, by which the method is exemplified, refers to the paradigm of economic-functional town classification often discussed in urban geography. While

³⁶ Der vorliegende Beitrag entstand im Rahmen eines Forschungsprojektes des Sonderforschungsbereichs 164 *Vergleichende geschichtliche Städteforschung*, Münster. Der Deutschen Forschungsgemeinschaft wird an dieser Stelle für die gewährte Unterstützung gedankt. Die Berechnungen wurden durchgeführt am Rechenzentrum der Ruhr-Universität Bochum unter Verwendung des dort implementierten Faktorenanalyseprogramms *AGFAP* (*Alberta General Factor Analysis Program*), des Faktorenanalyseprogramms aus dem *BMDP*-Paket sowie eines Programms zur Faktorisierung von Zusammenhangsmatrizen von N. de Lange. Für die Unterstützung bei der Durchführung der Berechnungen danke ich Herrn Wiss. Ass. N. de Lange, Institut für Geographie Münster, sowie Herrn cand. geogr. M. Kern und Herrn cand. geogr. D. Kalcevic, beide Bochum.

traditional methods of classification require the selection of only a few attributes which a priori are assumed to be essential, factor analysis can take a great amount of variates into account. It combines statistically correlating attributes to reduce a set of many variates to a few basic uncorrelated factorial dimensions.

The analysis starts from the percentages of 105 different occupations within the 37 cities (over 100 000 inhabitants) of the German Empire in 1907 (matrix 37×105). First of all a "usual" principal components analysis is computed. On behalf of this a standardization to z-values is made, and all paired correlations between the occupations are computed (matrix 105×105). From this matrix 7 factors are extracted with a clear profile after they have been rotated to simple structure. Each of these uncorrelated factors represents a bundle of occupations with a similar spatial distribution within the urban system and can be seen as one of the main factorial dimensions of the economic functional structure within the system of German cities about 1907. According to the position of cities on these dimensions (= factor scores) the main functions of the cities can be evaluated and displayed on maps.

A critical examination of this "usual" principal components analysis points out some faults which can largely be eliminated by alternative variants of analysis. At several points factor analysis requires from the revisor decisions concerning the technical procedure which in turn have far-reaching effects on the result. Thus it has to be warned of a recipe-like application.

From these possible variants only some are to be discussed here. First an oblique rotation of the principal components is taken, because of theoretical reasons correlations between the basic factorial dimensions can be assumed. The result reveals a better adjustment of the factors to the bundles of variates, but the correlations between the factors still remain very low.

As a next step it is proved which alterations appear in the factor pattern in case the z-standardization of the raw data is omitted and the analysis starts from mean-centered raw data (deviation scores). In this case the variance-covariance-matrix is factorized instead of the correlation-matrix. As another variant the mean-centralization is omitted and the resultative cross-products-matrix is factorized.

Whereas these variants can only be mentioned, finally the result of another raw-data-analysis is presented: The percentages of the occupations in the whole German Empire are subtracted from all raw data of the 105 occupations. Thus urban functions are not primarily referred to (often only accidental) relative differences between the cities but to the whole Empire. Again the cross-products-matrix is formed (105×105), subjected to an principal components analysis, and finally a solution of eight (orthogonally rotated) factors is chosen. The resultative factors represent all different structural rural-urban dimensions. The first factor is of particular interest because it can be seen as a complex basic urban dimension which alone represents 65.6 % of the sum of deviation squares. Thereof some special functions with specific locational distribution patterns are separated, as for example coal-mining, textile industry, engineering industry.

With respect to the method the high flexibility of factor analysis is pointed out, which — depending on aims and assumptions — shows quite different results. The

ascertained basic factorial dimensions of the economic functional structure of the urban system can form a starting-point for further research as for example about connections to variates of urban, population and economic growth as well as to social and municipal-political characteristics. On the way to a theory of historical urban development factor analysis along with other methods of multivariate statistics seems to be a valuable instrument of analysis to structuralize large data-sets and to formulate hypotheses. Its possible applications in the field of urban history seem not yet to be exhausted.