

Neue Methoden der Analyse historischer Daten

Best, Heinrich (Ed.); Thome, Helmut (Ed.)

Veröffentlichungsversion / Published Version

Sammelwerk / collection

Empfohlene Zitierung / Suggested Citation:

Best, H., & Thome, H. (Hrsg.). (1991). *Neue Methoden der Analyse historischer Daten* (Historisch-Sozialwissenschaftliche Forschungen : quantitative sozialwissenschaftliche Analysen von historischen und prozeßproduzierten Daten, 23). Sankt Katharinen: Scripta Mercaturae Verl. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-338007>

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.

HSF

HISTORISCH - SOZIALWISSENSCHAFTLICHE FORSCHUNGEN

**Quantitative sozialwissenschaftliche Analysen
von historischen und prozeß-produzierten Daten**

Herausgegeben von

**Heinrich Best, Wolfgang Bick, Paul J. Müller,
Herbert Reinke, Wilhelm H. Schröder**

**Gesellschaft für Historische Sozialforschung e. V.
in Verbindung mit dem
Zentrum für Historische Sozialforschung**

Band 23



HEINRICH BEST, HELMUT THOME
(HRSG.)

NEUE METHODEN DER ANALYSE
HISTORISCHER DATEN

SCRIPTA MERCATURAE VERLAG

Heinrich Best, Helmut Thome (Hrsg.)

Neue Methoden der Analyse historischer Daten

St. Katharinen 1991

Copyright: Scripta Mercaturae Verlag

D - 6551 St. Katharinen

ISBN 3 - 922661 - 84 - X

INHALT

Teil I: Formale Modelle als Grundlage multivariater Analysen und der Entdeckung latenter Strukturen

Ulrich Küsters

Analyse von qualitativen abhängigen Variablen 9

Heinrich Best und Joseph E. Zimmermann

Dimensionen politischer Konflikte. Die Analyse von namentlichen Abstimmungen in Parlamenten mit dem Verfahren der Mokken-Skalierung 41

Jörg Blasius

Intergenerationale und intragenerationale Mobilität von Parlamentariern der Frankfurter Nationalversammlung. Eine Anwendung der Korrespondenzanalyse 80

Rainer Schnell

Realisierung von Missing-Data-Ersetzungstechniken innerhalb statistischer Programmpakete und ihre Leistungsfähigkeit 105

Johannes Gordesch

Probleme formaler Modelle in den historischen Wissenschaften 138

Teil II: Analyse von Ereignisdaten und Zeitreihen

Hans-Jürgen Andreß, Heinrich Best und Kurt Sombert

Die Mandatsdauer im Deutschen Reichstag 1867 - 1918: Eine Anwendung neuer Methoden der Analyse historischer Verlaufsdaten 171

Winfried Stier

Grundlagen, Konzepte und neue Methoden der Zeitreihenanalyse 201

Volker Müller-Benedict

Lange Wellen im Bildungssystem 239

Rainer Metz und Helmut Thome

Verläuft der langfristige Wertewandel zyklisch? Eine zeitreihenanalytische
Überprüfung der Thesen von Namenwirth und Weber 269

Teil III: Computereinsatz in der Textanalyse

Cornelia Züll und Peter Ph. Mohler

Der General Inquirer III: Ein Dinosaurier mit Zukunft 303

Klaus Mudersbach

Erschließung historischer Texte mit Hilfe linguistischer Methoden 318

Hans-Christoph Hobohm

Historia ex machina. Der EDV-Einsatz in der Geschichtswissenschaft und
eine Erinnerung an seine Voraussetzungen 363

Autorenverzeichnis 376

Vorwort

Seit Anfang der 70er Jahre hat sich in der Bundesrepublik die "Historische Sozialforschung" (HSF) als Teildisziplin der Geschichtswissenschaft etabliert. Von anderen Teildisziplinen und Forschungsrichtungen innerhalb ihres Faches grenzt sie sich weniger durch einen speziellen Gegenstandsbereich (etwa der Sozialgeschichte) als durch eine bestimmte methodologische Orientierung ab. Entgegen der These von der Unvereinbarkeit geisteswissenschaftlicher und naturwissenschaftlicher Forschungslogik beruft sie sich auf die einheitswissenschaftlichen Prinzipien und den Wahrheitsbegriff der analytischen Wissenschaftstheorie, und sie benutzt theoretische Konzepte und empirische Methoden, wie sie vor allem in der Soziologie entwickelt oder dorthin aus noch anderen Disziplinen importiert worden sind. Ohne die Tradition der erzählerischen Rekonstruktion und der sinnstiftenden Deutung von Ereignisabläufen in Gänze zurückzuweisen, fordert sie, auch in der historischen Forschung auf überprüfbare, erklärungskräftige Theorien hinzuarbeiten - auf der Basis von Massendaten, die mit Hilfe formaler Modelle und statistischer Verfahren analysiert werden¹.

Mit diesem Anspruch befinden sich die Vertreter der Historischen Sozialforschung innerhalb ihres Faches in einer Minderheitsposition. Das läßt sich unschwer an den Lehrprogrammen geschichtswissenschaftlicher Institute an den Universitäten ablesen. Lehrveranstaltungen über statistische und andere formale Methoden (insbesondere Logik und Mathematik) sind nach wie vor Ausnahmen und gehören selten zum Pflichtteil eines geschichtswissenschaftlichen Curriculums.

Dem Zentrum für Historische Sozialforschung kommt deshalb mit seinen alljährlich (im "Herbstseminar") angebotenen "Grundkursen" zur empirischen Methodenlehre immer noch eine überregionale Ausbildungsfunktion zu, die eigentlich von den Universitäten übernommen werden müßte. Sinnvoller wäre es, das ZHSF könnte sich (neben seinen anderen Funktionen) verstärkt um die Vermittlung "fortgeschrittener" Methoden bemühen.

Die oft gepriesenen "einfachen" Methoden (wie prozentuierte Kontingenztabellen, bivariate Korrelations- und Regressionsrechnung) reichen auch

¹ Siehe die ausführlichere Darstellung dieser Konzeption in Heinrich Best u. Wilhelm Heinz Schröder, Quantitative historische Sozialforschung, in: Christian Meier u. Jörn Rüsen (Hrsg.), Theorie der Geschichte. Beiträge zur Historik, Band 5, München: dtv, 1988, S. 235-266.

in der historischen Forschungspraxis nicht aus. Sie werden der üblichen Komplexität theoretischer Konzepte und den unterschiedlichen Niveaus verfügbarer oder erreichbarer Daten nicht gerecht; sie schöpfen den informativen Gehalt der Daten nicht optimal aus.

Erklärende Theorien sind im Kern Aussagen über kausal interpretierte, unterschiedlich stark (oft auch zeitspezifisch) gewichtete Relationen von qualitativen oder/und metrischen Variablen. Wegen des möglichen Einflusses von "Drittvariablen" sind bivariate Assoziationsmaße und Korrelationskoeffizienten äußerst schlechte Indikatoren für kausale Beziehungen. Um die strukturellen Beziehungen zwischen mehreren Größen aufdecken oder entsprechende Hypothesen testen zu können, bedarf es multivariater Analysemethoden, mit denen man eine Vielzahl von Variablen unter Berücksichtigung von (zufälligen) Fehlereinflüssen bzw. Störgrößen simultan betrachten kann. Eine Gruppe von Modellen, die speziell auf die Analyse qualitativer abhängiger Variablen zugeschnitten sind, wird in dem Beitrag von Ulrich Küsters vorgestellt. Dabei handelt es sich um Regressionsmodelle, mit denen sich mehrdimensionale Kontingenztabelle analysieren lassen ².

Statistische Modelle definieren objektivierte Kriterien, nach denen unter explizit gemachten Voraussetzungen die Übereinstimmung oder Verträglichkeit zwischen theoretischen Annahmen und beobachteter Realität beurteilt werden kann. Anzuwenden sind sie nicht nur in der Kausalanalyse, sondern auch in anderen Feldern und Phasen des Forschungsprozesses. Dazu gehört die oft schwierige "Übersetzung" von theoretischen Begriffen (wie Industrialisierung, legitime Herrschaft, Klassenbewußtsein) in empirische (meßbare) Indikatoren. Bestimmte Aspekte dieser in der Forschungspraxis oft nur intuitiv geleisteten Zuordnung werden durch die verschiedenen Skalierungsverfahren objektivierbar (unter der Voraussetzung, daß für ein theoretisches Konstrukt mehrere Indikatoren erhoben worden sind). Dabei haben sich - wie auch in der Kausalanalyse - deterministische Modelle (wie die Guttman-Skalierung), die "Fehler" in den Beobachtungsgrößen oder Variablenbeziehungen nicht explizit berücksichtigen, in vielen Anwendungssituationen als zu restriktiv erwiesen. Heinrich Best und Josef E. Zimmermann stellen in ihrem Bei-

² Zwei weitere Veröffentlichungen des ZHSF können hierzu einführend oder ergänzend gelesen werden: Jürgen Sensch, Methoden der multivariaten Analyse nominal- und ordinalskalierten Daten auf der Basis mehrdimensionaler Kontingenztabelle, in: *Historical Social Research/Historische Sozialforschung* Nr. 42/43 (1987), S. 7-171; Helmut Thome, Grundkurs Statistik für Historiker, Teil II: Induktive Statistik und Regressionsanalyse. *Historical Social Research/Historische Sozialforschung*, Supplement No. 3 (1990), Kap. 11 und 12.

trag ein probabilistisches Skalierungsmodell ("Mokken-Skalierung") vor, mit dem sie das Abstimmungsverhalten der Abgeordneten und daran anknüpfend die Konfliktstrukturen der Frankfurter Nationalversammlung und der französischen Assemblée Nationale Constituante analysieren³.

Ebenfalls um die Herausarbeitung latenter, nicht direkt beobachtbarer Strukturen geht es in dem Aufsatz von Jörg Blasius über die inter- und intragenerationale Mobilität von Parlamentariern der Frankfurter Nationalversammlung. Je größer die Anzahl der Berufs- oder Statuskategorien, die in die Mobilitätstafeln eingehen, desto weniger gelingt es, die Muster von Konstanz und Bewegung, starken und schwachen Mobilitätsströmen allein durch "optische Interpretation" der entsprechenden Tabellen zu identifizieren. Blasius demonstriert, wie sich mit der Methode der Korrespondenzanalyse (als Alternative zur Multidimensionalen Skalierung) Mobilitätsdaten in fruchtbarer Weise auswerten lassen. Bei diesem datenreduktiven Verfahren handelt es sich um eine Art Hauptkomponentenanalyse aneinandergereihter Kreuztabellen, wobei die Zusammenhänge numerisch und graphisch dargestellt werden. Es ist auf qualitative (nominal skalierte) Daten anwendbar und hat gegenüber der Multidimensionalen Skalierung den Vorteil, daß die Originaldaten direkt in die Dimensionenanalyse eingehen, ohne daß sie (wie bei der MDS) zunächst in (Un)ähnlichkeitsmaße umgerechnet werden müssen.

Die Anwendung statistischer Modelle wird häufig durch fehlende Werte ("missing data" - MD) behindert; andererseits lassen sich nur auf der Basis von Modellen Konsequenzen fehlender Werte abschätzen und Strategien des kontrollierten Umgangs mit MD entwickeln. Rainer Schnell stellt in seiner Abhandlung eine Reihe derartiger "MD-Techniken" vor und berichtet über eine von ihm durchgeführte Simulationsstudie, in der er die Leistungsfähigkeit verschiedener Strategien testet. Als ein wichtiges Ergebnis seiner Untersuchung hebt er hervor, daß es kein allgemein "bestes" Verfahren gibt, sondern daß die Wahl einer optimalen Strategie stets die Kenntnis desjenigen Prozesses voraussetzt, der die fehlenden Werte erzeugt.

Die ersten vier Beiträge liefern reichhaltiges Anschauungsmaterial über den Gebrauch formaler/statistischer Modelle in der historischen Sozialforschung. Dieses "Denken in Modellen", das viele Historiker befremdet und zur Ablehnung reizt, wird in dem Beitrag von Johannes Gordesch in eine breitere forschungslogische Perspektive gesetzt. Er erläutert die verschiedenen Facetten und Anwendungsmöglichkeiten unterschiedlicher Typen formaler Model-

³ Zu Skalierungsverfahren siehe auch den einführenden Band von Johann Bacher, Einführung in die Logik der Skalierungsverfahren, in: Historical Social Research/Historische Sozialforschung, Vol. 15 (1990), S. 4-170.

le, weist auf Widersprüche und Probleme hin, argumentiert aber auch entschieden für den Nutzen der Modelltechnik in den verstehenden Geisteswissenschaften.

Von besonderem Interesse für Historiker dürften Modelle sein, in denen die Zeitdimension sozialen Geschehens explizit berücksichtigt wird. Hierfür stehen, je nach Datenstruktur und Fragestellung, verschiedene Modellklassen zur Verfügung. Eine Gruppe von Modellen firmiert unter verschiedenen Titeln wie "Event History Analysis", "Analyse von Ereignisdaten", "Regressionsmodelle für Übergangsraten". Im Kern geht es dabei um die Analyse von Zustandswechseln, die bei Individuen, Kollektiven oder Organisationen beobachtet werden. Beispiele sind berufliche und politische Karrieren oder die verschiedenen Entwicklungsstadien politischer Parteien und Regime. Mit Hilfe dieser Modelle können nicht nur zeitspezifische Übergangsraten für den Wechsel von einem "Zustand" zu einem anderen, sondern auch zeitspezifische Einflüsse externer Variablen auf die Übergangsraten geschätzt werden. Wie sich diese Modelle fruchtbar in der historischen Parlamentarismusforschung einsetzen lassen, demonstriert der Beitrag von Hans-Jürgen Andreß, Heinrich Best und Kurt Sombert. Trotz ihres suggestiven Namens sind Methoden der "Event History Analysis" in der historischen Sozialforschung (anders als in der Soziologie) bisher kaum angewandt worden. Da aber viele soziale Prozesse als Zustandswechsel konzipiert werden können, dürfte sich hier ein weites Anwendungsfeld erschließen lassen.

Obwohl in der historischen Forschung eine große Zahl von Zeitreihen gesammelt worden sind, hat sie statistische Methoden der Zeitreihenanalyse bisher nur selten eingesetzt. Außerhalb der Wirtschaftsgeschichte und der historischen Demographie beschränken sich die Analysen von Zeitreihendaten vorwiegend auf qualitative Interpretationen der graphisch dargestellten Verlaufsformen, auf grobe Trendbestimmungen und die Berechnung bivariater Korrelationskoeffizienten bestenfalls trendbereinigter Reihen. Diese Verfahren sind für die Analyse kausaler Zusammenhänge bei weitem nicht ausreichend. Das verwirrend reichhaltige und komplexe Angebot der Statistiker an weiterführenden zeitreihenanalytischen Methoden schreckt jedoch viele Historiker von den Mühen einer Rezeption ab. Der Beitrag von Winfried Stier erleichtert den Zugang zu diesen Methoden, indem er in einem vergleichenden Überblick grundlegende Modellannahmen und Verfahrensweisen der klassischen Komponentenerlegung, des Filter-Designs, der Box/Jenkins-Methode und der Spektralanalyse erläutert sowie Fragestellungen erhellte, die mit diesen Ansätzen jeweils zu bearbeiten sind.

Die kombinierte Anwendung unterschiedlicher zeitreihenanalytischer Verfahren in einem sozialhistorischen Forschungskontext demonstriert Volker Müller-Benedict in seiner Untersuchung über "Lange Wellen im Bildungssystem". Dabei legt er besonderes Gewicht darauf, den Aussagegehalt zeitreihenanalytischer Ergebnisse in den Gesamtzusammenhang einer sozialhistorischen Argumentationsweise einzubetten.

Nicht nur in der Wirtschaftsgeschichte oder der Bildungsökonomie, auch in anderen Forschungsgebieten werden Zyklentheorien diskutiert, so z. B. in Untersuchungen über den kulturellen Wandel. Hier haben die amerikanischen Sozialwissenschaftler Namenwirth und Weber mit ihren aus quantifizierenden Analysen abgeleiteten Thesen über den zyklischen Charakter des langfristigen Wertewandels international Aufmerksamkeit gefunden⁴. Rainer Metz und Helmut Thome diskutieren in ihrem Artikel methodische Probleme, die mit dem von Namenwirth und Weber gewählten Verfahren (Anpassen von trigonometrischen Funktionen an Zeitreihen) verbunden sind und überprüfen (mit überwiegend negativem Resultat) Teilergebnisse der Weberschen Analyse. Dabei wenden sie Methoden des Filter-Design-Ansatzes an.

Die Analysen von Namenwirth und Weber beruhen im wesentlichen auf Zeitreihen von Wertindikatoren, die sie aus den Texten von Partei- bzw. Wahlprogrammen und einer langen Serie von britischen Thronreden gewonnen haben. Das dabei benutzte computergestützte Verfahren quantitativer Inhaltsanalyse wird in dem Beitrag von Peter Ph. Mohler und Cornelia Züll näher beschrieben.

Auf der Basis theoretisch abgeleiteter Klassifikationsschemata und Kategoriensysteme ermöglicht das Programmsystem des "General Inquirer" Inferenzen von Texten auf soziale Wirklichkeiten. Zwar reichen die Entwicklungsarbeiten an diesem Verfahren bis zum Anfang der 60er Jahre zurück, doch ist erst seit 1989 eine stabile und gut dokumentierte Standardversion für IBM-Computer verfügbar. Damit wurde dieser "Dinosaurier" endlich dem Evolutionsdruck immer anspruchsvollerer Nutzererwartungen angepaßt. Eines der wichtigsten Anwendungsgebiete des "General Inquirer" sind Langzeituntersuchungen, in denen Wertereihen gewonnen werden, die bis zu 300 Jahre umfassen. Hier setzt aber auch die Kritik von Sprachwissenschaftlern und Historikern an, die einwenden, daß in den verfügbaren Diktionären der Gehalts- und Gestaltwandel der Sprache sowie die Veränderung der in den

⁴ Siehe J. Zvi Namenwirth u. Robert Philip Weber, *Dynamics of Culture*, Boston u.a.: Allen & Urwin, 1987.

Texten verschlüsselten Kognitions- und Wertesysteme nicht angemessen berücksichtigt werden.

Damit ist das "Verstehensproblem" des Historikers angesprochen, das Mudersbach durch eine formale und computerunterstützte Methode historio-graphisch-linguistischer Texterschließung angeht. In seinem Verfahren werden die Entscheidungen bei der Textbewertung bilanziert und ein Interpretationsergebnis berechnet. Subprogramme legen die Entscheidungspfade der Textinterpretationen mit ihren Verzweigungen aus, quantifizieren Gewichtung- und Präferenzentscheidungen und erzeugen eine graphische Veranschaulichung semantischer Netze. Mudersbachs Formalisierung des Verstehensprozesses zielt darauf, "Kriterien geisteswissenschaftlicher Wissenschaftlichkeit" anzuwenden, durch die individuengebundene Entscheidungen auf dem Weg zum Interpretationsergebnis und das Ergebnis selbst explizit dargestellt und für andere Experten intersubjektiv nachvollziehbar werden.

Auch der Beitrag von Hobohm stellt das Spannungsverhältnis zwischen der "Rigorousität der Maschinen" und den Unschärfen des historisch-philologischen Erkenntnisprozesses in den Mittelpunkt. Die geisteswissenschaftlichen Traditionen des Subjektivismus und Methodenpluralismus stehen nach Hobohm quer zu den Kodifikations- und Stringenzerfordernissen einer "automatisierten Argumentation" in Expertensystemen oder der formalen Modellierung komplexer Strukturen und Prozesse. Allerdings scheinen neuere Entwicklungen im Bereich der Mathematik und deren Adaptionen in der Informatik, wie etwa die Theorie der unscharfen Mengen, die Theorie der fraktalen Objekte sowie Katastrophen- und Chaostheorien eine Annäherung zwischen geisteswissenschaftlichen Orientierungen und dem naturwissenschaftlichen Paradigma bewirken zu können.

Die Beiträge in diesem Sammelband sind überwiegend aus Vorträgen hervorgegangen, die auf der Cologne Computer Conference 1988 gehalten wurden⁵. Das Hauptanliegen dieses Kongresses war es, die methodischen und methodologischen Konsequenzen informationstechnologischer Entwicklungen in den Geistes- und Sozialwissenschaften darzustellen. Besonders prägnant läßt sich dieser Zusammenhang im Bereich der statistischen Analyse- und Darstellungsverfahren erkennen. Entwicklungslinien und -fortschritte werden deutlich, wenn man den vorliegenden Sammelband mit der in der gleichen Reihe erschienenen Publikation einer ersten Konferenz über "The Uses of

⁵ Eine Auswahl von Beiträgen zu diesem Kongreß wurde veröffentlicht in: Heinrich Best u.a. (eds.), *Computers in the Humanities and the Social Sciences*, München u.a.: K.G. Saur, 1991.

Historical and Process Produced Data" vergleicht, die 1977 von der Arbeitsgemeinschaft QUANTUM und der amerikanischen Social Science History Association ebenfalls in Köln veranstaltet wurde⁶. Schon damals standen das Problem einer engeren Verzahnung zwischen theoretischen Annahmen und statistischen Verfahren sowie Fragen der Abbildungsqualität historisch-sozialwissenschaftlicher Forschungsdaten im Vordergrund. Die Entwicklung der vergangenen zehn Jahre zeigt deutlich eine verbesserte "Modelltreue" der Analyseverfahren und eine zunehmende "Sensibilität" von Datenbanksystemen in der Handhabung der Unschärfen und Komplexität historischer Daten.

Heinrich Best

Helmut Thome

⁶ Jerome M. Clubb u. Erwin K. Scheuch (eds.), *Historical Social Research* (Historische Sozialwissenschaftliche Forschungen, Bd. 6), Stuttgart: Klett, 1980.

TEIL I:

Formale Modelle als Grundlage multivariater Analysen und der Entdeckung latenter Strukturen

Analyse von qualitativen abhängigen Variablen

von Ulrich Küsters

1. Einführung und Übersicht

Dieser Artikel beschreibt die wichtigsten Regressionsmodelle zur Analyse von vermuteten Kausalbeziehungen zwischen nicht-metrischen abhängigen Variablen und unabhängigen Variablen beliebigen Meßniveaus. Regressionsmodelle für metrisch skalierte abhängige Variablen sind weitgehend bekannt. Leider erfüllen die meisten sozialwissenschaftlichen Datensätze nicht die notwendigen Annahmen des gewöhnlichen Regressionsmodells, da die abhängige Variable häufig ordinal oder nominal skaliert ist. Ein typisches Beispiel aus dem historischen Bereich ist die Analyse der Determinanten von nationalsozialistischen Aktivitäten von deutschen Lehrern im Dritten Reich in Abhängigkeit von sozioökonomischen Variablen wie Konfessionszugehörigkeit, Geschlecht, Schichtzugehörigkeit, Geburtskohorte etc. Bei einer derartigen Analyse, die etwa bei Jarausch und Arminger (1989) eingehend behandelt wird, ist die abhängige Variable nominalskaliert, da die verschiedenen Ausprägungen der abhängigen Variablen unterschiedliche NS-Organisationen wie NSDAP, SA, SS etc. repräsentieren, die nicht in einer natürlichen Weise geordnet werden können. Die unabhängigen Variablen weisen hingegen verschiedene Meßniveaus auf. Konfessionszugehörigkeit wird üblicherweise als nominalskalierte Variable aufgefaßt, während eine Variable wie die Größe des Herkunftsortes metrisch skaliert ist.

Wie dieses Beispiel zeigt, muß bei der statistischen Analyse nichtmetrischer Variablen beachtet werden, daß der Wertebereich der abhängigen Variablen eingeschränkt ist. Im klassischen Regressionsmodell wird hingegen unterstellt, daß die abhängige Variable alle Werte auf der reellen Zahlenachse annehmen kann. Daher müssen spezielle Modelle zur Analyse von nichtmetrischen Variablen herangezogen werden, die explizit den beschränkten Wertebereich der abhängigen Variablen berücksichtigen.

Im folgenden wird ein knapper Überblick über die wichtigsten Ansätze zur Analyse nichtmetrischer Variablen gegeben. Zunächst werden die Probleme des kleinsten Quadrateschätzers bei der Analyse von nichtmetrischen Variab-

len skizziert. Anschließend werden Konstruktionsprinzipien zum Aufbau nichtmetrischer Regressionsmodelle beschrieben. Diese Konstruktionsprinzipien basieren im wesentlichen auf der Anwendung des Zufallsnutzenmaximierungsprinzips für nominalskalierte Variablen und des Schwellenwertkonzepts für ordinale und zensierte Variablen. Innerhalb beider Modellklassen werden die wichtigsten Vertreter kurz angegeben. Ein weiterer Abschnitt skizziert einige Probleme bei der Spezifikation und der Schätzung nichtmetrischer Variablenmodelle. Dieser Abschnitt enthält auch einige Hinweise auf statistische Programmsysteme. Zum Abschluß wird ein historischer Datensatz von 743 Abgeordneten aus der Frankfurter Nationalversammlung von 1848/49 exemplarisch analysiert, um die Anwendung qualitativer Variablenmodelle in der empirischen Geschichtsforschung zu illustrieren.¹ Methoden zur Analyse von multivariaten qualitativen Variablen sowie zeitreihenanalytische Methoden werden in diesem Übersichtsartikel nicht behandelt. Detaillierte Beschreibungen der verschiedenen Methoden zur Analyse nichtmetrischer Variablen findet der Leser etwa in den Monographien von Amemiya (1985) und Madala (1983).

2. Probleme bei der Anwendung des gewöhnlichen Regressionsmodells auf nichtmetrische Variablen

Die Schätzung des gewöhnlichen Regressionsmodells

$$y_t = x_t \beta + \varepsilon_t \quad t = 1, \dots, n$$

basiert auf einer Zufallsstichprobe $\{y_t, x_t\}_{t=1, \dots, n}$ vom Umfang n , wobei y_t eine eindimensionale metrisch skalierte abhängige Variable repräsentiert, während x_t ein $1 \times k$ Vektor von unabhängigen Variablen ist, dessen Komponenten entweder metrisch oder dichotom (d.h. $\{0,1\}$) skaliert sind. t ist der Fallindex. Das lineare Regressionsmodell wird häufig in Matrixschreibweise durch $y = X\beta + \varepsilon$ dargestellt. In der Regel werden im klassischen

¹ An dieser Stelle möchte ich mich herzlich bei H. Best, Universität Köln, bedanken, der mir freundlicherweise eine Reihe von Datensätzen zur Verfügung stellte. Herrn R. Ponemereu, vom Zentralarchiv für empirische Sozialforschung, Abteilung Zentrum für historische Sozialforschung Köln, danke ich für das Zusammenstellen eines Datensatzes. G. Armingier (Bergische Universität Wuppertal) und H. Thome (Zentrum für Historische Sozialforschung Köln) gaben mir hilfreiche Kommentare zu einer Vorversion dieser Arbeit.

Regressionsmodell folgende Annahmen unterstellt:

1. $\{\varepsilon_t\}_{t=1, \dots, n}$ ist eine Folge von stochastisch unabhängigen und identisch verteilten Fehlertermen mit Erwartungswert $E(\varepsilon_t) = 0$ und endlicher Varianz $V(\varepsilon_t) = \sigma^2 > 0$.
2. Die Fehlerterme ε_t sind von den Regressoren x_t stochastisch unabhängig.
3. Die Designmatrix X hat vollen Spaltenrang. Damit werden lineare Abhängigkeiten zwischen den erklärenden Variablen ausgeschlossen.

Unterstellt man zusätzlich zu diesen fundamentalen Annahmen noch einige asymptotische Regularitätsbedingungen (siehe etwa Schmidt 1976), so folgt, daß der gewöhnliche kleinste Quadrateschätzer (KQS)

$$\hat{\beta} = (X^T X)^{-1} X^T y$$

asymptotisch konsistent und effizient innerhalb der Klasse der asymptotisch normalverteilten Schätzer für β ist. In einer endlichen Stichprobe ist der kleinste Quadrateschätzer $\hat{\beta}$ erwartungstreu für β .

Offensichtlich hat der kleinste Quadrateschätzer wünschenswerte Eigenschaften in Hinblick auf statistische Qualitätskriterien. Dennoch gibt es eine Reihe von Gründen, den kleinsten Quadrateschätzer für Regressionsmodelle mit nichtmetrischen abhängigen Variablen *nicht* zu verwenden. Zur Illustration der Problematik wird lediglich der Spezialfall einer binären abhängigen Variablen y_t mit den beiden Ausprägungen 0 und 1 betrachtet. Das Kernproblem des KQS liegt in der Tatsache, daß die KQS-Prädiktoren $\hat{y}_t = x_t \hat{\beta}$ nur selten einen der beiden logisch möglichen Werte $\{0, 1\}$ annehmen. Daher benötigt man ein Interpretationsmodell für die Regressionsgleichung

$$y_t = x_t \beta + \varepsilon_t, \quad t = 1, \dots, n$$

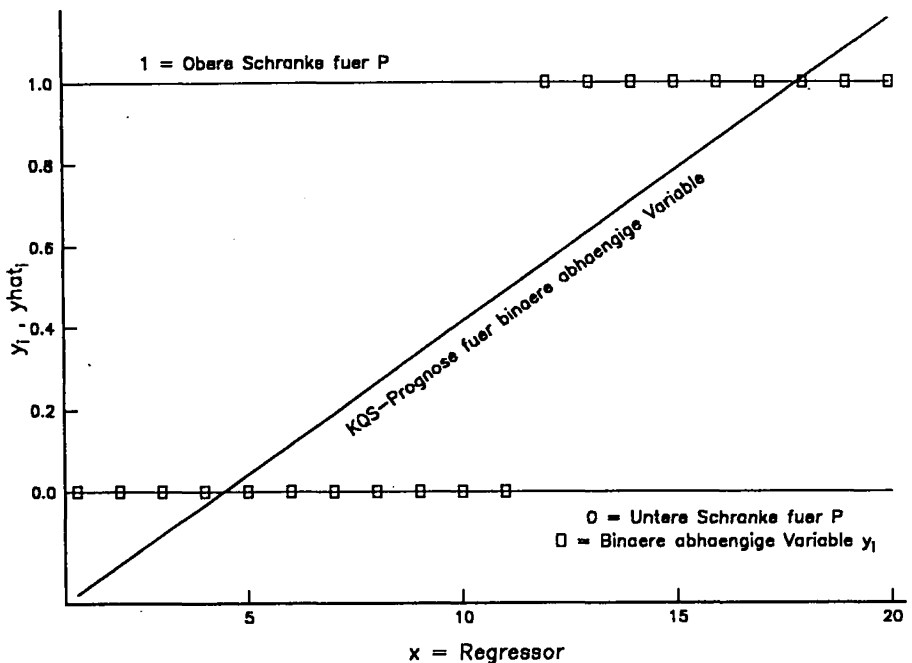
und das Prädiktormodell

$$\hat{y}_t = x_t \hat{\beta}, \quad t = 1, \dots, n$$

Ein einfaches Interpretationsmodell erhält man, indem man die prognostizierten Werte \hat{y}_t als die geschätzte konditionale Wahrscheinlichkeit der Ausprägung $y_t = 1$ bei gegebenem Regressor x_t auffaßt. Dieses Interpreta-

tionsmodell funktioniert jedoch nur dann, wenn der Regressionsvektor x_i in der Nähe der Mittelwerte der unabhängigen Variablen liegt. Für bestimmte Wertebereiche der unabhängigen Variablen treten jedoch Prädiktoren außerhalb des zulässigen Wertebereichs (0, 1) auf, wie man leicht anhand der Abbildung 1, in der ein einfaches Regressionsmodell mit einem Regressor eingezeichnet ist, ersehen kann. Konsequenterweise kann man daher das gewöhnliche Regressionsmodell nicht auf dichotome abhängige Variablen anwenden. Außerdem werden bei dichotomen abhängigen Variablen einige fundamentale Annahmen des linearen Regressionsmodells verletzt. Binäre Variablen sind Bernoulli-verteilt, so daß Heteroskedastizität auftritt, während im klassischen Regressionsmodell homoskedastische Fehlerterme unterstellt werden. Dieses Problem ist jedoch von untergeordneter Wichtigkeit, da man die Heteroskedastizität der Fehler durch verallgemeinerte kleinste Quadrateschätzer berücksichtigen kann. Das Kernproblem ist das Interpretationsmodell. Interpretationsmuster der Form *eine Erhöhung von x_i um eine Einheit verursacht eine Erhöhung der prognostizierten Wahrscheinlichkeit \hat{y}_i um $\hat{\beta}$ Einheiten* sind zwar recht einfach zu handhaben. Jedoch ist völlig unklar, wie prognostizierte *negative Wahrscheinlichkeiten* und *Wahrscheinlichkeiten größer als 1* interpretiert werden können.

Abbildung 1: KQS bei binären Daten



3. Konstruktionsprinzipien für Modelle zur Analyse nichtmetrischer Variablen

Fast alle Modelle zur Analyse von nicht-metrischen abhängigen Variablen werden auf der Basis der folgenden Prinzipien konstruiert:

1. Das Prädiktormodell wird grundsätzlich für Wahrscheinlichkeiten und nicht für Ereignisse formuliert.
2. Der Regressionsterm wird weiterhin linear durch $x_t\beta$ formuliert, so daß man ein einfaches lineares Interpretationsmuster verwenden kann. Damit muß man keine Parameterrestriktionen für β einführen.

Zur Erfüllung dieser Forderungen konzentriert man sich bei der Konstruktion auf Transformationsmodelle. Im folgenden wird der Spezialfall von dichotomen Variablen zuerst behandelt. Anschließend werden Modelle für nominalskalierte, ordinalskalierte und zensierte Variablen dargestellt.

3.1 Modelle zur Analyse von dichotomen Variablen

Zur Sicherung der Interpretation der prognostizierten Werte als Wahrscheinlichkeiten werden binäre Modelle aus zwei Komponenten zusammengesetzt:

1. Ein lineares Prädiktormodell der Form

$$\eta_t = x_t\beta, \quad t = 1, \dots, n$$

wird verwendet, um eine lineare Interpretation der Regressionskoeffizienten β_t bezüglich der latenten Prädiktorvariablen η_t zu ermöglichen.

2. Der lineare Prädiktor η_t wird durch eine nichtlineare Transformation mit einer auf das Intervall $(0, 1)$ beschränkten Wahrscheinlichkeit für die Ausprägung $y_t = 1$ verbunden, damit die prognostizierten Werte η_t als Indikatoren für steigende bzw. fallende Wahrscheinlichkeiten interpretiert werden können.

$$P(y_t = 1|x_t) = F(x_t\beta), \quad t = 1, \dots, n$$

Die Funktion $F: \mathbb{R} \rightarrow (0, 1)$ ist eine monoton steigende stetige Abbildung, die die Wahrscheinlichkeit P auf das Intervall $(0, 1)$ beschränkt. Somit ist F eine Verteilungsfunktion. Üblicherweise wird eine der beiden folgenden Verteilungsfunktionen verwendet:

1. Die kumulierte Standardnormalverteilung

$$F(z) = \Phi(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}x^2\right\} dx$$

ergibt das dichotome Probitmodell, das vorwiegend in der Psychometrie und Biometrie angewandt wird.

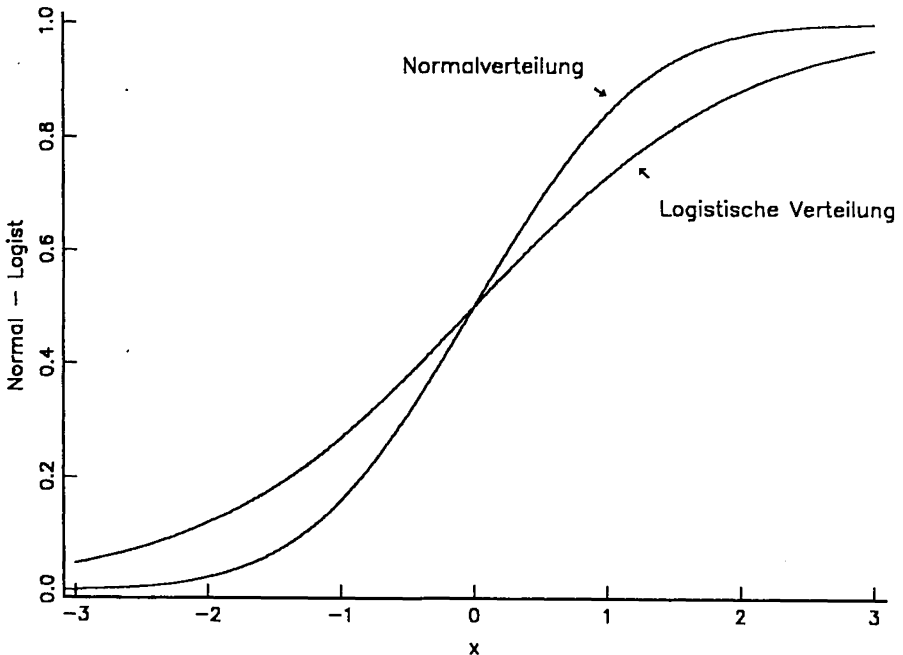
2. Die kumulierte logistische Funktion

$$F(z) = L(z) = \frac{1}{1 + \exp(-z)} = \frac{\exp(z)}{1 + \exp(z)}$$

ergibt das dichotome Logitmodell, das vorwiegend in der Soziologie und der Ökonometrie verwendet wird.

Beide Verteilungsfunktionen sind in Abbildung 2 eingetragen und stimmen mit Ausnahme der Ränder nahezu überein, sofern man die Dichte auf identische Varianzen skaliert. Daher erhält man bei praktischen Analysen bis auf einen Skalenfaktor nahezu identische Ergebnisse, so daß die Selektion zwischen beiden Verfahren im wesentlichen eine Frage der verfügbaren Software ist.

Abbildung 2: Normalverteilung – Logistische Verteilung



Bisher wurde davon ausgegangen, daß die unabhängigen Variablen vollständig metrisch skaliert sind. Daher wird in diesem Abschnitt skizziert, welche Modifikationen eingeführt werden müssen, wenn zusätzlich oder ausschließlich nominalskalierte Regressoren auftreten. Der einfachste Ansatz zur Berücksichtigung von nominalskalierten Regressoren ist die Verwendung von Designmatrizen, wie sie seit langem bei der metrischen Varianzanalyse verwendet werden. Zur Analyse einer metrischen abhängigen Variablen y_t in Abhängigkeit von einer ungeordneten qualitativen Variablen x_{t1} mit C Ausprägungen wird die Regressionsgleichung durch Mittelwerte strukturiert:

$$y_t = \mu + x_{t1}\alpha_1 + x_{t2}\alpha_2 + \dots + x_{tC}\alpha_C + \varepsilon_t, \quad t = 1, \dots, n$$

x_{ti} ist eine Dummy-Variable (Indikatorvariable) mit

$$x_{ti} = \begin{cases} 1 & \text{wenn die abhängige Variable } x_{t1} \text{ bei Fall } t \\ 0 & \text{sonst} \end{cases} \quad \left. \begin{array}{l} \\ \end{array} \right\} \begin{array}{l} \\ \text{Ausprägung } i \text{ annimmt} \end{array}$$

Damit repräsentiert der zusammengesetzte Term $\mu + \alpha_i$ den Mittelwert der abhängigen Variablen y_t unter der Bedingung, daß die unabhängige Variable x_{t1} die Ausprägung i annimmt. Allerdings enthält das Modell $C + 1$ Parameter, obwohl aus den Daten lediglich C konditionale Mittelwerte geschätzt werden können. Daher muß eine zusätzliche Parameterrestriktion eingeführt werden.

Meistens wird einer der beiden folgenden Ansätze² verwendet: Am einfachsten ist die Restriktion $\sum_{i=1, \dots, C} \alpha_i = 0$, so daß μ dem Mittelwert von y_t in der gesamten Stichprobe entspricht. Dieser erste Ansatz wird in der Literatur als zentrierte Reparametrisierung bezeichnet. Leider treten bei diesem Ansatz statistische Probleme auf, wenn Stichprobennullen oder strukturelle Nullen (siehe unten) auftreten. Der zweite Ansatz, der als Eckpunktparametrisierung bezeichnet wird, restringiert den Parameter der ersten Kategorie auf $\alpha_1 = 0$. Unter dieser Reparametrisierung korrespondiert der geschätzte Mittelwert μ zum Mittelwert von y_t innerhalb der Stichprobenpartition mit $x_{t1} = 1$, während die geschätzten Parameter α_i als Mittelwertverschiebungen gegenüber der ersten Kategorie interpretiert werden können. Beide Reparametrisierungsformen können einfach in das lineare Regressionsmodell eingebettet werden. Die Eckpunktparametrisierung wird in Regressionsschreibweise durch folgende Designmatrix repräsentiert:

² Komplexere Designmatrizen findet der Leser in Bock (1975).

$$\begin{pmatrix} \text{Prädiktor für Kategorie 1} \\ \text{Prädiktor für Kategorie 2} \\ \text{Prädiktor für Kategorie 3} \\ \text{Prädiktor für Kategorie 4} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix} \times \begin{pmatrix} \mu \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{pmatrix} = \begin{pmatrix} \mu \\ \mu + \alpha_2 \\ \mu + \alpha_3 \\ \mu + \alpha_4 \end{pmatrix}$$

Beide Ansätze können direkt auf den Fall mit mehreren unabhängigen Variablen verallgemeinert werden. Angenommen, zwei kategoriale Variablen x_{i1} und x_{i2} sollen zur Erklärung der Subgruppenmittelwerte einer metrischen abhängigen Variablen y_i herangezogen werden. Ein adäquates Prädiktormodell ist durch folgende Gleichung gegeben:

$$\hat{y}_i = \mu + \sum_{i=1}^{C_1} (\alpha_i x_{i1}) + \sum_{j=1}^{C_2} (\gamma_j x_{i2j}) + \sum_{i=1}^{C_1} \sum_{j=1}^{C_2} (\alpha \gamma_{ij} x_{i1i} x_{i2j})$$

Die Parameter α_i und γ_j repräsentieren die direkten Effekte der Dummy-Variablen x_{i1} bzw. x_{i2j} auf y_i , während die sogenannten Interaktionseffekte $\alpha \gamma_{ij}$ die zusätzlichen kombinierten Effekte beider Variablen messen, die nicht durch die direkten (additiven) Effekte berücksichtigt werden. Auch im multivariaten Fall müssen Parameterrestriktionen eingeführt werden, um die Schätzbarkeit der Parameter zu sichern. Bei einer Eckpunktparametrisierung wird üblicherweise die Identifikationsrestriktion $\alpha_i = \gamma_j = \alpha \gamma_{ij} = 0$ für alle Indizes mit $i = 1$ oder $j = 1$ verwendet. Somit kann man die varianzanalytische Designformulierung auch direkt für binäre Modelle verwenden, da die Struktur der linearen Prädiktorform $\eta_i = x_i \beta$ (bzw. $\eta = X\beta$) in binären Modellen mit der Prädiktorform der Varianzanalyse übereinstimmt.

3.2 Modelle zur Analyse von nominalskalierten abhängigen Variablen

Konzeptionell basieren die meisten Regressionsmodelle zur Analyse ungeordneter kategorialer Variablen auf einer Anwendung des Zufallsnutzenmaximierungsprinzips, bei dem jeder Ausprägung (Kategorie) $i \in \{1, \dots, C\}$ der abhängigen Variablen ein Zufallsnutzen³ U_{ii} zugeordnet wird, der durch

³ Das Zufallsnutzenmaximierungsprinzip wird in dieser Arbeit lediglich als mathematisches Konzept benutzt. Zur Modellierung von Regressionsmodellen mit qualitativen abhängigen Variablen kann man den Term U_{ii} genauso gut als eine latente Variable auffassen, die die relative Neigung zu Kategorie $y_i = i$ bei gegebenem Regressor x_i charakterisiert. Allerdings implizieren statistische Zufallsnutzenmaximierungsmodelle eine Reihe von nutzentheoretischen Eigenschaften, die in der Wahltheorie (McFadden 1981) ausführlich behandelt werden.

eine additive Kombination einer linearen Funktion $x_t \beta_i$ mit einem Fehlerterm ε_{it} parametrisiert wird.

$$U_{it} = x_t \beta_i + \varepsilon_{it}, \quad i = 1, \dots, C; \quad t = 1, \dots, n$$

Die Parameter β_i und die Fehlerterme ε_{it} stellen kategorienspezifische Größen dar. Nach dem Zufallsnutzenmaximierungsprinzip ist die konditionale Selektionswahrscheinlichkeit der Kategorie $y_t = i$ bei gegebenem x_t durch den Ausdruck

$$P(y_t = i | x_t) = P(U_{it} > U_{jt} \text{ für alle } j = 1, \dots, C \text{ mit } j \neq i)$$

definiert. Man beachte, daß die Zufallsnutzenterme U_{it} beliebige Werte auf der reellen Zahlenachse annehmen können. Dennoch sind alle Selektionswahrscheinlichkeiten auf das Intervall $(0, 1)$ beschränkt. In Analogie zu binären Logit- und Probitmodellen werden verschiedene Modelle durch unterschiedliche Verteilungsannahmen für ε_{it} generiert. Zwei typische Vertreter dieser Modellklasse sind:

1. Spezifiziert man die Verteilung des Fehlertermvektors $\varepsilon_t = (\varepsilon_{t1}, \dots, \varepsilon_{tC})^T$ durch eine C -dimensionale Normalverteilung mit Erwartungswert 0 und Varianz-Kovarianzmatrix Σ , so erhält man das multinomiale Probitmodell (Daganzo 1979).

2. Das multinomiale Logitmodell (McFadden 1974, Bock 1975) erhält man hingegen unter der Annahme, daß die einzelnen Komponenten ε_{it} des Fehlertermvektors ε_t stochastisch unabhängig und identisch verteilt sind und einer Extremwertverteilung vom Typ I mit der Verteilungsfunktion

$$F(z) = \exp(-\exp(-z))$$

folgen.

Der wesentliche Vorteil des multinomialen Probitmodells gegenüber dem multinomialen Logitmodell besteht in der weitaus größeren Flexibilität, da Ähnlichkeiten zwischen den verschiedenen Ausprägungen durch positive Korrelationsparameter in Σ berücksichtigt werden können. Ein wesentlicher Nachteil ist hingegen die Tatsache, daß multinomiale Probitmodelle numerisch nur sehr aufwendig für abhängige Variablen mit mehr als 4 Kategorien (= Ausprägungen) geschätzt werden können, da bei der Schätzung eines Modells mit C Kategorien $C - 1$ dimensionale Normalverteilungsintegrale für jeden Fall berechnet werden müssen. Daher wird das multinomiale Logitmodell üblicherweise vorgezogen, da die Schätzung bedeutend einfacher ist. Dies ist darauf zurückzuführen, daß die Selektionswahrscheinlichkeiten im multinomialen Logitmodell explizit durch den Ausdruck

$$P(y_i = i | x_i) = \frac{\exp(x_i \beta_i)}{1 + \sum_{l=2}^C \exp(x_i \beta_l)}, \quad i = 2, \dots, C$$

gegeben sind, während eine geschlossene Formel für die Selektionswahrscheinlichkeiten des multinomialen Probitmodells nicht zur Verfügung steht. Zur Sicherung der Parameterschätzbarkeit wird üblicherweise die Restriktion $\beta_1 = 0$ verwendet. Das multinomiale Logitmodell gehört zur Klasse der Exponentialfamilienmodelle. Somit ist die Log-Likelihood-Funktion konkav, so daß bei der numerischen Bestimmung des Maximum-Likelihood-Schätzers nur relativ wenige Probleme auftreten.

Individualdaten lassen sich ohne Informationsverlust zu Kontingenztabeln aggregieren, wenn alle Variablen nominalskaliert sind. Zur Analyse von Kontingenztabeln werden häufig loglineare Modelle (Bishop, Fienberg und Holland 1975, Jarausch, Arminger und Thaller 1985) herangezogen, mit denen die absoluten Häufigkeiten der einzelnen Zellen modelliert werden. Damit gehören loglineare Modelle in ihrer allgemeinsten Form zur Klasse der symmetrischen Modelle, da keine Variable explizit als abhängig ausgewiesen wird. Durch geeignete Parametrisierungsansätze kann man jedoch auch die Regressionsparameter β_i eines multinomialen Logitmodells indirekt durch die Schätzung eines loglinearen Modells bestimmen. Als Beispiel einer derartigen Parametrisierung wird eine dreidimensionale Kontingenztabelle mit drei Variablen $\{Z_1, Z_2, Z_3\}$ und den Zellenbesetzungen n_{ijk} betrachtet. Die Parametrisierung des loglinearen Modells zur exakten Reproduktion der gemeinsamen, dreidimensionalen Verteilung der Variablen $\{Z_1, Z_2, Z_3\}$ lautet

$$\ln(n_{ijk}) = \alpha_i + \gamma_j + \delta_k + \alpha\gamma_{ij} + \alpha\delta_{ik} + \gamma\delta_{jk} + \alpha\gamma\delta_{ijk}$$

Dabei sind $i = 1, \dots, C_{Z_1}$; $j = 1, \dots, C_{Z_2}$ und $k = 1, \dots, C_{Z_3}$ die Laufindizes der Ausprägungen der drei Variablen. In Analogie zur Varianzanalyse müssen die Effektparameter restringiert werden, um die Schätzbarkeit der Parameter zu sichern. Bei einer Eckpunktreparametrisierung werden üblicherweise die zu den ersten Kategorien korrespondierenden Effekte $\alpha_1, \gamma_1, \delta_1, \alpha\gamma_{ij}, \alpha\delta_{jk}, \gamma\delta_{ik}$ und $\alpha\gamma\delta_{ijk}$ für $i, j, k = 1$ auf Null restringiert. Ein Modell, mit dem die gemeinsame Verteilung aller Variablen exakt reproduziert werden kann, wird als saturiert (gesättigt) bezeichnet. Bei einer statistischen Modellierung werden i.d.R. einige Parameter auf Null restringiert, so daß das Modell die gemeinsame Verteilung nicht exakt reproduziert. Zur Erklärung der dreidimensionalen, gemeinsamen Verteilung der drei Variablen $\{Z_1, Z_2, Z_3\}$

durch die drei bivariaten Verteilungen $\{Z1, Z2\}$, $\{Z1, Z3\}$ und $\{Z2, Z3\}$ verwendet man das reduzierte Modell

$$\ln(n_{ijk}) = \alpha_i + \gamma_j + \delta_k + \alpha\gamma_{ij} + \alpha\delta_{ik} + \gamma\delta_{jk} + \varepsilon,$$

in dem alle Interaktionseffekte dritter Ordnung $\alpha\gamma\delta_{ijk}$ auf Null restringiert werden. Möchte man die dreidimensionale Verteilung durch die univariaten marginalen Verteilungen modellieren (Unabhängigkeitsmodell), so werden nur die Haupteffekte α_i , γ_j und δ_k zur Erklärung der gemeinsamen Häufigkeiten n_{ijk} geschätzt. Die bisher angegebenen Parametrisierungsformen von loglinearen Modellen weisen keiner Variablen eine besondere Bedeutung zu, so daß man diese Modelle in Analogie zur Korrelationsanalyse als reine Zusammenhangsmodelle interpretieren kann. Möchte man hingegen die Variable Z1 durch die beiden Variablen $\{Z2, Z3\}$ erklären, so kann man das loglineare Modell mit Haupteffekten und Interaktionseffekten erster Ordnung in folgender Weise als multinomiales Logitmodell mit den Regressoren $\{Z2, Z3\}$ auffassen:

$$P(Z1 = i | Z2 = j, Z3 = k) = \frac{n_{ijk}}{C_{Z1}} = \frac{\exp(\alpha_i + \gamma_j + \delta_k + \alpha\gamma_{ij} + \alpha\delta_{ik} + \gamma\delta_{jk})}{\sum_{l=1}^{C_{Z1}} \exp(\alpha_l + \gamma_j + \delta_k + \alpha\gamma_{lj} + \alpha\delta_{lk} + \gamma\delta_{jk})} = \frac{\exp(\alpha_i + \alpha\gamma_{ij} + \alpha\delta_{ik})}{\sum_{l=1}^{C_{Z1}} \exp(\alpha_l + \alpha\gamma_{lj} + \alpha\delta_{lk})}$$

Damit kann man die Parameter α_i , $\alpha\gamma_{ij}$ und $\alpha\delta_{ik}$ als Regressionsparameter $\beta_i = (\alpha_i, \alpha\gamma_{ij}, \alpha\delta_{ik})^T$ eines multinomialen Logitmodells auffassen, in dem die konditionalen Selektionswahrscheinlichkeiten der abhängigen Variablen Z1 durch die beiden unabhängigen kategorialen Variablen Z2 und Z3 erklärt werden. Somit können Regressionsparameter von multinomialen Logitmodellen grundsätzlich durch die Schätzung eines äquivalenten loglinearen

Modells bestimmt werden, sofern alle unabhängigen Variablen nominalskaliert sind⁴.

Wendet man das multinomiale Probitmodell und das multinomiale Logitmodell auf eine binäre abhängige Variable an, so erhält man die korrespondierenden binären Probit- bzw. Logitmodelle. Man beachte jedoch, daß sich beide Modelle im Gegensatz zum binären Fall *erheblich* unterscheiden, wenn die abhängige Variable mehr als zwei Kategorien besitzt. Heuristisch kann dieser Unterschied an folgendem Musterbeispiel erklärt werden: Angenommen, man möchte die Konsumentenwahl zwischen drei verschiedenen Transportmodi (z.B. Bus, Bahn und Auto) zum Arbeitsplatz durch eine Reihe von exogenen Regressoren erklären. Bus und Bahn gehören zur Klasse der öffentlichen Verkehrsmittel, die in der Regel durch gemeinsame oder ähnliche Attribute charakterisierbar sind, die nicht durch unabhängige Variablen gemessen werden. Der Transportkomfort ist ein klassisches Beispiel. Derartige Ähnlichkeiten zwischen Teilmengen von Alternativen stellen unbeobachtbare Heterogenitätsfaktoren dar, die in einem multinomialen Probitmodell relativ einfach durch eine positive Fehlertermkorrelation in Σ zwischen den Ausprägungen Bus und Bahn modelliert werden können. Innerhalb eines multinomialen Logitmodells kann ein derartiger Heterogenitätseffekt jedoch nicht berücksichtigt werden, da grundsätzlich unterstellt wird, daß die Fehlerterme stochastisch unabhängig sind. Aufgrund der erheblichen numerischen Probleme bei der Schätzung des multinomialen Probitmodells wurden in der Literatur eine Reihe von Modifikationen des multinomialen Logitmodells vorgeschlagen, die zumindest eine partielle Berücksichtigung von Ähnlichkeiten zwischen den verschiedenen Kategorien erlauben. Die wichtigsten Vertreter dieser Modellklasse sind die genesteten (= geschachtelten) Logitmodelle und die Tree-Extreme-Value-Modelle von McFadden (1981).

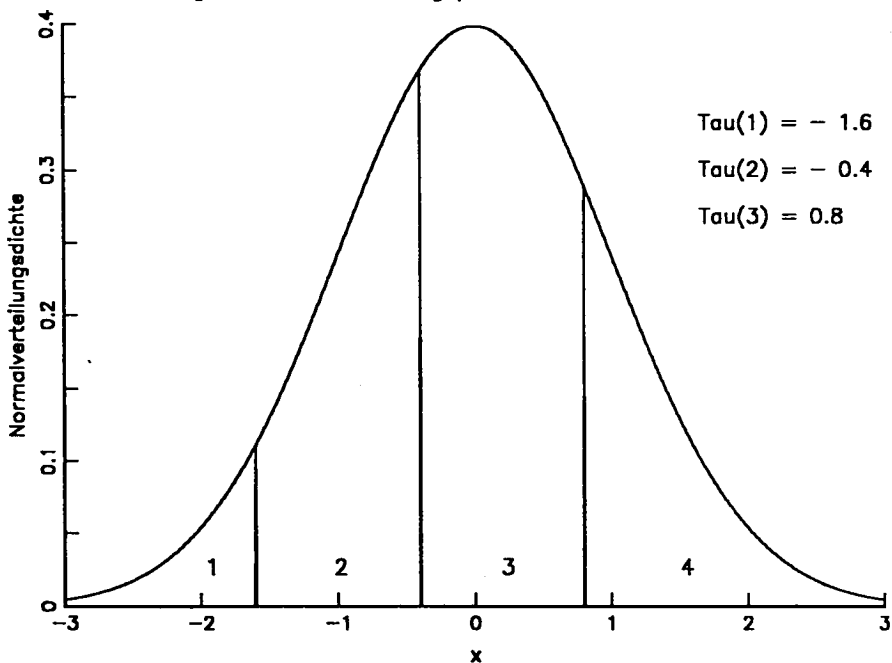
3.3 Modelle zur Analyse von ordinalskalierten Variablen

Zufallsnutzenmaximierungsmodelle zur Erklärung der Variation von ungeordneten qualitativen Variablen basieren auf der Tatsache, daß die einzelnen Ausprägungen der abhängigen Variablen *nicht* auf einer Rangskala angeordnet werden können. Bei praktischen Analysen treten jedoch relativ häufig abhängige Variablen auf, die in einer natürlichen Rangfolge angeord-

⁴ Der Satz von Birch (siehe Bishop, Fienberg und Holland 1975) liefert eine mathematische Grundlage für die Äquivalenz beider Ansätze.

net werden können. Ein typisches Beispiel ist die Messung von Einstellungen auf einer Likert-Skala. Eine spezielle Modellklasse, die eine derartige zusätzliche Ranginformation ausnutzt, verwendet das Schwellenwertkonzept. Schwellenwertmodelle basieren auf der grundlegenden Annahme, daß eine unbeobachtbare (latente) Variable η_i existiert, die alle Werte auf der reellen Zahlenachse annehmen kann. Diese latente Zufallsvariable wird durch ein lineares Regressionsmodell parametrisiert. Steigende Werte von η_i können als verstärkte relative Neigung von Kategorie $y_i = i$ im Verhältnis zu einer Referenzkategorie (etwa $y_i = 1$ oder $y_i = C$) interpretiert werden. Da die Ausprägungen geordnet sind, kann man lediglich die niedrigste oder die höchste Kategorie als Referenzkategorie wählen. Zur Verbindung der latenten Variablen η_i mit den diskreten Ausprägungen $y_i = 1, \dots, C$ wird die reelle Zahlenachse R durch $C - 1$ geordnete Schwellenwerte $\tau_1 < \tau_2 < \dots < \tau_{(C-1)}$ partitioniert, wie man aus Abbildung 3 ersehen kann.

Abbildung 3: Normalverteilung partitioniert durch Schwellenwerte



Die abhängige Variable y_t nimmt genau dann den Wert i an, wenn die latente Variable η_t in das halboffene Intervall (τ_{i-1}, τ_i) fällt. Bei festem Regressor x_t wird die stochastische Variation von η_t durch die stochastische Variation des Fehlerterms ε_t verursacht. Formal können Schwellenwertmodelle durch folgende Gleichungen repräsentiert werden:

$$\eta_t = x_t \beta + \varepsilon_t, \quad t = 1, \dots, n$$

und

$$y_t = i \Leftrightarrow \eta_t \in (\tau_{i-1}, \tau_i], \quad i = 1, 2, \dots, C, \quad t = 1, \dots, n.$$

Dabei wird angenommen, daß die unbekanntenen und damit zu schätzenden Schwellenwerte durch $-\infty = \tau_0 < \tau_1 < \dots < \tau_{C-1} < \tau_C = +\infty$ geordnet sind.

Unterschiedliche Modelle werden wiederum durch verschiedene Fehlertermannahmen generiert. Unterstellt man, daß die Fehlerterme ε_t stochastisch unabhängig und identisch standard-normalverteilt sind, so erhält man das ordinale Probitmodell. Dies entspricht der Verteilung in Abbildung 3. Ist der Fehlerterm ε_t logistisch verteilt, so erhält man das ordinale Logitmodell (Bock 1975). Ähnliche Modelle können durch andere Verteilungsannahmen für ε_t erzeugt werden (McCullagh 1980). Übrigens unterstellen alle erwähnten Modelle, daß die Folge der Fehlerterme $\{\varepsilon_t\}_{t=1, \dots, n}$ stochastisch unabhängig ist.

Eine natürliche Rangfolge zwischen den Kategorien der abhängigen Variablen reicht im allgemeinen nicht zur Anwendung eines ordinalen Schwellenwertmodells aus. Diese Einschränkung ergibt sich aus der einfachen Überlegung, daß die Effekte der unabhängigen Variablen einen U-Effekt in dem Sinn bewirken können, daß sowohl extrem hohe als auch extrem niedrige Werte einer unabhängigen Variablen zu hohen Neigungen bezüglich einer Kategorie der abhängigen Variablen führen können, während mittlere Werte der unabhängigen Variablen zu niedrigen Neigungen führen. In einer derartigen Situation stehen zwei Analysealternativen zur Verfügung. Erstens kann man in Analogie zur nichtlinearen metrischen Regression ein in x_t nichtlineares Prädiktormodell für η_t formulieren. Andererseits kann man aber auch einfach die natürliche Rangfolge zwischen den Kategorien der abhängigen Variablen ignorieren und ein Modell zur Analyse von nominalskalierten abhängigen Variablen verwenden. Das stereotype Modell von Anderson (1984) basiert auf einer Modifikation des multinomialen Logitmodells, in dem spezielle Parametrisierungen zur Berücksichtigung von Parameteranordnungen

verwendet werden.

Gruppierte metrische abhängige Variablen, bei denen lediglich die Klassenzugehörigkeit zu einem Intervall bekannt ist, werden in der Regel analog zu ordinalen Variablen behandelt, indem die Schwellenwerte auf die festen Intervallgrenzen restringiert werden. Zusätzlich muß die Varianz des Fehlertermvektors freigesetzt werden (Stewart 1983).

3.4 Modelle zur Analyse von zensierten Variablen

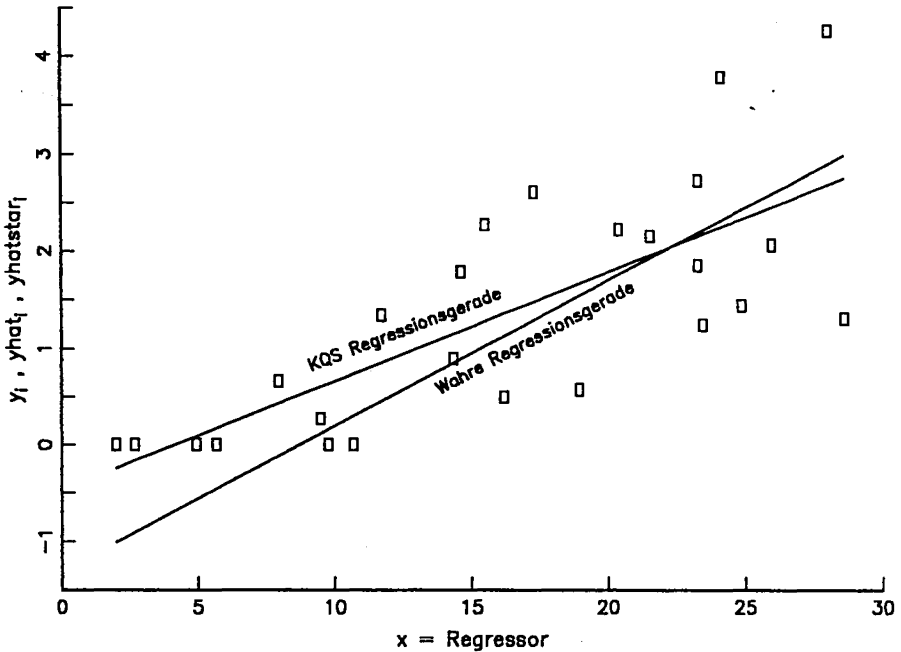
Bei einer zensierten Variablen y_i kann der exakte Wert der abhängigen Variablen lediglich für eine Teilmenge der Stichprobenelemente $t = 1, \dots, n$ beobachtet werden. Beobachtungen innerhalb der Komplementärmenge werden hingegen nur durch eine Indikatorvariable (etwa durch 0 aus dem Indikatorset $\{0, 1\}$) repräsentiert, ohne daß der exakte Wert der abhängigen Variablen bekannt ist. Derartige Variablen treten oft bei Untersuchungen zum Verhalten von Konsumenten beim Kauf von dauerhaften Gebrauchsgütern auf, bei denen die abhängige Variable den Kaufbetrag eines bestimmten Gebrauchsgutes repräsentiert. Eine ganze Reihe von dauerhaften Konsumgütern (etwa Autos und Waschmaschinen) werden nicht von allen Konsumenten innerhalb einer fixen Haushaltsperiode (etwa ein Jahr) gekauft, sondern nur in Abständen von mehreren Jahren erneuert. Damit repräsentiert die Ausprägung 0 der Indikatorvariablen das Unterlassen eines Kaufaktes. Wird ein Kauf getätigt, so sind die Beträge grundsätzlich positiv. Wendet man den gewöhnlichen KQS auf zensierte Daten an, so erhält man inkonsistente und verzerrte Schätzungen, wie man anhand der Abbildung 4 ersehen kann.

Offensichtlich ist der gewöhnliche KQS nicht erwartungstreu und konsistent, da die auf Null zensierten Werte den Steigungsparameter β_1 nach unten verzerren. Verwendet man lediglich diejenigen Stichprobenelemente mit positiven y_i Beobachtungen, so wird dieser Fehler zwar korrigiert. Allerdings ist diese Vorgehensweise ineffizient, da man auf die in den zensierten Fällen enthaltene Information verzichtet.

Ein adäquates Modell zur Behandlung von zensierten Variablen⁵ ist das Tobit-Modell von Tobin (1959), das eine Mischung des gewöhnlichen Regressionsmodells mit einem binären Probitmodell darstellt. Bei der Modellierung von Tobitmodellen wird folgende Grundstruktur (Amemiya 1973) unterstellt:

⁵ Man beachte, daß der hier verwendete Zensierungsbegriff nur sehr wenig mit den Zensierungsbegriffen in der "Failure Time Series" Literatur zu tun hat.

Abbildung 4: KQS bei zensierten Daten



$$\eta_t = x_t \beta + \varepsilon_t, \quad t = 1, \dots, n$$

ε_t ist eine Folge von stochastisch unabhängigen und identisch normalverteilten Zufallsvariablen mit Erwartungswert 0 und unrestringierter Varianz $\sigma^2 > 0$. Die Meßrelation wird durch folgende Gleichung definiert:

$$y_t = \begin{cases} \eta_t & \text{wenn } \eta_t \text{ einen bekannten Schwellenwert } \tau \text{ überschreitet} \\ \tau & \text{sonst} \end{cases}$$

Im Kontext von Untersuchungen zur Konsumentennachfrage kann die partiell beobachtbare latente Variable η_t als marginale Kaufneigung interpretiert werden. Das oben skizzierte Tobit-Modell ist lediglich ein exemplarisches Beispiel einer weiten Klasse von Tobit-Modellen, die eingehend in Amemiya (1985) dargestellt wird. Eine wichtige Verallgemeinerung parametrisiert den festen Schwellenwert τ durch eine stochastische oder deterministische Regressionsgleichung.

4. Schätzung und Modellspezifikation

Mit Ausnahme des multinomialen Probitmodells ist die numerische Schätzung der Parameter in qualitativen Variablenmodellen recht einfach, da die Loglikelihood-Funktionen in der Regel ein eindeutiges Maximum besitzen, so daß man relativ einfache numerische Optimierungsverfahren wie Newton-Raphson und Fisher's Scoring (Thisted 1988) verwenden kann. Modellspezifische Probleme findet der Leser in den Monographien von Maddala (1983) und Amemiya (1985). Daher beschränkt sich dieser Abschnitt auf eine knappe Diskussion von speziellen Datensatzkonstellationen, Modellspezifikationsprozeduren und auf eine selektive Angabe problemadäquater Softwarepakete.

4.1 Strukturelle Nullen und Stichprobennullen

Bei Datensätzen mit qualitativen Variablen tritt sehr häufig das Problem auf, daß eine Teilmenge der Ausprägungen innerhalb von Teilstichproben, die sich durch gemeinsame Regressoren charakterisieren lassen, nicht besetzt sind. Die Ursachen für derartige Nullbesetzungen lassen sich durch folgende Unterscheidung klassifizieren:

1. Innerhalb einer Teilstichprobe ist eine bestimmte Ausprägung der abhängigen Variablen logisch unmöglich. Ein einfaches Beispiel dieses Falles, der in der Literatur als strukturelle Null bezeichnet wird, ist die Tatsache, daß Frauen in nationalsozialistischen Organisationen wie der SS nicht zugelassen wurden. Daher kann die Variable Geschlechtszugehörigkeit nicht zur Erklärung der SS-Mitgliedschaft herangezogen werden.
2. Weitaus häufiger tritt jedoch der Fall auf, daß bestimmte (i.d.R. seltene) Kombinationen von Ausprägungen der abhängigen und der unabhängigen Variablen aufgrund unzureichendem Stichprobenumfangs nicht besetzt sind. Dieser Fall wird als Stichprobennull bezeichnet.

Nullbesetzungen können unter bestimmten Parametrisierungen zu einer Nichtexistenz des Maximum-Likelihood-Schätzers führen, obwohl sich dieses Problem bei Stichprobennullen durch eine selektive Nacherhebung oder durch eine Vergrößerung des Stichprobenumfangs eliminieren läßt. Das resultierende Schätzproblem wird im folgenden an einem einfachen Beispiel einer Vierfeldertafel illustriert, die theoretisch mit Hilfe eines binären Logitmodells analysiert wird:

Geschlecht / Organisation	SS-Mitglieder	keine SS-Mitglieder
Männlich	n_{11}	n_{12}
Weibliche	0	n_{22}

Man beachte, daß die relative Häufigkeit einer SS-Mitgliedschaft innerhalb der Gruppe der Männer durch $n_{11}/(n_{11}+n_{12})$ gegeben ist, während die relative Häufigkeit einer SS-Mitgliedschaft von Frauen *a priori* auf Null restringiert ist. Eine Parametrisierung zur Schätzung der Effekte von Geschlecht auf Mitgliedschaft mit einem binären Logitmodell ergibt folgende Modellstruktur:

$$P(\text{SS Mitglied} | \text{Männlich}) = \frac{\exp(\mu)}{1 + \exp(\mu)}$$

$$P(\text{SS Mitglied} | \text{Weiblich}) = \frac{\exp(\mu + \alpha)}{1 + \exp(\mu + \alpha)}$$

$$P(\text{Kein SS Mitglied} | \text{Männlich}) = 1 - P(\text{SS Mitglied} | \text{Männlich})$$

$$P(\text{Kein SS Mitglied} | \text{Weiblich}) = 1 - P(\text{SS Mitglied} | \text{Weiblich})$$

Offensichtlich können die Parameter μ und α exakt an die Daten angepaßt werden, da in der allgemeinen Formulierung des Modells nur zwei freie Wahrscheinlichkeiten, nämlich $P(\text{SS Mitglied} | \text{Männlich})$ und $P(\text{SS Mitglied} | \text{Weiblich})$ auftreten. Daher kann der Maximum-Likelihood-Schätzer von μ analytisch durch den Ausdruck

$$\hat{\mu} = \ln\left(\frac{n_{11}}{n_{12}}\right)$$

bestimmt werden. Ist μ gegeben, so kann man die absolute Häufigkeit von $n_{21} = 0$ nur dann exakt an

$$P(\text{SS Mitglied} | \text{Weiblich}) = \frac{\exp(\mu + \alpha)}{1 + \exp(\mu + \alpha)}$$

anpassen, wenn man den Parameterschätzer für α gegen $-\infty$ konvergieren läßt. Somit existiert in diesem Modell kein endlicher Maximum-Likelihood-Schätzer für (μ, α) . Verwendet man zur numerischen Bestimmung des Maximum-Likelihood-Schätzers ein iteratives Optimierungsverfahren wie das Newton-Raphson-Verfahren, so kann man diesen Effekt sehr leicht an einer

Divergenz einer Parametersubgruppe entdecken. Falls ein infinites Maximum-Likelihood-Schätzer auftritt, so muß man die divergierenden Parameter durch eine Reparametrisierung aus der Modellspezifikation eliminieren. Leider sind eine Reihe von Softwarepaketen nicht in der Lage, eine Parameterdivergenz zu diagnostizieren. Dies liegt zum Teil daran, daß einige Programmsysteme ungeeignete Schätzverfahren wie den einstufigen GSK-Schätzer verwenden. Anderere Systeme benutzen hingegen ungeeignete Abbruchkriterien wie Gradientenabstände zu Null. Ein Anwender muß daher diese Probleme bei der Auswahl und der Anwendung von Software berücksichtigen. Erwähnenswert ist die Tatsache, daß eine Nichtexistenz des ML-Schätzers grundsätzlich an eine bestimmte Parametrisierung gebunden ist. Restringiert man im obigen Beispiel den Regressionsparameter α auf Null, so existiert der eindimensionale ML-Schätzer μ , so daß man unter gewissen Einschränkungen das Existenzproblem des ML-Schätzers durch eine eingeschränkte Parametrisierung umgehen kann. Allerdings besteht dabei die Gefahr, daß man ein inhaltlich sinnloses Modell schätzt. Übrigens kann das Problem der Nichtexistenz von Maximum-Likelihood-Schätzern nicht nur bei der Analyse von Kontingenztafeln, sondern auch bei der Analyse von Individualdaten mit metrischen und/oder nicht-metrischen Regressoren auftreten. Eine detaillierte Analyse der statistischen Implikationen von strukturellen Nullen und Stichprobennullen findet der Leser bei Albert und Anderson (1984).

4.2 Modellspezifikationsprobleme

Die Auswahl einer geeigneten Modellklasse basiert in der Regel auf theoretischen a priori Entscheidungen, bei denen etwa die Skalenniveaus der abhängigen und der unabhängigen Variablen neben anderen Daten- und Stichprobeneigenschaften einbezogen werden. Ist ein geeignetes Modell oder eine geeignete Modellklasse ausgewählt worden, so verbleibt für den Anwender das Problem, eine Auswahl zwischen geeigneten Regressoren und Designmatrixformulierungen (= Parametrisierung von Haupt- und Interaktionseffekten) zu treffen. Für diesen Zweck stehen drei unterschiedliche, aber eng verwandte Modellspezifikationsprozeduren zur Verfügung:

1. Zur Evaluation von Parametrisierungen werden häufig formale Signifikanztests herangezogen, mit denen man auf der Basis eines a priori gewählten Signifikanzniveaus *wichtige* Parameter bzw. Variablen einschließt und *unwichtige* ausschließt. Bei dieser Vorgehensweise kann eine Reihe von Problemen entstehen, obwohl diese (rein mechanische) Technik recht

einfach anzuwenden ist. Das erste Problem ist die Selektion einer geeigneten Evaluationsstrategie. Einige Statistiker propagieren die *Bottom-UP*-Technik, bei der man - beginnend mit dem einfachsten Modell (dem Mittelwertmodell) - sukzessiv immer weitere Regressoren bzw. Effekte einschließt. Andere wiederum starten mit dem komplexesten Modell (einem saturierten Modell) und eliminieren sequentiell insignifikante Parameter. Obwohl beide Evaluationstrategien Vorteile bieten, steht jeder Anwender vor dem Problem, sich für eine der beiden Strategien zu entscheiden, die übrigens fast immer zu unterschiedlichen Ergebnissen führen. Ein weiteres Problem bei der Anwendung von Signifikanztests ist das multiple Testproblem (Savin 1984). Theoretisch geht man bei der Herleitung von Signifikanztests davon aus, daß lediglich *ein einziger* Test an einem Datensatz durchgeführt wird. Dies entspricht einer rein konfirmatorischen Fragestellung. Bei explorativen Untersuchungen muß man Tests jedoch in der Regel wiederholt anwenden. Daher muß man im Prinzip multiple Testgrenzen auf der Basis der Bonferroni-Ungleichung oder aufgrund ähnlicher Wahrscheinlichkeitsabschätzungen verwenden. Außerdem muß sich ein Anwender für ein adäquates Signifikanzniveau entscheiden. Diese Auswahl ist recht problematisch, da a priori keine hinreichenden Gründe zur Festsetzung von Signifikanzniveaus zur Verfügung stehen.

2. Als weiteres Hilfsmittel zur Modellselektion werden häufig Determinationsmaße herangezogen, die in Analogie zum Bestimmtheitsmaß R^2 der klassischen Regressionsanalyse hergeleitet werden. Klassisches Problem dieser Determinationsmaße ist, daß der Einbezug zusätzlicher Parameter grundsätzlich zu einem Anstieg des Maßes führt. Daher muß man ebenso wie bei der Anwendung von Signifikanztest Diskriminationsgrenzen verwenden, deren Festlegung ein nichttriviales Problem darstellt. Die geeignetste Verwendung von Determinationsmaßen ist der Vergleich von nicht-genesteten (nicht-geschachtelten) Modellen, bei denen man mit Ausnahme von Cox-Tests keine Signifikanztests heranziehen kann.

3. Zur Aufdeckung der Effekte von Ausreißern, Nichtlinearitäten und unberücksichtigten Variablen kann man häufig eine graphische Repräsentation von Residuen verwenden. Allerdings erfordert die Inspektion von Residuen einige Erfahrung in der empirischen Datenanalyse. Außerdem müssen Residuen in Abhängigkeit vom ausgewählten Modell geeignet skaliert werden. Daher steht kein eindeutig definiertes Residuum für alle Modelle zur Verfügung, das für alle Modelle analog zu interpretierende Aussagen liefert.

Allgemeine Strategien zur Selektion von Submodellen findet man in McCullagh und Nelder (1983) sowie in Arminger und Küsters (1986), in denen auch weitere Literaturhinweise angegeben sind.

4.3 Softwarepakete zur Analyse von qualitativen Variablen

Mittlerweise enthalten fast alle statistischen Softwarepakete einen oder mehrere Schätzer und Methoden zur Analyse von nichtmetrischen abhängigen Variablen. Einen detaillierten Vergleich von Softwarepaketen zum Stand 1985 findet der Leser in Arminger und Küsters (1986). Die Auswahl eines Systems hängt im wesentlichen von den individuellen Bedürfnissen eines Nutzers in Hinblick auf Kriterien wie Problemgröße, Parametrisierungsoptionen und Programmflexibilität ab. Großrechner-Systeme sind in der Regel in Bezug auf die Problemgröße nahezu unbeschränkt, während PC-Programme tendenziell einfacher zu benutzen sind. Daher beschränken sich die folgenden Bemerkungen auf Kerneigenschaften von ausgewählten und teilweise weitverbreiteten Systemen:

1. GLIM (Payne 1985, McCullagh und Nelder 1983) kann zur Schätzung von binären Modellen wie das Logit- und das Probitmodell, aber auch zur Schätzung von komplementären Log-Log-Modellen herangezogen werden. Designmatrizen zur Parametrisierung von nominalskalierten unabhängigen Variablen mit Eckpunkteffekten kann man mit GLIM durch eine sehr einfach handhabbare FIT-Direktive spezifizieren. Zum Aufbau der Haupt- und Interaktionseffekte für zwei kategoriale unabhängige Variablen A und B muß man lediglich den Befehl $A * B$ (bzw. $A + B + A.B$) angeben, durch den die Designmatrix automatisch aufgebaut wird. Außerdem kann man mit GLIM Maximum-Likelihood-Schätzer für loglineare Modelle zur Analyse von mehrdimensionalen Kontingenztabellen berechnen, so daß man die Parameter von multinomialen Logitmodellen indirekt durch die Spezifikation von äquivalenten loglinearen Modellen bestimmen kann, sofern alle Variablen nominalskaliert sind. Das Programmpaket GLIM steht für eine sehr breite Reihe von Maschinen zur Verfügung, zu denen unter anderem Großrechner von IBM, DEC und CDC sowie PC's gehören.
2. LIMDEP (Greene 1986) kann fast alle in diesem Artikel erwähnten Modelle schätzen. Die einzige Ausnahme ist die Schätzung des multinomialen Probitmodells, das allerdings auch mit keinem anderen Standard-

paket geschätzt werden kann. Trotz dieser breiten Spannweite ist es mit LIMDEP relativ zeitraubend, kategoriale unabhängige Variablen in die Analysen einzubeziehen, da Designmatrizen nicht automatisch aufgebaut werden können. Anstelle dessen muß man alle Designmatrizen mit Hilfe der Matrix-Makro-Kommandos in LIMDEP selbst aufbauen, was bei komplexen Designs recht mühsam ist. Innerhalb von LIMDEP wird als Schätzverfahren durchweg Maximum-Likelihood verwendet. LIMDEP steht unter anderem für Großrechner von IBM und DEC sowie für PC's zur Verfügung.

3. SPSS enthält eine Option zur Schätzung von loglinearen Modellen mit zentrierten Effekten. Dieser Ansatz ist recht nützlich, wenn man eine Kontingenztabelle ohne Stichprobennullen und strukturelle Nullen analysieren möchte.

4. SAS enthält eine Reihe von Prozeduren zur Schätzung von binären und ordinalen Modellen sowie von loglinearen Modellen und multinomialen Logitmodellen. Designmatrizen können mit einem ähnlichen Syntax wie in GLIM aufgebaut werden. Allerdings werden zentrierte Reparametrisierungen verwendet, so daß man Probleme mit Nullbesetzungen bekommen kann. SAS verwendet eine automatische Aggregation der Daten, die zu Problemen führen kann (Küsters 1988b). Im Gegensatz zu den anderen Paketen ermöglicht SAS die recht einfache Durchführung von Tests auf marginale Homogenität in Kontingenztafeln. Aufgrund der in SAS enthaltenen Matrixsprache kann man allerdings alle in diesem Artikel erwähnten Schätzer selbst implementieren, sofern man einige grundlegende numerische Verfahren kennt.

5. BMDP enthält die Optionen zur Schätzung von loglinearen Modellen und binären Modellen. Unter anderem enthält BMDP auch eine Routine zur automatischen Erzeugung von zentrierten Designmatrizen.

Multinomiale Probitmodelle können mit keinem der erwähnten Programmsysteme geschätzt werden, da der numerische Aufwand zu hoch ist. Mit Ausnahme dieses Modells kann man daher lediglich mit LIMDEP alle in diesem Artikel behandelten Modelle schätzen. Aufgrund der Probleme bei der Erstellung von Designmatrizen ist GLIM jedoch bedeutend benutzerfreundlicher, wenn alle Variablen nominalskaliert sind, so daß man ein loglineares Modell zur Schätzung heranziehen kann.

Eine weitere Alternative zur Schätzung von Maximum-Likelihood-Schätzern besteht in der Verwendung einer allgemeinen Matrixsprache wie GAUSS (Edlefsen und Jones 1988), die eine Reihe von Bibliotheksmodulen

enthält, die u. a. auch binäre und ordinale Probit- und Logitmodelle sowie loglineare Modelle und multinomiale Logitmodelle umfaßt. GAUSS ist einfach zu erlernen, so daß der Benutzer die einzelnen Applikationsmodule leicht an seine Bedürfnisse anpassen kann. Die Programmierung mit GAUSS sowie eine Anwendung des binären Probitmodells findet der Leser in Küsters und Armingier (1989). Allerdings ist GAUSS nur für PC's implementiert, so daß man bei größeren Datensätzen auf andere Programmiersprachen wie etwa FORTRAN oder APL2 umsteigen muß.

5. Beispiel

In diesem Abschnitt wird die Schätzung eines binominalen Logitmodells anhand eines historischen Datensatzes demonstriert. Der Datensatz wurde der umfangreichen Studie über *Struktur und Handeln parlamentarischer Führungsgruppen in Deutschland und Frankreich 1848/49* von Best (1986) entnommen. Ziel der Analyse ist die Bestimmung der wichtigsten sozioökonomischen und politischen Einflußfaktoren auf das Abstimmungsverhalten der Abgeordneten der Frankfurter Nationalversammlung von 1848/49, das mit Hilfe eines hier nur kurz erläuterten Verfahrens auf eine dichotome Links-Rechts-Variable reduziert wurde. Die folgende Analyse verwendet die in Thome (1989) beschriebene vierdimensionale Kontingenztabelle⁶ und basiert auf den Datensätzen von 743 Abgeordneten⁷. Folgende Variablen werden verwendet:

- KONF Konfessionszugehörigkeit mit den Ausprägungen:
 1 Katholisch.
 2 Nicht katholisch (Vorwiegend Protestantisch).
- VK Verfassungskontext vor 1848 mit den Ausprägungen:
 1 Abgeordneter repräsentierte einen Wahlkreis eines Einzelstaates, der bis mindestens 1848 als absolute Monarchie verfaßt war.
 2 Abgeordneter repräsentierte einen Wahlkreis eines Einzelstaates, der schon vor 1848 eine Verfassung erhalten hatte.

⁶ Siehe Thome (1989), S. 130.

⁷ Insgesamt waren in der Nationalversammlung 809 Abgeordnete vertreten. Abgeordnete mit fehlenden Werten wurden aus der Analyse ausgeschlossen.

- POLERF** Politische Erfahrungen des Abgeordneten vor 1848 mit den Ausprägungen:
 1 Vor 1848 ausschließlich in politischen Ämtern tätig.
 2 Vor 1848 in politischen Ämtern tätig *und* an illegalen Aktivitäten beteiligt.
 3 Vor 1848 ausschließlich an illegalen Aktivitäten beteiligt.
 4 Keine politischen Erfahrungen vor 1848.
- LRO** Links-Rechts-Orientierung mit den Ausprägungen:
 1 Rechts.
 2 Links.

Die Links-Rechts-Skala wurde mit Hilfe von Indikatoren konstruiert, die das Abstimmungsverhalten zu Fragen der Gültigkeit von Grundrechten, Ausschluß von Wahlrechten, Versammlungsverboten bei Gefahren für öffentliche Sicherheit und Ordnung etc. beschreiben. Die Daten sind in der folgenden Kontingenztafel, die auch eine Stichprobennull enthält, aufgelistet.

	KONF1				KONF2			
	VK1		VK2		VK1		VK2	
	LR01	LR02	LR01	LR02	LR01	LR02	LR01	LR02
POLERF1	11	6	19	13	29	8	23	35
POLERF2	5	1	0	8	15	3	19	25
POLERF3	13	22	4	2	29	16	12	23
POLERF4	103	72	26	11	94	35	23	38

Mit Hilfe einer deskriptiven Analyse durch die bivariaten Kontingenztafeln (KONF, LRO), (VR, LRO) und (POLERF, LRO) kann man lediglich erkennen, daß der Verfassungstext VK einen erheblichen Einfluß auf die ideologische Orientierung LRO der Parlamentarier ausübt. Die folgenden Tabellen geben die konditionalen Verteilungen von LRO bei gegebenen unabhängigen Variablen an:

	LR01	LR02			LR01	LR02	
KONF1	0.5728	0.4272		VK1	0.6472	0.3528	
KONF2	0.5714	0.4286		VK2	0.4484	0.5516	

	LR01	LR02	
POLERF1	0.5694	0.4306	
POLERF2	0.5132	0.4868	
POLERF3	0.4793	0.5207	
POLERF4	0.6119	0.3881	

Die konditionalen Verteilungen für LRO stimmen somit für katholische und nichtkatholische Abgeordnete fast überein. Daher stützt diese Tabelle zunächst die These, daß die Konfessionszugehörigkeit keinen Einfluß auf die politische Orientierung aufwies. Die folgende dreidimensionale Tabelle, in der die konditionalen Verteilungen von LRO für alle Kombinationen zwischen Verfassungskontext und Konfession angegeben wird, zeigt hingegen, daß diese Hypothese nicht korrekt ist.

	KONF1		KONF2		
	LR01	LR02	LR01	LR02	
VK1	0.5665	0.4335	0.7293	0.2707	
VK2	0.5904	0.4096	0.3889	0.6111	

Die auf VK und KONF konditionierte Verteilung der Links-Rechts-Orientierung LRO ist gegenüber der konditionierenden Variablen VK nahezu invariant. Somit reagieren katholische Abgeordnete offenbar nur in einem sehr geringen Ausmaß auf den Verfassungskontext. Bei nichtkatholischen Abgeordneten (vorwiegend Protestanten) ist der Verfassungskontext hingegen ein dominanter Faktor für die politische Orientierung. Möglicherweise nahmen

nichtkatholische Abgeordnete ihre Umgebung stärker als katholische Abgeordnete wahr.

Die bivariate Kontingenztabelle (POLERF, LRO) zeigt einen Zusammenhang zwischen der politischen Erfahrung und der Rechts-Links-Orientierung. Abgeordnete ohne politische Erfahrung neigten eher zu einem rechtsorientierten Abstimmungsverhalten, während Abgeordnete mit einer subversiven Vergangenheit eher eine linke Orientierung aufwiesen. Die Stärke dieses Effektes muß allerdings durch ein multivariates Modell beurteilt werden, in dem alle unabhängigen Variablen berücksichtigt werden.

Zur Bestimmung der gemeinsamen Wirkung der Einflußfaktoren wurde eine Sequenz von binären Logitmodellen mit GLIM 3.77 geschätzt. Der folgende Ausschnitt aus der Log-Datei enthält die Schätzung des ausgewählten Modells. [i] und [o] sind Prefixmarker für Eingabekommandos bzw. Glim-Ausgaben. Die unabhängigen Variablen KONF, VK und POLERF wurden mit den Ausprägungsziffern kodiert, während R und L die Häufigkeiten von rechts - bzw. linksorientierten Abgeordneten für jede Kombination der unabhängigen Variablen angeben. Die Anzahl der Kombinationen wird in der UNITS Direktive festgelegt, während die Anzahl der Ausprägungen jeder unabhängigen Variablen durch die FAC Anweisung spezifiziert wird. Durch die Anweisung CAL TOTAL=R+L wird die Häufigkeit TOTAL jeder Kombination der unabhängigen Variablen berechnet, die in der Anweisungsfolge ERROR B TOTAL und YVAR L zur Spezifikation eines binären Logitmodells mit klassifizierten Daten benötigt wird. Durch die FIT Anweisung wird ein Modell geschätzt, bei dem die unabhängigen Variablen durch einen varianzanalytischen Design automatisch expandiert werden. Die abhängige Variable ist der Anteil der linken Abgeordneten innerhalb jeder Kombination der unabhängigen Variablen. Das Kommando FIT KONF*VK+POLERF bewirkt die Schätzung eines Modells, das alle Haupteffekte von KONF, VK und POLERF enthält und in dem zusätzlich alle Interaktionseffekte zwischen KONF und VK berücksichtigt werden. Durch die DISPLAY Anweisung können verschiedene Schätzer und Statistiken selektiv ausgegeben werden.

```

[i] $UNITS 16
[i] $FAC KONF 2 VK 2 POLERF 4
[i] $DATA KONF VK POLERF R L
[i] $READ
[i] 1 1 1 11 6
[i] 1 1 2 5 1
[i] 1 1 3 13 22
[i] 1 1 4 103 72
[i] 1 2 1 19 13
[i] 1 2 2 0 8
[i] 1 2 3 4 2
[i] 1 2 4 26 11
[i] 2 1 1 29 8
[i] 2 1 2 15 3
[i] 2 1 3 29 16
[i] 2 1 4 94 35
[i] 2 2 1 23 35
[i] 2 2 2 19 25
[i] 2 2 3 12 23
[i] 2 2 4 23 38
[i] $CAL TOTAL=R+L
[i] $ERROR B TOTAL
[i] $YVAR L
    >>> Schaetzung von verschiedenen Zwischenmodellen <<<
[i] $COM AUSGEWAEHLTES MODELL
[i] $CAL P3=%IF(%EQ(POLERF,3),1,0)$
[i] $COM KONF*VK+P3 --> LRO
[i] $FIT KONF*VK+P3$
[o] scaled deviance = 22.396 at cycle 3
[o]          d.f. = 11
[i] $DISPLAY ME
[o] Current model:
[o]   number of units is 16
[o]   y-variate L

```

```

[o] weight      *
[o] offset      *
[o] probability distribution is BINOMIAL
[o]   with binomial denominator TOTA
[o]           link function is LOGIT
[o]           scale parameter is 1.000
[o] terms = 1 + KONF + VK + P3 + KONF.VK
[o]           estimate      s.e.      parameter
[o] 1      -0.3465      0.1366      1
[o] 2      -0.7550      0.2005      KONF(2)
[o] 3      -0.05727     0.2605      VK(2)
[o] 4       0.5161      0.2094      P3
[o] 5       1.525       0.3337      KONF(2).VK(2)
[o]           scale parameter taken as 1.000

```

Das ausgewählte Modell $KONF * VK + (POLERF=3)$ zeigt einige interessante Effekte, die man teilweise auch schon bei der Analyse der bi- und tri-variablen Tabellen vermuten konnte. Der Parameter mit dem Namen 1 und dem Schätzwert -0.3465 repräsentiert die Neigung zu linken Abstimmungspositionen innerhalb der Referenzkategorie, die durch die Kombination der unabhängigen Variablen $KONF=1$, $VK=1$ und $POLERF=1$ gegeben ist. Das negative Vorzeichen bedeutet, daß diejenigen katholischen Abgeordneten aus absolutistischen Staaten, deren politische Erfahrungen vor 1848 sich auf die Ausübung offizieller Ämter beschränkte, wahrheitlich nach rechts tendierten. Der Parameter $KONF(2)$ mit dem Schätzwert -0.7550 gibt die Verschiebung der Kombination $KONF=2$, $VK=1$ und $POLERF=1$ gegenüber der Referenzkategorie an. Dies impliziert, daß nichtkatholische Abgeordnete gegenüber katholischen Abgeordneten innerhalb der Gruppe der Parlamentarier mit monarchistischem Verfassungskontext und politischen Erfahrungen durch politische Ämter eine noch höhere Neigung zu rechten Abstimmungspositionen aufwiesen. Der nichtsignifikante Parameterschätzer für $VK(2)$ wird lediglich benötigt, um den signifikanten Interaktionseffekt $KONF(2) \cdot VK(2)$ zu schätzen. Ein Einfluß des Verfassungskontexts innerhalb der Gruppe der Parlamentarier mit katholischer Konfession auf die Links-Rechts-Orientierung ist hingegen nicht feststellbar. Innerhalb der Gruppe der Parlamentarier, die einen Wahlkreis repräsentierten, der vor 1848 eine Verfassung erhielt, ist hingegen ein interessanter Effekt erkennbar. Nichtkatholische Abgeordnete neigten innerhalb dieser Gruppe

eher zu linken Abstimmungspositionen, wie man aus dem positiven Parameterschätzer 1.525 für KONF (2) .VK (2) schließen kann.

Eine Auswirkung des politischen Erfahrungshorizontes zeigt sich im Datensatz in einer signifikanten Form lediglich bei denjenigen Abgeordneten, die sich vor 1848 subversiv betätigten. Abgeordneten mit subversivem Hintergrund neigten offensichtlich zu einem verstärkt linken Abstimmungsverhalten, wie der zur Dummy-Variablen P3 korrespondierende Parameterschätzer 0.5161 zeigt. Abgeordnete aus den anderen Gruppen (POLERF=1), (POLERF=4) sind hingegen relativ homogen, wenn man die Variablen KONF und VK berücksichtigt. Somit ist die aus der marginalen Kontingenztafel zwischen POLERF und LRO ablesbare Rangfolge POLERF=4, POLERF=1, POLERF=2, POLERF=4 ein Scheineffekt, der lediglich durch die Aggregation über alle Ausprägungen der Variablen KONF und VK entsteht.

Für die nicht abgedruckten Modelle wurden folgende Devianzstatistiken berechnet. Differenzen von Devianzstatistiken zwischen genesteten Modellen sind approximativ χ^2 verteilt, wobei die Anzahl der Freiheitsgrade durch die Differenz der Freiheitsgrade zwischen den konkurrierenden Modellen bestimmt wird.

Modell	Devianz	df (Freiheitsgrade)
1	79.744	15
KONF	79.742	14
VK	51.574	14
POLERF	71.839	12
KONF+VK	50.369	13
KONF*VK	28.493	12
VK+POLERF	44.756	11
VK*POLERF	37.052	8
KONF+VK+POLERF	43.010	10
KONF*VK+POLERF	22.296	9
KONF*VK+(POLERF=3)	22.396	11

Die Insignifikanz zwischen den Kategorien POLERF=1, POLERF=2 und POLERF=4 ist auch aus der Devianzdifferenz zwischen den Modellen KONF*VK+POLERF (3) und KONF*VK+POLERF ersichtlich, die bei 2 Freiheitsgraden 0.1 beträgt und somit nicht signifikant ist. Übrigens lassen sich die Parameter im saturierten Modell KONF*VK*POLERF nicht bestimmen, da die Stichprobennull in der Kontingenztafel zu einer Nichtexistenz des ML-Schätzers im saturierten Modell führt.

6. Schlußfolgerungen

Regressionsmodelle zur Analyse von nichtmetrischen abhängigen Variablen können beim gegenwärtigen Stand der statistischen Modellierung ohne erhebliche Probleme angewendet werden. Allerdings sind Schätzer für die meisten Modelle bis zum jetzigen Zeitpunkt noch nicht in einer benutzerfreundlichen Form implementiert. Dies erzwingt beim Anwender eine Auseinandersetzung mit den typischen Analyseproblemen wie Stichprobennullstellen etc., da automatische Hinweise auf Datenprobleme von gegenwärtigen Softwareprodukten noch nicht gegeben werden.

Literatur

- Albert, A. und Anderson, J.A. (1984): On the Existence of Maximum Likelihood Estimates in Logistic Regression Models. *Biometrika* 71, 1 - 10.
- Amemiya, T. (1973): Regression Analysis when the Dependent Variable is Truncated Normal. *Econometrica* 41, 997- 1016.
- Amemiya, T. (1985): *Advanced Econometrics*. Oxford.
- Anderson, J.A. (1984): Regression and Ordered Categorical Variables (with discussion). *Journal of the Royal Statistical Society B*, 1 -30.
- Arminger, G. und Küsters, U. (1986): *Statistische Verfahren zur Analyse qualitativer Variablen*. Forschungsberichte der Bundesanstalt für Straßenwesen 147. Aachen.
- Best, H. (1990): *Die Männer von Bildung und Besitz. Struktur und Handeln parlamentarischer Führungsgruppen in Deutschland und Frankreich 1848/49*. Bd. 90 der Schriften zur Geschichte des Parlamentarismus und der politischen Parteien. Düsseldorf.
- Bishop, Y.M.M., Fienberg, S.E. und Holland, P.W. (1975): *Discrete Multivariate Analysis - Theory and Practice*. Cambridge.
- Bock, R.D. (1975): *Multivariate Statistical Methods in Behavioral Research*. New York.
- Daganzo, C. (1979): *Multinomial Probit - The Theory and Its Application to Demand Forecasting*. New York.

- Edlefsen, L. und Jones, S. (1986): GAUSS-Programming Language Manual. Kent, Aptech Systems Inc.
- Fahrmeir, L. und Hamerle, A. (1984, Hrsg.): Multivariate statistische Verfahren. Berlin.
- Greene, W. (1986): LIMDEP-Reference-Manual. New York University.
- Jarausch, K.H. und Arminger, G. (1988): The German Teaching Profession and Nazi Party Membership: A Demographic Logit Model. *Journal of Interdisciplinary History*, 1989, 197-225.
- Jarausch, K.H., Arminger, G. und Thaller, M. (1985): Quantitative Methoden in der Geschichtswissenschaft - Eine Einführung in die Forschung, Datenverarbeitung und Statistik. Darmstadt.
- Küsters, U. (1988a): LIMDEP - Ein vielseitiges Programmsystem zur Schätzung von Regressionsmodellen für abhängige Variablen beliebigen Meßniveaus. In: Faulbaum, F. und Uehlinger, H.M.: Fortschritte der Statistik-Software 1. Stuttgart.
- Küsters, U. (1988b): Analyse von qualitativen Variablen. In: Faulbaum, F.: Datenanalyse mit SAS. Erscheint beim Gustav Fischer, Stuttgart.
- Küsters, U. und Arminger, G. (1989): Programmieren in GAUSS - Eine Einführung in das Programmieren statistischer und numerischer Algorithmen. Stuttgart.
- Maddala, G.S. (1983): Limited-dependent and Qualitative Variables in Econometrics. Cambridge.
- McCullagh, P. (1980): Regression models for ordinal data (with discussion). *Journal of the Royal Statistical Society B20*, 109- 142.
- McCullagh, P. und Nelder, J.A. (1983): Generalized Linear Models. London.
- McFadden, D. (1974): Conditional Logit Analysis of Qualitative Choice Behavior. In: Zarembka, P. (ed.): *Frontiers in Econometrics*. New York, 105-142.
- McFadden, D. (1981): Econometric Models of Probabilistic Choice. In: Manski, C.F. und McFadden, D. (Hrsg.): *Structural Analysis of Discrete Data with Econometric Applications*. Cambridge, Massachusetts, 198-272.
- Payne, C.D. (1985, Ed.): The GLIM System Release 3.77. Oxford, NAG-Royal Statistical Society.
- Röding, M., Küsters, U. und Arminger, G. (1985): KALOS-Programmbeschreibung - Ein interaktives Programmsystem zur Analyse kategorialer Logitmodelle. Köln, Zentralarchiv für empirische Sozialforschung.
- Savin, N.E. (1984). Multiple Hypothesis Testing. In: Griliches, Z. und Intriligator, M.D.: *Handbook of Econometrics - Volume II*. Amsterdam, 827-879.

- Schmidt, P. (1976). *Econometrics*. New York.
- Stewart, M.B. (1983): On the least squares estimation when the dependent variable is grouped. *Review of Economic Studies* L, 737 -753.
- Thisted, R.A. (1988): *Elements of Statistical Computing*. New York.
- Thome, H. (1989): *Grundkurs Statistik für Historiker, Teil I: Deskriptive Statistik*. *Historische Sozialforschung, Beiheft No. 2.*, 1 - 147.
- Tobin, J. (1958): Estimation of relationships for limited dependent variables. *Econometrica* 26, 24-36.

Dimensionen politischer Konflikte

Die Analyse von namentlichen Abstimmungen in Parlamenten mit dem Verfahren der Mokken-Skalierung

Heinrich Best u. Ernst J. Zimmermann

1. Methoden und Modelle der Analyse namentlicher Abstimmungen

Es ist ein häufig beschworenes Dilemma der quantifizierenden historischen Sozialforschung, daß sie auf "vorgefundene" Daten angewiesen ist und ihre Indikatoren typisch nicht aus direkten, durch den Forscher selbst kontrollierten Meßoperationen gewinnen kann. Diese Beschränkung hat weitreichende Konsequenzen. So werden Individualdaten in der historischen Sozialforschung häufig auf niedrigem Meßniveau erhoben, die von ihr ausgewerteten Massenquellen sind weit überwiegend Erzeugnisse "öffentlicher Buchführung", die eher institutionell definierte Rollen und formale Strukturen als die intentionalen und motivationalen Aspekte menschlichen Handelns abzubilden vermag - Ausnahmen bestätigen auch hier die Regel.¹

Das ist eine fatale Diagnose, und es ist verständlich, warum historische Sozialforscher mit erheblichem Einsatz bemüht sind, das hermetische Gehäuse formaler Strukturanalysen zu sprengen. Die Auswertung der mentalitätsgeschichtlich ertragreichen Massenquellen oder die Anwendung von Verfahren computerunterstützter Inhaltsanalyse auf große Textmengen eröffnen Auswege. Seltener werden statistische Modelle angewendet, die hinter der Fassade manifester Beobachtungswerte verborgene "latente Strukturen" und "konzeptuelle Schemata" historischer Akteure aufzudecken vermögen. Zu diesem Verfahren zählen neben anderen die "intrinsischen" oder "indirekten" Skalierungsmethoden, die "numerische Strukturen, gemäß modellspezifischen Annahmen, in andere numerische Strukturen transformieren".² Sie sind insbesondere dann anwendbar, wenn mangels eines externen Kriteriums die Gültigkeit der Messung eines latenten Konstrukts nur aus den Wechselbeziehungen zwischen den Indikatoren selbst erschlossen werden kann.

Die Voraussetzung für den Einsatz indirekter Skalierungsverfahren ist die

¹ Vgl. hierzu u.a. die Zusammenfassung der "qualitativen Kritik" und der "quantitativen Selbstkritik" in Konrad H. Jarausch et al., *Quantitative Methoden in der Geschichtswissenschaft* (Darmstadt 1985), S. 195-203.

² Erwin K. Scheuch u. Helmut Zehnpfennig, "Skalierungsverfahren in der Sozialforschung", in: *Handbuch der empirischen Sozialforschung*, Bd. 3a (3. Aufl., Stuttgart 1974), S. 99.

Überlieferung möglichst umfangreicher Serien gleichartiger "Indikatoren", "Stimuli" oder "Items". Sie ist in historischen Untersuchungen häufiger gegeben als gemeinhin angenommen wird. Ein besonders wichtiges Anwendungsgebiet ist das formale Abstimmungsverhalten in Organisationen, vor allem in rechtsprechenden und legislativen Körperschaften.³ Wird das Entscheidungsverhalten der einzelnen Mitglieder dabei in allen, zumindest aber in einer größeren Anzahl von Abstimmungen protokolliert und werden diese Voten nicht vollständig durch feste Fraktions- und Koalitionsbindungen prädeterniniert, dann ist es sinnvoll, nach "latenten" kognitiven oder normativen Konzepten zu suchen, die das Abstimmungsverhalten strukturierten. Tatsächlich zeigt sich in einschlägigen Untersuchungen regelmäßig, daß sich die beteiligten Akteure in wiederkehrenden Abstimmungssituationen nicht jedesmal "ad hoc" entscheiden, sondern mehr oder weniger prägnanten Abstimmungsmustern folgen. Die Frage, worin diese Regelmäßigkeiten begründet sind, wird dann allerdings sehr unterschiedlich und zum Teil kontrovers beantwortet.

Vor allem in den USA, Großbritannien, den Niederlanden und den skandinavischen Ländern war die Analyse namentlicher Abstimmungen unter der Bezeichnung "roll-call analysis" eines der ersten und ist nach wie vor eines der wichtigsten Anwendungsgebiete quantitativer Methoden in der Geschichtswissenschaft und in der historischen Soziologie. Sie wird eingesetzt, um die grundlegenden politischen Konfliktlinien in Parlamenten aufzudecken, den Einfluß externer Akteure auf legislative Entscheidungen zu identifizieren, den "Lebenszyklus" ideologischer Konflikte zu verfolgen und Prozesse der Koalitions- und Fraktionsbildung zu rekonstruieren.⁴ Die Bedeutung serieller Abstimmungsanalysen für die quantifizierende Geschichtsforschung hat William O. Aydelotte mit geradezu emphatischen Worten hervorgehoben. Seine sehr optimistische Bewertung des empirischen Gehalts und der Qualität von Abstimmungsdaten ist aber auch ein Beleg für die gelegentliche Naivität, in der sie vor allem in der Frühphase nach ihrer "Entdeckung" als Massenquelle verwendet wurden:⁵ "Yet there exists a source, a body of materials, ... which is

³ Vgl. u.a. Melissa P. Collie, "Voting Behaviour in Legislatures", in: Gerhard Loewenberg et al. (Hrsg.), *Handbook of Legislative Research* (Cambridge/Mass. u. London 1985), S. 471-518. Darin auch weitere Hinweise auf die umfangreiche amerikanische Literatur zur legislative behavior-Forschung.

⁴ Aage R. Clausen, "Some Basic Approaches to the Measurement of Roll-Call Voting", in: Donald F. Freeman (Hrsg.), *Foundation of Political Science* (New York u. London 1977), S. 313.

⁵ "Voting Patterns in the British House of Commons in the 1840's", in: *Comparative Studies in Society and History*, 5 (1962-1963), S. 134f.

not only readily available but is also fuller and more reliable than most sources which historians have at their disposal. This is the division lists. ... The peculiar value of these lists is that they contain expressions of opinion upon important subjects from men who are no longer available for questioning. It does not strain terms too much to say that, imaginatively used, they can constitute a kind of questionnaire which one may submit to the dead on which they will give us their replies. Further, the information about the votes of members of parliament is perhaps more complete and more certain than anything else that is known about them. The social background, their economic interests and their relations with their constituents can be studied with profit, matters the information is and always will be fragmentary and there are nuances which it is difficult now to recapture. By contrast, the stands which those men publicly adopted on the major issues of the day are documented by the division lists with a wealth of detail, repeated corroboration and, reason to believe, a relatively high degree of accuracy".

Es ist bemerkenswert, daß Verfahren der seriellen Abstimmungsanalyse in der Bundesrepublik Deutschland bislang noch kaum eingesetzt wurden. Die Gründe dafür können hier nicht ausführlich erörtert werden. Sie sind jedenfalls nicht in der geringeren Dichte der Überlieferung oder der minderen Relevanz von namentlichen Abstimmungen in den historischen Parlamenten West- und Mitteleuropas zu suchen. Ausschlaggebend war wohl der Umstand, daß angesichts der vor allem in der heutigen Bundesrepublik Deutschland hochdisziplinierten Parlamente kein Ausstrahlungseffekt von der gegenwartsbezogenen Forschung ausgeht, der historische Untersuchungen mit Abstimmungsdaten hätte stimulieren können.⁶ In dieser Hinsicht unterscheidet sich die Forschungssituation in der Tat grundlegend von der in den USA oder anderer europäischer Länder, in denen seit langem in historischen und gegenwartsbezogenen Untersuchungen Abstimmungsanalysen erfolgreich angewendet werden. Die sozialwissenschaftliche Relevanz eines Untersuchungsgegenstandes bestimmt sich jedoch nicht an seinem Gegenwartsbezug, sondern an seinem Gehalt verallgemeinerungsfähiger Erkenntnisse - wobei für Historiker selbst ohne solche Generalisierungsabsichten das Abgelebte ein würdiger Gegenstand des Interesses ist.

Im folgenden soll nun eine Forschungssituation vorgestellt werden, in der sich mit Abstimmungsanalysen Prozesse der Strukturbildung in Parlamenten

⁶ Ausführlich zur Forschungssituation und zur Methodik der Abstimmungsanalyse jetzt Heinrich Best, *Die Männer von Bildung und Besitz. Struktur und Handeln parlamentarischer Führungsgruppen in Deutschland und Frankreich 1848/49* (Düsseldorf 1990), S.484-531.

untersuchen lassen: "die Stunde Null" nach der Ausdehnung der Partizipationsrechte infolge von Revolutionen. Solche Versammlungen werden als Interaktionssysteme betrachtet, die noch ganz am Anfang ihrer Systemgeschichte stehen, in der sich noch wenig an systemspezifischen Institutionalisierungen, Erfahrungen und Erinnerungen, oder, um es mit Luhmann zu sagen: an "aufbewahrter Selektivität", abgelagert hat.⁷ Das Jahr 1848 war eine solche "Stunde Null", in der sich zum ersten Mal in Mittel- und Westeuropa das allgemeine Männerwahlrecht mit parlamentarischer Regierungsweise, einer Formierung von Parlamentsfraktionen, einer breiten politischen Mobilisierung der Bevölkerung und Anfängen organisierter Massenparteien verband. Auch die Abgeordneten der beiden konstituierenden Versammlungen, die im Mai 1848 in Paris und Frankfurt zusammentraten, waren, was die Parlamentserfahrung angeht, weit überwiegend homines novi: Nur jeweils etwa 20% der Mitglieder beider Konstituanten hatten schon vor dem Mai 1848 in einer Kammer oder einem Landtag gesessen.⁸

Hier interessiert nun vor allem die soziale Morphologie der Konfliktstrukturen, die sich in der Situation relativer Freiheit eines institutionellen Neubeginns ausdifferenzierten. Dazu wählten wir das Verfahren der seriellen Analyse namentlicher Abstimmungen, für das in beiden Untersuchungsgruppen die Anwendungsbedingungen günstig sind. So wurde in der Pariser Konstituante insgesamt 259 mal abgestimmt. Die Geschäftsordnung sah dies vor, wenn mindestens zwanzig Abgeordnete einen entsprechenden Antrag unterstützten. Obligatorisch waren namentliche Abstimmungen vor allem bei unmittelbar ausgabenwirksamen Beschlüssen. In der Frankfurter Nationalversammlung erfolgte eine offene Abstimmung durch Namensaufruf, wenn dies am Schluß der Debatte von mindestens fünfzig Abgeordneten verlangt wurde. Zwar war das Quorum hier strenger als in der Pariser Konstituante, doch wurde in der Paulskirche bei einer kaum längeren Sitzungsperiode noch häufiger - über insgesamt 298 Vorlagen - namentlich abgestimmt. Wie im Fall der französischen Nationalversammlung wurden die Voten der Abgeordneten protokolliert und anschließend in den offiziellen Sitzungsprotokollen publiziert. Insgesamt liefert das namentliche Abstimmungsverhalten für beide Untersuchungsgruppen mehr als 470000 Datenpunkte zum parlamentarischen Entscheidungshandeln der Abgeordneten.

⁷ "Weltzeit und Systemgeschichte. Über Beziehungen zwischen Zeithorizonten und sozialen Strukturen gesellschaftlicher Systeme", in: Peter Christian Ludz (Hrsg.), Soziologie und Sozialgeschichte (Opladen 1973), S. 85.

⁸ Best, Bildung und Besitz, S. 170.

Diese Fülle ist zwar eine besonders beeindruckende Widerlegung des Vorurteils von der relativen Datenarmut der historischen Sozialforschung, doch wäre es vorschnell zu glauben, daß man hier über einen unproblematischen Indikator verfügt, der einen ungebrochenen empirischen Zugang zu einer Analyse der politischen Präferenzen der Abgeordneten eröffnet. Tatsächlich sind namentliche Abstimmungen der Ausdruck eines hochkontingenten sozialen Handelns, in dem die ideologischen Präferenzen der Abgeordneten nur ein Bestimmungsmoment unter anderen sind. Abgeordnete sind nutzenmaximierende Akteure, die unter verfügbaren strategischen Alternativen wählen, um ihre Ziele zu erreichen. Junktimpositionen und Obstruktion sind Beispiele für taktisches Verhalten, die nahelegen, daß Abweichungen von der persönlichen Präferenzskala häufige Vorkommnisse in parlamentarischen Abstimmungen sind.⁹ Solchen Situationen ist gemeinsam, daß Abgeordnete einem Entscheidungskalkül folgen, das andere Akteure in ihrem Umfeld einbezieht, die ebenfalls unmittelbaren Einfluß auf das Ergebnis der Abstimmung haben. Deshalb gilt: während die Folgenlosigkeit der Antworten für die Befragten die Voraussetzung für die Gültigkeit von Ergebnissen in der Standard-Interviewsituation ist, sind die politische Wirkung und die absichtsvolle Sichtbarkeit des Abstimmungsverhaltens Kennzeichen von Entscheidungssituationen in parlamentarischen Versammlungen.

Für die Analyse des namentlichen Abstimmungsverhaltens resultiert hieraus das schwer lösbare Problem, die "strategische Komponente" in den Voten ermitteln zu müssen, wenn man aus ihnen Indikatoren für die politische Orientierung der Abgeordneten gewinnen will. Ein solches Vorgehen setzt die Konstruktion eines allgemeinen entscheidungstheoretischen Modells voraus, auf dessen Grundlage das Stimmverhalten von Abgeordneten unter der Voraussetzung vorhergesagt werden kann, daß sie entsprechend ihren persönlichen Präferenzskalen votieren. Abweichungen zwischen dem danach erwarteten Abstimmungsmuster und dem tatsächlich beobachteten Abstimmungsverhalten lassen dann auf eine strategische Komponente schließen. Erst in den vergangenen Jahren wurden Modelle parlamentarischen Abstimmungsverhaltens entwickelt, die die Elemente von Entscheidungssituationen und deren konditionelle Verknüpfung soweit spezifizieren, daß ein solches Vorgehen aussichtsreich wird.¹⁰ Zuvor waren Abstimmungsanalysen zumeist kaum mehr als beliebige Kombinationen eines Datentyps mit einem Methodenre-

⁹ J.A. Frerejohn u. M.P. Fiorina, "Purpose Models of Legislative Voting", in: *American Economic Review*, 65 (1975), S. 407.

¹⁰ Vgl. insbes. Menno Wolters, *Interspace Politics* (Leiden 1984).

pertoire, das nach Vorkenntnis und Vorliebe des einzelnen Forschers eingesetzt wurde.

Ausgehend von dem entscheidungstheoretischen Axiom, daß das zentrale Element einer Entscheidung die Bewertung eines Items nach einem Kriterium ist, wird Abstimmungsverhalten in Analogie zu Ansätzen moderner Theorien des Wählens als ein mehrdimensionales räumliches Modell konzipiert. Die Grundform des Kalküls bei einer Abstimmung ist danach die Bewertung der Distanz zwischen dem "idealen Punkt" des Abgeordneten auf seiner Präferenzskala und dem Punkt, den er einer Abstimmungsvorlage zuordnet. Die gerade noch tolerierte Abweichung vom Idealpunkt ist die kritische Distanz.

Während bei dem so modellierten Bewertungsprozeß die Entscheidung für oder gegen eine Vorlage ein interner Vorgang ist, wird bei der Berücksichtigung einer expliziten Alternative ein externes (kognitives) Kriterium eingeführt. Herbert Weisberg, auf den die hier skizzierte Unterscheidung zurückgeht, spricht von einem "cumulative scale model", wenn eine Alternative bei der Entscheidung berücksichtigt wird, von einem "proximity scale model", wenn dies nicht der Fall ist.¹¹ Der Unterschied zwischen beiden Entscheidungsmodellen ist insofern empirisch bedeutsam, als "proximity voting" die Möglichkeit einschließt, daß die Abgeordneten auf beiden Flügelpositionen übereinstimmend alle Abstimmungsvorlagen ablehnen, weil sie außerhalb des Bereichs ihrer kritischen Distanzen liegen. Dies ist dann der Fall, wenn die Abstimmungsgegenstände im mittleren Bereich des politischen Kontinuums lokalisiert sind, die Abgeordneten aber über die gesamte Spannweite des Konfliktraumes streuen. Das hat mißliche Folgen, wenn die politischen Positionen von Abgeordneten durch ihr Abstimmungsverhalten bestimmt werden sollen: "Les extrêmes se touchent", die Flügelpositionen werden sich dann nicht unterscheiden lassen, selbst dann, wenn alle Abgeordneten bei jeder Abstimmung ihren persönlichen Präferenzen entsprechend votieren und strategisches Abstimmungsverhalten nicht ins Spiel kommt. Wenn sich Abgeordnete zwischen expliziten oder impliziten Alternativen entscheiden, ist eine solche Konstellation hingegen ausgeschlossen.

Der Status quo ist die in der parlamentarischen Praxis häufigste, wenn auch fast immer implizite Alternative bei Abstimmungen. Bleibt der Status quo über Serien von Voten die relevante Entscheidungsalternative und markiert er zugleich einen der beiden Pole der Präferenzskalen aller Abgeord-

¹¹ "Scaling Models for Legislative Voting", in: *American Political Science Review*, 66 (1972), S. 1306-1315.

neten, ist der Sonderfall der "Unipolarität" gegeben. Das ist ein Umstand, der das Problem der Ermittlung von politischen Orientierungen durch Abstimmungsanalysen in entscheidender Weise vereinfacht. Denn unter der weiteren Voraussetzung, daß Abgeordnete und Abstimmungsgegenstände über den gleichen Bereich des politischen Kontinuums streuen - wofür in unserer Untersuchungssituation die Beobachtung spricht, daß die Vorlagen von den Abgeordneten selbst formuliert und eingebracht wurden - lassen sich dann Abstimmungen und Abgeordnete in eine korrespondierende Rangordnung bringen. Die resultierende Skala hätte idealiter die Eigenschaften einer Guttman-Skala.¹² Deren Meßmodell wurde von Stouffer prägnant beschrieben: "Es muß möglich sein, die einzelnen Indikatoren so zu ordnen, daß im Idealfall alle Personen, die eine gegebene Frage (hier: eine gegebene Abstimmungsvorlage) zustimmend beantworten, höhere Ränge einnehmen als Personen, die die gleiche Frage ablehnend beantworten. Aus dem Rang oder dem Skalenprodukt einer Person können wir exakt schließen, welche Indikatoren sie bevorzugten".¹³

In unserer Untersuchungssituation stellt sich nun die Frage, ob sich die namentlichen Abstimmungen einer oder mehreren kumulativen Skalen zuordnen lassen und wie sich die Abgeordneten auf die Skalenpositionen verteilen. Als Verfahren zur Rekonstruktion der parlamentarischen Konfliktstrukturen haben wir die Mokken-Skalierung eingesetzt. Es ist dies eine stochastische Version des Guttman-Modells, das ja von der rigorosen Annahme einer deterministischen Relation zwischen der empirisch zugänglichen Beobachtungsebene und der Konstruktebene ausgeht. Das ist eine wenig realistische Annahme, und Guttman-Skalen, die dem Kriterium folgen, daß einem Beobachtungspunkt ein und nur ein Punkt auf einem latenten Kontinuum entspricht, sind in der empirischen Sozialforschung dann auch extrem selten. "Latent-structure" Modelle wie die Mokken-Skalierung gehen demgegenüber davon aus, daß für die Abweichungen zwischen dem Guttman-Modell und der beobachteten Realität außer Mehrdimensionalität auch unscharfe Valenzen der Items verantwortlich sein können.¹⁴ Danach besteht zwischen einem Indikator und einer latenten Einstellungsdimension eine probabilistische Beziehung: Der Indikator wird nicht immer und unter allen Bedingungen die gleiche Information über das Konzept liefern. Die Mokken-

¹² In M. Wolters verzweigter Systematik von Abstimmungsmodellen entspricht die Guttman-Skala dem "unipolar unidimensional individualized fixed alternative model". *Interspace Politics*, S. 50.

¹³ Samuel A. Stouffer, *Measurement and Prediction*, Princeton 1950, S. 5,

¹⁴ C.H. Coombs, *A Theory of Data* (2. Aufl. New York u.a. 1967), S. 280.

Skalierung läßt sich dementsprechend als eine Variante des Guttman-Modells mit einem freien Parameter formalisieren.

Nur am Rande sei vermerkt, daß wahrscheinlichkeitstheoretische Modellierungen dieser Art mit Theorien unscharfer Mengen eng verwandt sind, die heute in Expertensystemen und speziell in der Entwicklung von Verfahren "quellennaher" Datenverarbeitung für Historiker eine große Rolle spielen.¹⁵ Sucht man nach geistesgeschichtlichen Anknüpfungspunkten für das Wirklichkeitsbild, das probabilistischen Skalierungsverfahren zugrundeliegt, dann kann man bis zu den Kategorien des "Wesentlichen" und des "Zufälligen" oder von "Wirklichkeit" und "Erscheinung" in der idealistischen Philosophie zurückgehen.

2. Das Verfahren der Mokken-Skalierung

Bevor wir das Verfahren der Mokken-Skalierung im folgenden näher beschreiben, ist zu klären, was unter dem Begriff der "Skala" verstanden werden soll. Ganz allgemein ist eine Skala ein Meßinstrument zur Erfassung der Ausprägungen einer bestimmten Eigenschaft von Objekten. Der Meßvorgang als solcher besteht dabei in der Zuordnung von Skalenwerten zu den Objekten, wobei die Relation der Skalenwerte der Relation der Objekte entsprechen soll. Vielfach wird in diesem Zusammenhang die Welt der Objekte auch als "empirisches Relativ" bezeichnet, während man die Welt der Skalenwerte "numerisches Relativ" nennt.¹⁶

Der Begriff des "numerischen Relativ" ist jedoch insofern problematisch, als Zahlen nicht notwendig die einzige Möglichkeit zur Darstellung von Skalenwerten sind. Im Falle von "Nominalskalen" ist die Verwendung von Zahlzeichen (Ziffern) zur Darstellung von Skalenwerten sogar irreführend - suggeriert sie doch Eigenschaften, die in diesem Fall den Skalenwerten nicht zukommen: Codiert man z.B. "politisch rechts" mit 1 und "politisch links" mit 2, so gilt zwar - wenn man 1 und 2 nicht als Ziffern, sondern als Zahlen auffaßt - daß 2 größer ist als 1, nur hat diese Relation kein empirisches Äquivalent, da eine Aussage "links" größer als "rechts" sinnlos ist. Noch absurder wäre eine Aussage wie: $2 \cdot \text{rechts} = 1 \cdot \text{links}$, obwohl $2 \cdot 1 = 1 \cdot 2$ arithmetisch sicherlich richtig ist. "Rechts" und "links" sind in diesem Zusammenhang lediglich zwei disjunkte Klassen von Objekten, die durch zwei beliebige verschiedene Sym-

¹⁵ Peter K. G. Günther, Die wahrscheinlichkeitstheoretischen Grundbestände der Testtheorie und verwandter formaler Gebilde, so auch der Latenten-Struktur-Theorie. Phil.Diss., Univ. Bremen, FB Psychologie, in Vorb.

¹⁶ Bernhard Orth, Einführung in die Theorie des Messens, 1974.

bole repräsentiert werden können. Die Verwendung zweier Ziffern ist hier nur eine von vielen Möglichkeiten - und mit Sicherheit eine problematische.

Bei der Messung von z.B. Längen oder Gewichten ist allerdings die Verwendung von Zahlen (dargestellt durch Ziffern) geradezu geboten, da sie die Eigenschaften des empirischen Relativs sehr gut abbilden können. Da hier sogar Aussagen wie "doppelt so lang" oder "halb so schwer" getroffen und repräsentiert werden können, spricht man in diesen Fällen von einer Verhältnis- oder Ratioskala. Wesentliches Merkmal von Ratioskalen ist, daß der Zahl 0 jegliches Fehlen der betreffenden Eigenschaft (z.B. Länge) zugeordnet ist. Hierin unterscheidet sich die Ratioskala von der Intervallskala. Für letztere gilt, daß zwar gleiche Abstände (Differenzen) zwischen zwei Skalenwerten (z.B. $3 - 1 = 4 - 2$) gleich große Unterschiede im empirischen Relativ repräsentieren, Verhältnisaussagen wie "doppelt so..." oder "halb so..." hingegen nicht aus den Skalenwerten abgeleitet werden können.

Ein Beispiel für eine Intervallskala ist die Temperaturskala nach Celsius. Eine Aussage wie: "Heute ist es doppelt so warm wie gestern" aufgrund der Messung von 4°C und 2°C ist nicht haltbar, da dies impliziert, daß bei 0°C keine Temperatur vorläge. "Keine Temperatur" aber ist - wie Chemie und Physik lehren - gleichbedeutend mit "keine Molekularbewegung" und dies entspricht etwa -273°C (oder 0°Kelvin - womit die Kelvin-Skala zu einer Ratioskala wird).

Sehen wir uns den Vorgang der Messung am Beispiel der Temperaturbestimmung mittels Thermometer etwas genauer an, so stellen wir fest, daß ein Thermometer im Grunde eine Längenskala aufweist, über die die Ausdehnung von Quecksilber erfaßt wird und von der auf das Ausmaß der Molekularbewegung - also die Temperatur - rückgeschlossen wird. Manifest und damit tatsächlich meßbar ist also nicht die Temperatur, sondern lediglich die Ausdehnung von z. B. Quecksilber, von der auf das Ausmaß eines hierfür verantwortlichen "latenten" Phänomens der Molekularbewegung geschlossen wird. Dies ist eine Situation, die in den empirischen Sozialwissenschaften vertraut ist. Ein Beispiel ist der Gegenstand dieses Beitrags: die politische Orientierung von Abgeordneten in historischen Parlamenten. Diese läßt sich nicht direkt erfassen, sondern kann nur aus manifestem Abstimmungsverhalten erschlossen werden. Politische Orientierung wird dabei als latente, das manifeste Verhalten determinierende Größe, als ein hypothetisches Konstrukt angesehen. Erschwerend ist im gegebenen Fall, daß das manifeste Abstimmungsverhalten durchaus auch andere Determinanten haben kann, wie etwa taktische Kalküle. Das Verhalten bei nur einer einzigen Abstimmung als Indikator für die politische Orientierung des jeweils Abstimmenden zu be-

trachten wäre deshalb genauso gefährlich, wie aus dem Verhalten mehrerer Personen bei nur einer Wahl auf die Determinanten des Wahlverhaltens schlechthin zu schließen. Zur Untersuchung der Determinanten des Abstimmungsverhaltens ist es also notwendig, mehr als eine Abstimmung einer Reihe von Abgeordneten heranzuziehen. Aber: Ist eine solche Sammlung von Abstimmungen *eo ipso* eine Skala?

Gibt man sich damit zufrieden, die Voten einer jeden Person in einem Antwortmustervektor zu protokollieren, diese untereinander auf Gleichheit zu prüfen und dabei lediglich "Vektoren gleich" bzw. "Vektoren ungleich" festzuhalten, so bildet die Menge der möglichen Antwortmustervektoren sicherlich eine (sehr umfangreiche) Nominalskala, nur ist der aus deren Anwendung zu erwartende Erkenntnisgewinn eher gering.

Will man nachweisen, daß politische Orientierungen wesentliche Determinanten des Abstimmungsverhaltens waren, ist es notwendig, zunächst aus der Gesamtmenge der Indikatoren (Abstimmungen) jene zu einem "Metaindikator" zusammenzufassen, die das hypothetische Konstrukt erfassen. Die "Metaindikatoren" sollen dabei den Charakter einer Dimension haben, entlang derer sich die konstituierenden Einzelindikatoren in einer eindeutigen Reihenfolge anordnen lassen. Gemeinsame Eigenschaften der Einzelindikatoren, hier also der Abstimmungsvorlagen, sollen dann zur inhaltlichen Identifikation der Dimension(en) herangezogen werden. Ferner soll die Position, die eine Person auf einer Dimension inne hat, eindeutig aus ihrem Antwortmustervektor bestimmt werden können, wobei dieser aus den Voten zu den jeweils konstituierenden Einzelindikatoren nach einer eindeutigen Vorschrift zusammengesetzt wird.

Diese Beschreibung der Aufgabe legt es nahe, hauptkomponentenanalytische Ansätze - wie die Faktorenanalyse - oder Verfahren aus dem Repertoire der Multidimensionalen Skalierung (MDS) heranzuziehen. Die Hauptkomponentenanalyse erfordert aber - da in ihr Datenwerte additiv verknüpft werden - Eingangsdaten von zumindest Intervallskalen-Niveau, und es wird - da dieses Verfahren lediglich die Parameter Varianz und Mittelwert berücksichtigt - Normalverteilung der Daten vorausgesetzt. Die vorliegenden Daten aber sind allenfalls ordinal (Ablehnung/Enthaltung/Zustimmung) - d.h. die Datenwerte einer jeden Abstimmung können zwar eindeutig je einer von drei Kategorien zugeordnet werden, welche ihrerseits in eine natürliche Ordnung zu bringen sind (sofern "Enthaltung" als eindeutiger Indikator für Indifferenz angesehen werden kann), eine Aussage über die Größe der Abstände

"Ablehnung - Enthaltung", "Enthaltung - Zustimmung" und "Ablehnung - Zustimmung" ist jedoch nicht möglich. Von daher verbietet sich streng genommen jedwede additive Verknüpfung von Datenwerten und damit die Berechnung von Mittelwerten und Varianzen. Zur Illustration: Eine Codierung Ablehnung = - 1, Enthaltung = 0, Zustimmung = + 1 bildet die uns bekannten Gegebenheiten im empirischen Relativ genauso gut oder schlecht ab wie eine Codierung Ablehnung = - 1, Enthaltung = 5, Zustimmung = 10 000.

Man könnte also auch sagen, unser Meßinstrument ist für den Einsatz hauptkomponentenanalytischer Verfahren zu ungenau. Hier liegt dann aber auch der Ansatzpunkt für eine heuristische Begründung, nach der diese Verfahren doch eingesetzt werden können: Codiert man nämlich Enthaltung = x , Ablehnung = $x-y$, Zustimmung = $x+y$ und setzt man voraus, daß Ablehnung ungefähr genauso weit von Enthaltung entfernt ist wie Zustimmung (Betrag) und betrachtet man die durch die Grobheit des Meßinstrumentes verursachten Fehler in den Einzelwerten als normalverteilte Zufallsgröße, so kann man den Einsatz von z.B. hauptkomponentenanalytischen Verfahren rechtfertigen.

Die Multidimensionale Skalierung wiederum wirft insbesondere das Problem auf, daß hier eine sog. Stressfunktion minimiert werden muß, die u.U. neben einem globalen Minimum etliche lokale Minima aufweisen kann, in welchen sich der Minimierungsalgorithmus "verfangen" kann, ohne daß dies feststellbar wäre.

Benötigt wird aufgrund der vorliegenden Datenqualität ein Verfahren, das es erlaubt, aus ordinalen Eingangsdaten Metaindikatoren zu bilden, ohne dabei Annahmen über die Meßeigenschaften der Daten zu überfordern oder Probleme verfahrenstechnischer Imponderabilien aufzuwerfen. Ein solches Verfahren ist die Skalenanalyse nach Mokken,¹⁷ die wiederum aufgefaßt werden kann als eine probabilistische Verallgemeinerung der Skalogrammanalyse nach Guttman.¹⁸ Im folgenden wollen wir zunächst kurz auf die Guttman-Skalogrammanalyse für dichotome Daten eingehen, von dieser dann auf die Mokken-Skalierung dichotomer Daten übergehen und diese schließlich auf polytome Daten ausweiten.

Nehmen wir also zunächst an, wir hätten es mit dichotomen Voten zu tun - es gebe also nur die Kategorien "Ablehnung = 0" und "Zustimmung = 1". Gäbe es nun nur eine einzige das Abstimmungsverhalten determinierende

¹⁷ R.J. Mokken, A Theory and Procedure of Scale Analysis (Den Haag 1971).

¹⁸ Vgl. Anm. 13.

latente Dimension, so müßten sich die Gesetzesvorlagen genauso wie die Abstimmenden entlang dieser Dimension in eine eindeutige Reihenfolge bringen lassen. Dies könnte wie folgt geschehen: Zunächst wird die Matrix aller aufgetretenen disjunkten Antwortmustervektoren aufgestellt, wobei diese Vektoren die Zeilen der Matrix bilden, so daß für jede Gesetzesvorlage genau eine Spalte existiert (vgl. Abb. 1a).

Abbildung 1a: Hypothetisches Abstimmungsmuster

		Vorlagen					
		1	2	3	4	5 ...	
M	1	0	1	1	1	0 ...	(1: Zustimmung, 0: Ablehnung)
u	2	0	1	0	1	0 ...	
s	3	0	0	0	1	0 ...	
t							
e	4	1	1	1	1	0 ...	
r	5	1	1	1	1	1 ...	
						
						

Kann nun eine Zeilen- und Spaltenpermutation wie in Abb. 1b dargestellt gefunden werden, so liegt eine perfekte Guttman-Skala vor. Die Skalenposition einer Gesetzesvorlage entspricht ihrer neuen Spaltennummer, die eines Abstimmenden der neuen Zeilennummer seines Antwortmustervektors in der permutierten Matrix.

Abbildung 1b: Hypothetisches Abstimmungsmuster bei perfekter Guttman-Skala

		Vorlagen					
		1	2	3	4	5	
		(4)	(2)	(3)	(1)	(5)	In Klammern: Zeilen- und Spaltennummern von Tab. 1a
M	1 (3)	1	0	0	0	0	
u	2 (2)	1	1	0	0	0	
s	3 (1)	1	1	1	0	0	
t							
e	4 (4)	1	1	1	1	0	
r	5 (5)	1	1	1	1	1	
						
						

Die Praxis zeigt jedoch, daß eine perfekte Guttman-Skala eher die Ausnahme ist - d.h. selbst die beste aller Permutationen resultiert in einer Matrix, die Abweichungen von obigem Schema enthält. Die Skala ist dann nicht perfekt. Definiert man z.B.

$$\text{REP} = \frac{E}{N * I} \quad (\text{Gleichung 1})$$

mit: E - Anzahl der Fehler (1 statt 0, 0 statt 1)
 N - Anzahl der Antwortmustervektoren (Abstimmende)
 I - Anzahl der Gesetzesvorlagen

so gelangt man zu einem möglichen Maß für die Güte einer Skala. Dieses sollte - einem Vorschlag von Guttman folgend - nicht kleiner als 0.90 sein (entspr. max. 10% Fehler über alle Vorlagen und Abstimmenden hinweg). Inzwischen wurden neben dem REP-Koeffizienten weitere Gütemaße für Guttman-Skalen entwickelt, ohne daß dies jedoch die Logik des Verfahrens verändert.

Die Skalogrammanalyse nach Guttman kann also wie folgt zusammengefaßt werden:

1. Bilde alle möglichen Spalten- und Zeilen- Permutationen der Matrix der disjunkten Antwortmustervektoren und wähle diejenige aus, für die der REP-Koeffizient maximal ist.
2. Prüfe, ob REP größer als 0.90 ist.
 Wenn ja: Fertig
 Wenn nein: Eliminiere eine Spalte und gehe zu 1.

Auch wenn über die Einführung beispielsweise des REP-Koeffizienten ein gewisses Maß an Fehlern zugelassen wird, geht dieser Ansatz von Guttman doch davon aus, daß die latente Eigenschaft das manifeste Verhalten eindeutig determiniert. Damit ist dieser Ansatz den deterministischen Verfahren zuzuordnen. Dieser Klasse stehen die probabilistischen Verfahren gegenüber, denen ein Denkmodell zugrunde liegt, demzufolge mit steigendem Ausmaß der latenten Eigenschaft die Wahrscheinlichkeit für ein bestimmtes manifestes Verhalten monoton steigt. Diese Wahrscheinlichkeit kann sich beliebig nahe an 1 heranbewegen, erreicht 1 jedoch nie ganz, so daß immer eine modellimmanente Möglichkeit für Abweichungen besteht, wiewohl diese Möglichkeit beliebig klein sein kann.

Als probabilistische Form der Guttman Skalogrammanalyse wird die Skalenanalyse nach Mokken angesehen, die umfassend erstmals 1971 vorgestellt

wurde. Einzug nimmt die Probabilistik hier durch die Verwendung des H-Koeffizienten als Maß für die Güte einer Skala. Der Algorithmus des Verfahrens kann wie folgt umrissen werden:¹⁹

- 1.: Suche die beiden "homogensten" Abstimmungen der (Rest-) Gesamtheit aller (noch nicht skalierten) Abstimmungen.
 - 1a: Prüfe, ob für diese überhaupt von "Homogenität" gesprochen werden kann.
 Wenn nein: Keine (weitere) Skalenbildung möglich. Fertig.
 Wenn ja: Betrachte die beiden Abstimmungen als "Anfang" einer Skala und
- 2.: Versuche, die Skala um eine Abstimmung zu erweitern.
 - 2a: Schließe alle Abstimmungen für diese Skala endgültig aus, die einen negativen "Homogenitätskoeffizienten" mit der soweit gebildeten Skala aufweisen.
 - 2b: Prüfe, ob es Items gibt, die zur bisherigen Skala "homogen" sind.
 Wenn nein: Keine Erweiterung mehr möglich.
 Skala fertig. Gehe zu 4.
 Wenn ja: Weiter bei 3.
- 3.: Wähle von den zur bisherigen Skala "homogenen" Abstimmungen diejenige aus, die zur "homogensten" erweiterten Skala führt. Gehe zu 2.
- 4.: Versuche, aus den noch nicht skalierten Abstimmungen eine weitere Skala zu bilden. Gehe wieder zu 1.

In dieser Beschreibung des Skalierungsalgorithmus' nach Mokken tritt wiederholt der Begriff der Homogenität auf. Diese wird definiert und quantifiziert durch Loevingers H-Koeffizienten,²⁰ der für zwei dichotome Variablen (Abstimmungen) i und j wie folgt dargestellt werden kann: Seien die beiden Abstimmungen kodiert als: 0 = Ablehnung, 1 = Zustimmung und werde der Vorlage i seltener zugestimmt als der Vorlage j . Dann lassen sich die zugehörigen (kombinierten) Zustimmungswahrscheinlichkeiten in der in Abb. 2 dargestellten Form in einer Vier-Felder-Tafel anordnen.

¹⁹ Vgl. Anm. 17.

²⁰ J. Loevinger, A systematic approach to the construction and evaluation of tests of ability. Psychological monographs, 61 (4).

Abbildung 2: Kombinierte Zustimmungswahrscheinlichkeiten
zweier Abstimmungen
Vorlage j

		0	1		
Vorlage i	0	p_{00}	p_{10}	$p_{1.}$	mit: $p_{1.} \leq p_{0.}$
	1	p_{10}	p_{11}	$p_{1.}$	
		$p_{.0}$	$p_{.1}$	$p_{..}$	

H_{ij} ist dann definiert als

$$H_{ij} = 1 - \frac{p_{10}}{p_{1.} * p_{.0}} \quad (\text{Gleichung 2})$$

oder - bei Schätzung der p_{ij} aus den Zellenhäufigkeiten n_{ij} der Stichprobe

$$H_{ij} = 1 - \frac{n_{10}}{(n_{1.} * n_{.0}) / n_{..}} \quad (n_{..} = N) \quad (\text{Gleichung 2'})$$

Nebenbei bemerkt entspricht dieser Koeffizient sowohl

$$PHI_{terr} = \frac{PHI}{PHI_{...}} \quad (\text{Gleichung 2''})$$

als auch dem DEL -Koeffizienten nach Hildebrand et al. mit (1;0) als Fehlerzelle²¹. Bei der Quantifizierung von Homogenität ist es nahezu zwingend, (1;0) als Fehlerzelle zu betrachten, da unter der Voraussetzung $p_{1.} \leq p_{0.}$ diese Zelle all jene Abstimmenden enthält, die der Vorlage mit der geringeren Annahmewahrscheinlichkeit zugestimmt haben, obwohl sie der Vorlage mit der global höheren Annahmewahrscheinlichkeit ihre Zustimmung entzogen haben - eine Konstellation, die bei eindeutiger Anordnung von Vorlagen und Abstimmenden entlang einer einzigen Dimension nicht auftreten dürfte. (Bilden zwei Vorlagen eine perfekte Guttman-Skala, so ist $n(1;0) = 0$ und H_{ij} wird 1).

²¹ Georg Rudinger, Friedrich M. Chaselon, Hans J. Henning, E. J. Zimmermann, Qualitative Methoden - Neue Wege sozialwissenschaftlicher Methodik, München 1985.

Anders betrachtet wird im H-Koeffizienten für zwei Items i und j, $p_{1.} \leq p_{.1}$, der Erwartungswert für die Fehlerwahrscheinlichkeit $p(1;0)$ in Beziehung gesetzt zu deren Erwartungswert unter der Nullhypothese (bei Vorhersage von $p(1;0)$ aufgrund der bloßen Kenntnis der Randverteilungen). Bezeichnet man diese mit E_{ij} bzw. mit EO_{ij} , so ist

$$H_{ij} = 1 - \frac{E_{ij}}{EO_{ij}} \quad (\text{Gleichung 3})$$

Hier wird deutlich, daß der H-Koeffizient 0 wird, wenn E_{ij} und EO_{ij} gleich sind, wenn also der Erwartungswert der Fehlerwahrscheinlichkeit unter der Homogenitätshypothese ($p(1;0)$) gleich deren Erwartungswert unter der Nullhypothese ($p(1;.) \cdot p(.;0)$) ist, so daß deren Quotient 1 ergibt. Es bleibt jedoch die Frage, inwieweit sich ein H-Koeffizient > 0 statistisch signifikant von 0 unterscheidet. Dies läßt sich auf die Frage reduzieren, inwieweit sich

$$E_{ij} - EO_{ij} = p_{10} - p_{1.} \cdot p_{.0}$$

signifikant von 0 unterscheidet. Hierzu konnte Mokken zeigen, daß folgende Teststatistik sich für steigende Stichprobengröße $N = n..$ einer Normalverteilung nähert:²²

$$\Delta^* = \frac{(p_{10} - p_{1.} \cdot p_{.0}) - 0}{S^2(p_{10})} = \frac{p_{10} - p_{1.} \cdot p_{.0}}{S^2(p_{10})} \quad (\text{Gleichung 4})$$

Bei Schätzung der Parameter aus einer Stichprobe ergibt sich hieraus:

$$\Delta^* = \frac{n_{10} - (n_{1.} \cdot n_{.0}) / N}{S^2(n_{10})} \quad (\text{Gleichung 4'})$$

Gleichung 2' und Gleichung 4' liefern uns alle Hilfsmittel zur Durchführung von Schritt 1 und Schritt 1a des o. a. Ablaufschemas der MOKKEN-Skalenanalyse:

- Die beiden homogensten Vorlagen sind diejenigen mit dem höchsten (positiven) H_{ij} -Koeffizienten (1.)

²² Vgl. Anm. 17.

- Dieser kann auf statistische Bedeutsamkeit (signifikante Abweichung von 0) geprüft werden (1a.)²³

Für die übrigen Schritte fehlen jedoch noch

- die Definition der Homogenität einer Abstimmung mit einer (Teil-)Skala (H_j)
- die Definition der Homogenität einer ganzen Skala (H) nebst deren Signifikanzprüfungen.

H wird definiert als

$$H = 1 - \frac{E_{..}}{E0_{..}} \quad (\text{Gleichung 5})$$

- Mit: $E_{..}$ - Summe der Fehlerwahrscheinlichkeiten $p(1;0)$ aller H_{ij} , $i < j$ der nach Zustimmungswahrscheinlichkeiten absteigend geordneten Vorlagen der Skala.
- $E0_{..}$ - Summe der zugehörigen Erwartungswerte unter der Nullhypothese.

H_j wird analog definiert zu:

$$H_j = 1 - \frac{E_{1.}}{E0_{1.}} \quad (\text{Gleichung 6})$$

- Mit: $E_{1.}$ - Summe der $p(1;0)$ aller H_{ij} , $i \neq j$
- $E0_{1.}$ - Summe der zugehörigen Erwartungswerte unter der Nullhypothese.

Auch für diese lassen sich approximativ normalverteilte Teststatistiken DELTA* ableiten, die einen Test auf signifikante Abweichung von 0 erlauben. In diesem Zusammenhang muß jedoch betont werden, daß eine signifikante Abweichung von Null lediglich besagt, daß die Wahrscheinlichkeit dafür, daß keine Homogenität vorliegt, sehr gering ist. Dies ist jedoch keinesfalls gleichbedeutend mit dem Vorliegen einer brauchbaren Skala. Mokken schlägt daher vor, a priori einen Mindest-Skalierbarkeitskoeffizienten für die zu findenden Skalen festzulegen.

²³ Da etliche simultane Tests durchgeführt werden, ist eine entsprechende Adjustierung des α -Niveaus (etwa durch $\alpha^* = \alpha/N$ Tests) erforderlich.

Hierzu klassifiziert er:

$H < 0.30$	"keine Skala"
$0.30 \leq H < 0.40$	"schwache Skala"
$0.40 \leq H < 0.50$	"mittelmäßige Skala"
$0.50 \leq H$	"starke Skala"

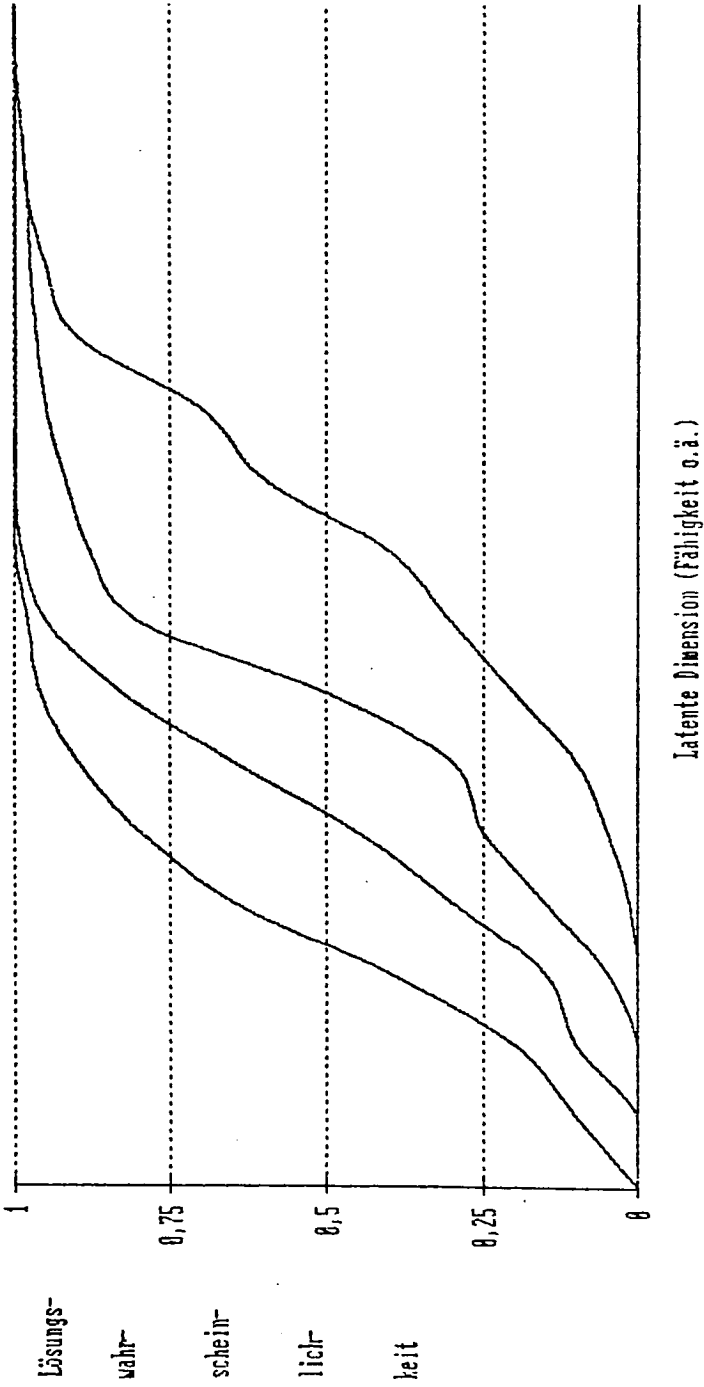
Allerdings wäre nach dem bislang Dargestellten eine solche Skala ungeachtet des H-Koeffizienten im Grunde nur ein Cluster homogener Variablen und nichts weiter. Um von einem Cluster zu einer Skala zu kommen, muß eine Ordnung in die konstituierenden Variablen - hier also in die Abstimmungen - gebracht werden, wobei sich als Ordnungskriterium das Komplement der Wahrscheinlichkeit für die 1-Kategorie - hier also der Zustimmungswahrscheinlichkeit für die Vorlagen - anbietet. Dabei gehen implizit folgende Annahmen ein:

1. Alle konstituierenden Variablen des Clusters erfassen die gleiche Eigenschaft.
2. Trägt man in einem Diagramm die Wahrscheinlichkeiten für die 1-Kategorien der konstituierenden Variablen gegen das Ausmaß der erfaßten Eigenschaft ab, so überschneiden sich die Funktionsgraphen der einzelnen Variablen (deren sog. "trace lines") nicht. D.h.: Für jede gegebene Wahrscheinlichkeit außer 0 und 1 der 1-Kategorie (hier: Zustimmungswahrscheinlichkeit) ist die Rangreihe der konstituierenden Variablen bezüglich des Ausmaßes der gemessenen Eigenschaft die gleiche (vgl. Abb. 3).

Sind diese beiden Annahmen haltbar und ordnet man die konstituierenden Variablen aufsteigend nach dem Komplement der Wahrscheinlichkeit der 1-Kategorie ($1-p_1$) bzw. absteigend nach p_1 , so ist der Skalenwert eines Untersuchungsobjektes identisch mit der Position der letzten 1 in seinem solchermaßen geordneten Antwortmustervektor.

Eine zweite Möglichkeit der Ermittlung des Skalenwertes eines Untersuchungsobjektes besteht darin, die Anzahl der Einsen im jeweiligen Antwortmustervektor auszuzählen. Im Idealfall - nämlich dann, wenn der geordnete Antwortmustervektor die Form $(1, \dots, 1, 0, \dots, 0)$ hat - sind die Ergebnisse beider Verfahren identisch. Bei Abweichungen vom Idealfall werden im ersten Verfahren die Skalenwerte eher überschätzt, während das zweite Verfahren eher unterschätzte Werte liefert.

Abbildung 3



Überträgt man diese beiden Vorschriften zur Ermittlung von Skalenwerten auf die vorliegenden Abstimmungsdaten, so können sie in entsprechender Terminologie wie folgt formuliert werden:

- a) Der Skalenwert eines Abgeordneten entspricht der Position der letzten Eins in seinem nach Zustimmungswahrscheinlichkeit der konstituierenden Vorlagen absteigend geordneten Abstimmungsmustervektor.
- b) Der Skalenwert eines Abgeordneten entspricht der Anzahl der Zustimmungen zu den konstituierenden Vorlagen.

Beide Verfahren liefern Skalenwerte auf ordinalem Meßniveau. Im Fall vieler fehlender Datenwerte liefern beide Verfahren in toto stark überschätzte Skalenwerte. Hier empfiehlt es sich, das zweite Verfahren dahingehend zu modifizieren, anstelle der einfachen Auszählung der Einsen deren relative Häufigkeit bezogen auf die Gesamtzahl der jeweils in einem Antwortvektor tatsächlich vorhandenen Daten zu benutzen.

In jedem Fall aber sollten die beiden o.a. impliziten Annahmen geprüft werden. Durch das Verfahren der Mokken-Skalierung ist automatisch gewährleistet, daß alle konstituierenden Variablen einer Skala die gleiche Eigenschaft oder ein "homogenes Korrelat" erfassen. Eine Unterscheidung von Eigenschaft und Korrelaten ist verfahrenstechnisch nicht möglich und kann daher nur inhaltlich-interpretativ erfolgen. Es empfiehlt sich daher, von vorneherein nur solche Variablen zu berücksichtigen, die unter einem inhaltlichen Gesichtspunkt gesehen vermutlich die gleichen latenten Dimensionen zu erfassen erlauben.

Die zweite der beiden o.a. impliziten Annahmen wird als doppelte Monotonie der trace lines (monoton steigend und sich nicht schneidend) bezeichnet. Zur Prüfung dieser Annahme genügt es nach Mokken,²⁴ die sog. P^{++} - und P^{-} -Matrices zu inspizieren. Bei der P^{++} -Matrix handelt es sich um die Matrix der p_{11} -Wahrscheinlichkeiten aller konstituierenden Variablen einer Skala. Die p_{11} -Wahrscheinlichkeit zweier Variablen i und j ist die Wahrscheinlichkeit für ein (1; 1) - Datenwertepaar in diesen beiden Variablen - in unserem Falle also die Wahrscheinlichkeit, in zwei Abstimmungen i und j gleichermaßen mit "ja" zu stimmen. Die P^{-} -Matrix ist analog die Matrix der Wahrscheinlichkeiten für (0;0)- Datenwertepaare. Ist nämlich doppelte Monotonie der trace lines gegeben und ordnet man Spalten und Zeilen dieser

²⁴ Vgl. Anm. 17.

beiden Matrices entsprechend der Wahrscheinlichkeiten für 1-Kodierungen in den Zeilen- und Spaltenvariablen absteigend an, so muß sich in der P^{++} -Matrix ein spalten- und zeilenweise monoton fallender Trend zeigen, während in der P^- -Matrix ein spalten- und zeilenweise monoton steigender Trend auftreten muß. Trendverletzungshäufigkeiten in einer Größenordnung von 3% und kleiner betrachtet Mokken dabei als irrelevant.²⁵

Ein wesentlicher Nachteil der Mokken-Analyse in der bislang dargestellten Form ist ihre Beschränkung auf dichotome Variablen. Diese Beschränkung wurde durch die Arbeiten von Molenaar aufgehoben.²⁶ Molenaar geht davon aus, daß mehrstufige (polytome) Variablen mit $M+1$ Kategorien als per se homogene "Mini-Skalen" bestehend aus M Schritten aufgefaßt werden können. Voraussetzung hierfür ist natürlich, daß die einzelnen Variablen mindestens ordinales Meßniveau aufweisen. Für K homogene Variablen mit jeweils M_k Schritten ergeben sich hieraus $\sum M_k$ Skalenschritte in der Gesamtskala - bildlich gesprochen durch "Verschachteln" oder "Ineinanderschieben/Aneinanderhängen" der Schritte der Einzelvariablen. Die Reihenfolge der Einzelvariablenschritte $S_{1,1-3}$, $S_{2,1-3}$ und $S_{3,1-3}$ dreier dreistufiger Variablen könnten z. B. wie folgt entlang einer gemeinsamen latenten Dimension angeordnet sein:

$$S_{1,1} S_{2,1} S_{2,2} S_{1,2} S_{2,3} S_{3,1} S_{3,2} S_{3,3} S_{1,3}$$

Allerdings sind die Schritte innerhalb einer einzigen Variablen keineswegs unabhängig voneinander. Würde man die Einzelvariablenschritte m und $m-1$ in zwei dichotome Dummy-Variablen codieren und diese in einer Vierfeldertafel gegenüberstellen, so könnte es nie vorkommen, daß die Zelle ($S_m = 1; S_{m-1} = 0$) eine Besetzung aufweist, da der ordinale Charakter der Variablen impliziert, daß immer dann, wenn Schritt m innerhalb einer Variablen vollzogen wurde (also die $(m+1)$ -te Stufe erreicht ist) auch alle vorangegangenen Schritte vollzogen wurden (also alle vorangegangenen Stufen bereits überschritten sind). Molenaar nennt diese Zellen daher folgerichtig "logisch unmögliche Fehlerzellen". Aus diesem Grunde aber ist es nicht möglich, eine Variable mit M_k Schritten einfach durch M_k dichotome Dummy-Variablen zu ersetzen und anschließend den o.a. Algorithmus für dichotome Variablen auf die Gesamtheit aller Dummy-Variablen anzusetzen. Betrachtet man dieses Problem vor dem Hintergrund der Arbeiten von Hildebrand et al. zum DEL-Koeffizienten, so reduziert sich dieses auf die Festlegung der Fehlerzel-

²⁵ Ebd., S. 277.

²⁶ Ivo W. Molenaar, Mokken scaling revisited, *Kwantitative Methoden* (3), 1982, S.145-164.

len in den bivariaten Kreuztabellen aller Variablenpaare sowie die Auswahl geeigneter Homogenitätsmaße.²⁷

Zur Festlegung der Fehlerzellen führt Molenaar das Konzept der "Cumulated Percentages Past" (CPP) ein, auf deren Basis in Übereinstimmung mit dem Guttman-Modell für mehrstufige Variablen die Sequenz der Wertepaare absteigender Auftretenswahrscheinlichkeit ermittelt wird. Unter den CPP sind dabei die kumulierten relativen Häufigkeiten der Randverteilungen der Kreuztabellen je zweier Variablen zu verstehen, wobei die Kumulation von der Endstufe mit der geringsten Auftretenswahrscheinlichkeit ausgehend erfolgt. Bei der oben beschriebenen Mokken-Analyse für dichotome Variablen wurde die Fehlerzelle immer a priori unabhängig von den Randverteilungen festgelegt und negative H_i hatten zur Folge, daß diese Variablen nicht in eine Skala aufgenommen wurden. Konsequenz dieser Vorgehensweise ist, daß lediglich unipolare Skalen generiert werden können. Bipolare Skalen treten daher immer in der Form von zwei unipolaren Skalen zutage - es sei denn, die konstituierenden Variablen eines der beiden Pole werden vorab umcodiert ($1 \rightarrow 0$, $0 \rightarrow 1$). Eine Orientierung an den relativen Häufigkeiten der Endstufen der Einzelvariablen resultiert jedoch in einer automatischen Umcodierung der konstituierenden Variablen eines evtl. negativen Poles - und zwar auch dann, wenn der Algorithmus für polytome Variablen auf dichotome Variablen angewendet wird.

Dieser Effekt ist zwar erwünscht, muß aber andererseits bei der Bestimmung der Skalenwerte einzelner Untersuchungsobjekte adäquat berücksichtigt werden. Die generelle Vorgehensweise der Ermittlung der Fehlerzellen im polytomen Fall läßt sich nur schwer allgemein darstellen, so daß sie anhand eines Beispielles illustriert werden soll. Abbildung 4 gibt die Kreuztabellen der relativen Häufigkeiten zweier Variablen i und j wieder. Betrachten wir zunächst die mit "Gesamt" bezeichneten Spalten- und Zeilensummen, so ist offensichtlich, daß für beide Variablen der Code 2 eine geringere Auftretenswahrscheinlichkeit hat als der Code 0. Die Ermittlung der CPP muß also durch Kumulation "von 2 nach 0" erfolgen - zeilenweise also von unten nach oben und spaltenweise von rechts nach links.

²⁷ D.K. Hildebrand, J.D. Laing & H. Rosenthal, Prediction Analysis of Cross Classifications, New York, Wiley, 1977.

Abbildung 4

	Variable i			Gesamt	CPP	
	0	1	2			
Variable	0	29	20	1	50	100
j	1	1	8	1	10	50
	2	1	12	27	40	40
Gesamt		31	40	29	100	
CPP		100	69	29		

(Diese Kumulationsrichtungen indizieren, daß zur Erreichung der Stufe 2 der Variablen ein höheres Ausmaß der latenten Eigenschaft erforderlich ist, als zur Erreichung der Stufe 0.) Die erste erlaubte Zelle ist nun die Zelle mit der CPP-Kombination (100; 100). (Die theoretische Wahrscheinlichkeit dafür, daß alle Untersuchungsobjekte diese oder eine höhere Stufe der aus beiden Variablen gebildeten Skala erreicht haben, ist 1.0). Nun muß entschieden werden, welches die nächste erlaubte Zelle ist. Zur Auswahl stehen (0; 1) und (1;0). Die zu beantwortende Frage ist im Grunde: Ist der Schritt von 0 nach 1 kleiner für Variable i oder für Variable j ? Ein Blick auf die CPP zeigt, daß bei Variable i 69% der Untersuchungsobjekte die Stufe 1 oder eine höhere erreicht haben, während bei Variable j dies nur für 50% zutrifft. Der Schritt Von 0 nach 1 ist daher wohl bei Variable i kleiner als bei Variable j. Die nächste erlaubte Zelle ist damit die Zelle (0;1). Von dieser kann nun entweder der Schritt nach 2 in Variable i der nächstkürzere sein, oder der nach 1 in Variable j. Als nächste erlaubte Zelle stehen also zur Auswahl (0;2) und (1;1). Betrachten wir auch hier wieder die korrespondierenden CPP-Werte, so sind dies 29% für (0;2) und 50% für (1;1). Daraus wird geschlossen, daß der Schritt nach j = 1 der kürzere ist, so daß die Wahl auf (1;1) fällt. Nun stehen zur Wahl die Zellen (1;2) und (2;1) mit den korrespondierenden CPP-Werten 29% und 40%. Da $40 > 29$, fällt die Wahl auf (2;1), von wo aus nur noch (2;2) zur Auswahl steht.

Als erlaubte Zellen sind damit festgelegt: (0;0), (0;1), (1;1), (2;1) und (2;2). Die übrigen Zellen - also (0;2), (1;0), (1;2), (2;0) - liegen damit als Fehlerzellen fest und der Homogenitätskoeffizient H_{ij} kann durch Summation über diese Zellen hinweg (in Zähler und Nenner) analog zu Gleichung 3 bestimmt werden. In gleicher Analogie (zusätzliche Summation über alle Fehlerzellen hinweg) können die Koeffizienten H_i und $H_{..}$ bestimmt werden.

Der im Vorangegangenen vorgestellten Vorgehensweise zur Ermittlung der Fehlerzellen liegt die Annahme zugrunde, daß die beiden Variablen sich

entlang einer Dimension skalieren lassen. Ist diese Annahme nicht haltbar, so ist die Besetzung der Fehlerzellen verhältnismäßig hoch und der H-Koeffizient wird schlecht. Die Annahme der unidimensionalen Skalierbarkeit findet ihren Niederschlag in der Vorschrift, sich von einer Ecke der Kreuztabelle nur durch Einzelschritte in der Horizontalen oder in der Vertikalen in Richtung auf die diagonal gegenüberliegende Ecke bis hin zu dieser bewegen zu können.

Da es Molenaar desweiteren gelang, auch die Teststatistik DELTA* entsprechend zu verallgemeinern, kann der eigentliche Skalierungsalgorithmus unverändert vom dichotomen Fall übernommen werden.

3. Die Anwendung der Mokken-Skalierung auf die namentlichen Abstimmungen in der Frankfurter Nationalversammlung (FNV) und der Assemblée nationale constituante (ANC): Ergebnisse und Interpretation

Das Verfahren der Mokken-Skalierung wurde exemplarisch auf zwei Datenbestände angewendet, in denen das Verhalten von jeweils 900 (ANC) und 766 (FNV) Abgeordneten in 90 (FNV) und 86 (ANC) namentlichen Abstimmungen erfaßt wurde. Diese Voten sind systematische Zufallsauswahlen aus den 259 (ANC) und 298 (FNV) namentlichen und in den offiziellen Parlamentsberichten dokumentierten Abstimmungen. Erfaßt wurde - nach einem Zufallsstart - jeweils jede dritte Abstimmung in zeitlicher Folge. Anhand eines Datensatzes, der nahezu alle Abstimmungen in der FNV enthält, konnte nachträglich die Stichprobenvalidität überprüft werden. Es zeigte sich, daß das ungewöhnliche Verfahren einer Zufallsauswahl von Items keine erheblichen Verzerrungen bewirkte: Die Hauptskalen, die aus verschiedenen Stichproben und dem Datensatz mit (nahezu) allen Abstimmungen gewonnen wurden, korrelierten mit Werten von $r = 0.95$ und 0.99 .²⁸

Vor der Anwendung der Mokken-Skalierung wurden die Primärdaten in zweierlei Weise aufbereitet:

(1) Zunächst wurden die Abstimmungen nach dem Kriterium der Zugehörigkeit (bzw. Affinität) der Antragsteller zu Fraktionen und Protofraktionen "gepolt". Entsprechende Informationen waren in den Parlamentsberichten und in den biographischen Materialien über die einzelnen Abgeordneten enthalten. Der Code +1 wurde der Ja-Stimme dann zugewiesen, wenn der Antrag von einem Abgeordneten der Rechten eingebracht wurde, der Code -1,

²⁸ Best, Bildung und Besitz, S. 498.

wenn ein Abgeordneter der Linken Antragsteller war. Bei Stimmenthaltung wurde Code 0 vergeben. In den folgenden Analysen werden die Daten in der Weise dichotomisiert, daß die Enthaltungen der Kategorie "Nichtzustimmung" zugewiesen wurden.²⁹

(2) Fehlende Werte, die auf Abwesenheit von Abgeordneten bei Abstimmungen zurückzuführen waren, wurden durch "Imputation" mittels multipler schrittweiser Regression geschätzt. Die Anwendung eines fall- und variablenorientierten Schätzverfahrens war geboten, da andere Methoden der Behandlung fehlender Werte, wie etwa deren pauschale Ersetzung durch einen "neutralen" Wert, zu schwerwiegenden Verzerrungen führen würden. So würden etwa Abgeordnete, die tatsächlich dem rechten oder linken Flügel angehört hatten, nach ihrem Abstimmungsverhalten der "Mitte" zugeordnet, wenn häufig bei Abwesenheit ein neutraler Wert vergeben worden wäre. Auch ergab bei der Schätzung fehlender Werte mit Regressionstechniken die Validitätsüberprüfung im Vergleich zu anderen missing-data-Techniken eine deutlich bessere Anpassung der Skalenwerte an die Kriteriumsvariable "Fraktionszugehörigkeit".³⁰

Die Analysen des Abstimmungsverhaltens mit dem Verfahren der Moken-Skalierung hatten ein prägnantes Ergebnis, das in den Grundzügen zwischen beiden Untersuchungsgruppen übereinstimmt. Wählt man den hohen Wert von $H > 0.50$ als Einschlußkriterium für die Aufnahme einer Abstimmung in eine Skala, dann lassen sich in der Frankfurter Nationalversammlung 79 von 90 (=88%) der Voten und in der Assemblée nationale constituante 66 von 86 Abstimmungen (=77%) jeweils einer Hauptskala zuordnen. Die für die Hauptskalen errechneten Gesamt-Homogenitätskoeffizienten erreichen sehr hohe Werte (ANC und FNV: $H = 0.76$), gleiches gilt für die Prüfstatistik DELTA* (ANC = 752.56; FNV = 946.91).

Es wurde jedoch schon darauf hingewiesen, daß die Homogenität eines Item-Satzes nicht hinreicht, um zu entscheiden, ob eine Skala vorliegt, die eine Eigenschaftsdimension abbildet. Dazu muß zusätzlich geklärt werden, ob die in die Hauptskala aufgenommenen Abstimmungen einem gemeinsamen "Bedeutungsuniversum" zugehören, das heißt, einem "Sinnzusammenhang, ... für den ein konkreter Satz von Indikatoren als operationale Definition verstanden werden kann".³¹

²⁹ Stimmenthaltungen waren in beiden Parlamenten außerordentlich selten und machten weniger als 0.3% der abgegebenen Voten aus. Ebd., S. 499.

³⁰ Heinrich Best, "Computing the Unmeasurable: Estimating Missing Values in Legislative Roll-Call Analysis", in: Peter Denley et al. (Hrsg.), *History and Computing*, Bd. 2 (Manchester u. New York 1989), S. 104-116.

In Abstimmungsanalysen ist es sonst üblich, diesen Sinnzusammenhang aus den manifesten Inhalten der Vorlagen zu erschließen. Das ist ein problematisches Vorgehen, weil sich dabei gewöhnlich ein hoher Arbeitsaufwand mit einem unerfreulich großen Ermessensspielraum bei der Deutung von Abstimmungstexten verbindet. Im gegebenen Fall verfügen wir jedoch über leistungsfähige externe Kriteriumsvariablen, die eine zuverlässigere Zuordnung ermöglichen: In beiden Versammlungen bildeten sich mehr oder weniger ephemere Fraktionen oder "Protofraktionen", die von den Beteiligten und den zeitgenössischen Beobachtern nach einem rechts-links-Schema geordnet wurden. Wenn die Skalen, wie erwartet, Positionen auf einem rechts-links-Kontinuum messen, dann müßten die Abgeordneten vom äußersten rechten Flügel über die Mitte zum linken Flügel hin abnehmende durchschnittliche Skalenwerte aufweisen. Die Skalenwerte der Abgeordneten wurden zuvor aus der Anzahl der positiven Voten gebildet, die für Vorlagen der Rechten abgegeben wurden. Wie erwähnt, ist dies ein konservatives Verfahren, weil es bei inkonsistentem Abstimmungsverhalten eher unterschätzte Werte liefert.³²

Das Ergebnis einer Varianzanalyse der Skalenwerte mit der Fraktionszugehörigkeit als Klassifikationsvariable unterstützt für beide Untersuchungsgruppen nachdrücklich die Annahme, daß die Skalen Positionen auf einem rechts-links-Kontinuum abbilden. Die durchschnittliche Anzahl positiver Reaktionen nimmt zu, je weiter rechts eine Fraktion oder eine Protofraktion plaziert war. Die historisch überlieferten und die aus dem Abstimmungsverhalten ermittelten Rangfolgen von Fraktionen und Protofraktionen sind nahezu identisch. Der in beiden Fällen sehr hohe ETA-Wert ($ANC = 0.80$; $FNV = 0.82$), der den Anteil der Summe der Abweichungsquadrate zwischen den Gruppen an den quadrierten Gesamtabweichungen angibt, zeigt an, daß die Gruppen relativ homogen sind. Wenn man - mit guten Gründen - jene Abgeordneten ausschließt, die keiner Fraktion angehörten oder für die die genauen Angaben zu den politischen Standorten im biographischen Material fehlen, verbessert sich dieser Wert erwartungsgemäß - im Fall der deutschen Untersuchungsgruppen sogar erheblich ($ANC = 0.84$; $FNV = 0.94$). In die gleiche Richtung weist die Inspektion der Standardabweichungen und der Quartilsabstände der Gruppenmittelwerte. Dabei fällt auf, daß die Streuung von den Flügelfraktionen zu den Mittelfraktionen hin zunimmt. Dies ist ein inhaltlich bedeutsamer Befund, der mit anderen Beobachtungen koinzidiert, nach denen die Fraktionsdisziplin der Abgeordneten bei den Mittelfraktionen geringer war.³³

³¹ Scheuch u. Zehnpfennig, Skalierungsverfahren, S. 117.

³² S.o., S. 60.

Alles in allem bestätigt aber die Korrelation der Skalen mit den Kriteriumsvariablen die Erwartung, daß ein links-rechts-Kontinuum in beiden Untersuchungsgruppen den gemeinsamen "Sinnzusammenhang" der Abstimmungen der Hauptskalen bildete. Die verbleibende Unschärfe ist - wie auch der Vergleich zwischen den Untersuchungsgruppen nahelegt - eher den Validitätsmängeln der Kriteriumsvariablen als denen der Skalen zuzuschreiben. Auch muß bedacht werden, daß die bei der Skalenbildung berücksichtigten Abstimmungen in einem langen Zeitraum stattfanden, während sich die Angaben über Gruppen- und Fraktionszugehörigkeiten auf bestimmte Zeitpunkte beziehen.

Eine Komplikation bei diesem Verfahren, das ja auf Mittelwertberechnungen beruht, besteht allerdings darin, daß Mokken-Skalen ebenso wie Guttman-Skalen lediglich die Meßeigenschaften von Ordinalskalen haben und demnach nur über Rangordnungen der Untersuchungseinheiten Auskunft geben. Wir haben uns jedoch aus drei Gründen berechtigt gefühlt, diese Beschränkung hier zu ignorieren. Zum einen werden unsere Skalen aus einer ungewöhnlich großen Anzahl von Items gebildet; die "ungemessenen" Abstände zwischen den Skalenwerten können also nicht allzu groß sein. Für diese Vermutung spricht auch, daß die Distanzen zwischen den Items, gemessen an ihrer "Schwierigkeit", nicht sehr groß sind. Zudem dient die Mittelwertberechnung hier ja lediglich der inhaltlichen Deutung und Validierung der Skalen, deren weitere Verwendung als metrische Indikatoren ist nicht beabsichtigt.

Aufschlüsse über die Mechanismen, die den Prozeß der Formierung von Konfliktgruppen steuerten, ergeben sich, wenn man die Besetzung der Skalenpositionen betrachtet (Abb. 5 u. 6). Dann zeigt sich ein markant bipolares Muster: In beiden Versammlungen konzentrierten sich die Fälle auf die äußeren Skalenpositionen. Dieses Ergebnis der Skalenanalysen unterstützt unsere Annahme, daß das Abstimmungsverhalten einer Entscheidungslogik folgte, bei der der status quo über Serien von Voten die gleichbleibende Alternative bildete. "Proximity-Voting" und strategisches Abstimmungsverhalten hätten dagegen zu überzufällig häufigen Inkonsistenzen in den Abstimmungsmustern oder einer stärkeren Besetzung des Mittelbereichs der Hauptskalen führen müssen. Die beobachteten Verteilungen sind dagegen ein starkes Argument für die These eines "natürlichen Bipartismus", wie sie u.a. Maurice Duverger vertreten hat: Das Ergebnis politischer Freiheit war nicht promiskuitive Pluralität, sondern ein rigoreses, wenn auch kompetitives, Kon-

³³ Best, Bildung und Besitz, S. 330 ff.

Abbildung 5

Abgeordnete der Frankfurter Nationalversammlung 1848/49:
 Skalenwerte auf der Links-Rechts-Skala (Mokken-Skala)
 79 Items (in der Darstellung wurden je zwei benachbarte Skalenpositionen zusammengefaßt)
 $\text{II} = 0.76$ $\text{Delta-Star} = 946.91$

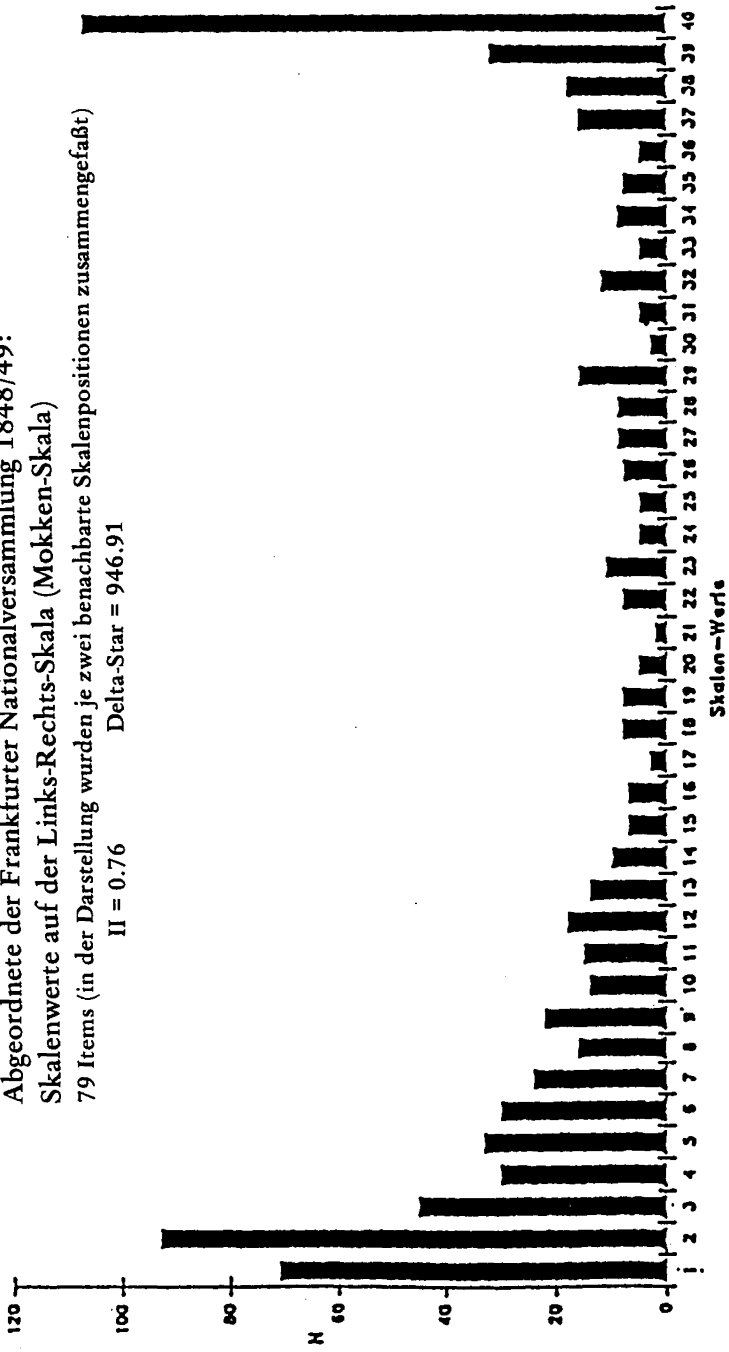


Abbildung 6

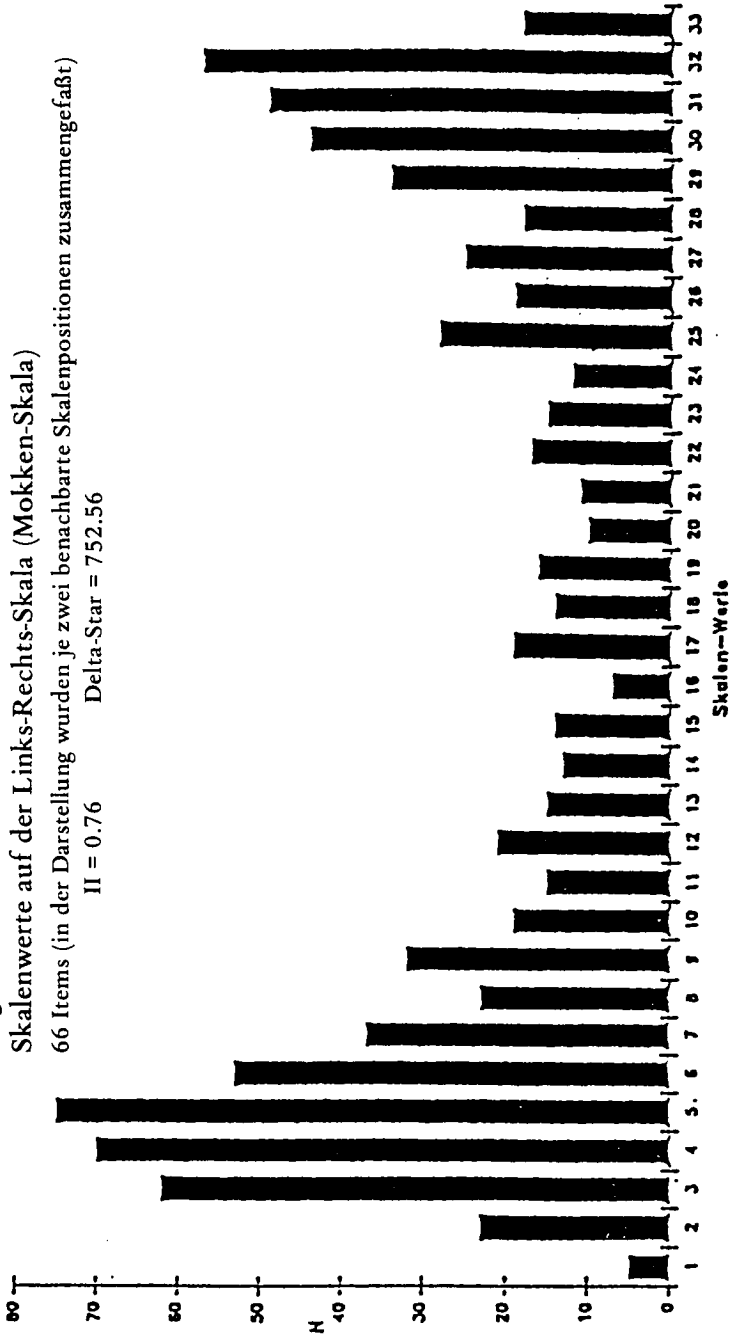
Abgeordnete der Assemblée Nationale Constituante 1848/49:

Skalenwerte auf der Links-Rechts-Skala (Mokken-Skala)

66 Items (in der Darstellung wurden je zwei benachbarte Skalenpositionen zusammengefaßt)

$\Pi = 0.76$

Delta-Star = 752.56



fliktmuster.³⁴

Die Frage stellt sich hier, warum in beiden Versammlungen, die in markant unterschiedlichen historisch-politischen Kontexten zusammentraten, isomorphe und rigide Konfliktstrukturen entstanden. Einen "Zwang zur Konsistenz", der in Befragungssituationen fehlt, setzte - wie erwähnt - sicherlich der Umstand, daß namentliche Abstimmungen öffentlich sind und Abweichungen der Abgeordneten von den Verhaltenserwartungen durch andere Akteure sanktioniert werden können. Doch erklärt dieser Sachverhalt nicht, warum sich in beiden Versammlungen bipolare Konfliktstrukturen ausprägten, was im übrigen ja auch den frühparlamentarischen Vorbehalten gegenüber "Parteiungen" jeder Art widersprach und die beobachteten Konstellationen zu einem Beispiel für ungeplante, ja unerwünschte Folge sozialen Handelns macht.

In einem anderen Zusammenhang wurde ein dynamisches Modell der Strukturbildung in parlamentarischen Versammlungen vorgeschlagen, das von der Überlegung ausgeht, daß die Formierung politischer Konfliktgruppen in Parlamenten den Koordinationserfordernissen kontinuierlicher Entscheidungsbildung unter kompetitiven Bedingungen folgt.³⁵ In vereinfachter Weise läßt es sich wie folgt darstellen: Ein primärer Prozeß der Gruppenbildung wird dadurch ausgelöst, daß Abstimmungen in Parlamenten - wie andere rationale Entscheidungen auch - Informations- und Koordinationskosten verursachen, die sehr hoch werden können, wenn sich Abgeordnete mit Serien von Vorlagen befassen müssen, zu deren Bewertung großes Fachwissen erforderlich ist oder die nur schwer absehbare politische Folgen haben. In solchen Situationen werden Abgeordnete versuchen, ihre Informationskosten zu senken, indem sie ihr Abstimmungsverhalten an solchen Kollegen orientieren, die sie für sachlich kompetent oder besonders vertrauenswürdig ansehen. Im Insider-Jargon der amerikanischen Parlamentarier wird dieser Vorgang mit "cue-taking" bezeichnet. Die Entstehung von Protofraktionen in Parlamenten läßt sich danach als das Ergebnis einer Spezialisierung und Arbeitsteilung deuten. (Proto-)Fraktionen bilden die Foren der Kommunikation, in denen Abgeordnete entweder ihr jeweiliges Expertenwissen austauschen oder Informationen über den Preis der Verpflichtung zu Konformität in Abstimmungen zu erlangen suchen.

³⁴ Les partis politiques. (Paris 1976 u.ö.), S. 293 ff. Auch Carl Schmitts Bestimmung des Politischen als ein polarisiertes Freund-Feind Verhältnis läßt sich hier als Referenz anführen.

³⁵ Ausführlich mit Literaturhinweisen: Best, Bildung und Besitz, S. 350-357.

Gelingt es, verschiedene Handlungsalternativen in einem niederdimensionalen Konflikttraum zu lokalisieren, führt dies zu einer weiteren kognitiven Entlastung der parlamentarischen Akteure. Robert Heiner hat darauf aufmerksam gemacht, daß unter turbulenten Handlungsbedingungen, wie sie ja in Revolutionen typischerweise bestehen, die Tendenzen zur Realitätsvereinfachung besonders ausgeprägt sind.³⁶

Solche Argumente begründen jedoch eher die Eindimensionalität der Konfliktstrukturen, weniger ihre Bipolarität, die wir vor allem dem Bemühen der Parlamentarier zurechnen, ihre Durchsetzungsmacht zu steigern. Die primären parlamentarischen Gruppen verstetigten sich und tendierten zur weiteren Aggregation infolge der parlamentarischen Mehrheitsregeln: stabile Mehrheiten über Zeit werden belohnt - das gilt selbst für die Verabschiedung von Einzelgesetzen mit zugehörigen Amendments.

Bedenkt man, daß politische Konflikte auch Kämpfe um Machtpositionen sind, wird eine weitere Knappheitsregel ins Spiel gebracht, die im Argument der "minimal winning coalition" formalisiert wurde:³⁷ Kleine Mehrheiten führen zu einer günstigeren Relation zwischen der Zahl der im Konkurrenzkampf um die Macht verfügbaren Positionen und der Zahl der durch Zugehörigkeit zur Mehrheitsgruppe zur Ämterappropriation berechtigten Aspiranten.

Schließlich gilt: eine bipolare Struktur erhöht die Mitgliederstabilität der parlamentarischen Konfliktgruppen, denn bei sonst gleichen Bedingungen ist in einem Zweiparteien-System für den Austritt oder das Ausscheiden im Durchschnitt ein höherer Preis zu zahlen als in einer Mehrparteien-Struktur.³⁸ Fassen wir die Argumente zusammen, dann erklärt sich die beobachtete eindimensional-bipolare Konfliktstruktur aus dem Zusammenwirken von drei Hauptfaktoren der Strukturbildung: (1) der Notwendigkeit, die Transaktionskosten bei der parlamentarischen Entscheidungsbildung zu senken, (2) der Tendenz, die Durchsetzungsmacht der parlamentarischen Konfliktgruppen zu steigern und dabei (3) einen möglichst starken Solidaritätsdruck auf die Gruppenmitglieder auszuüben.

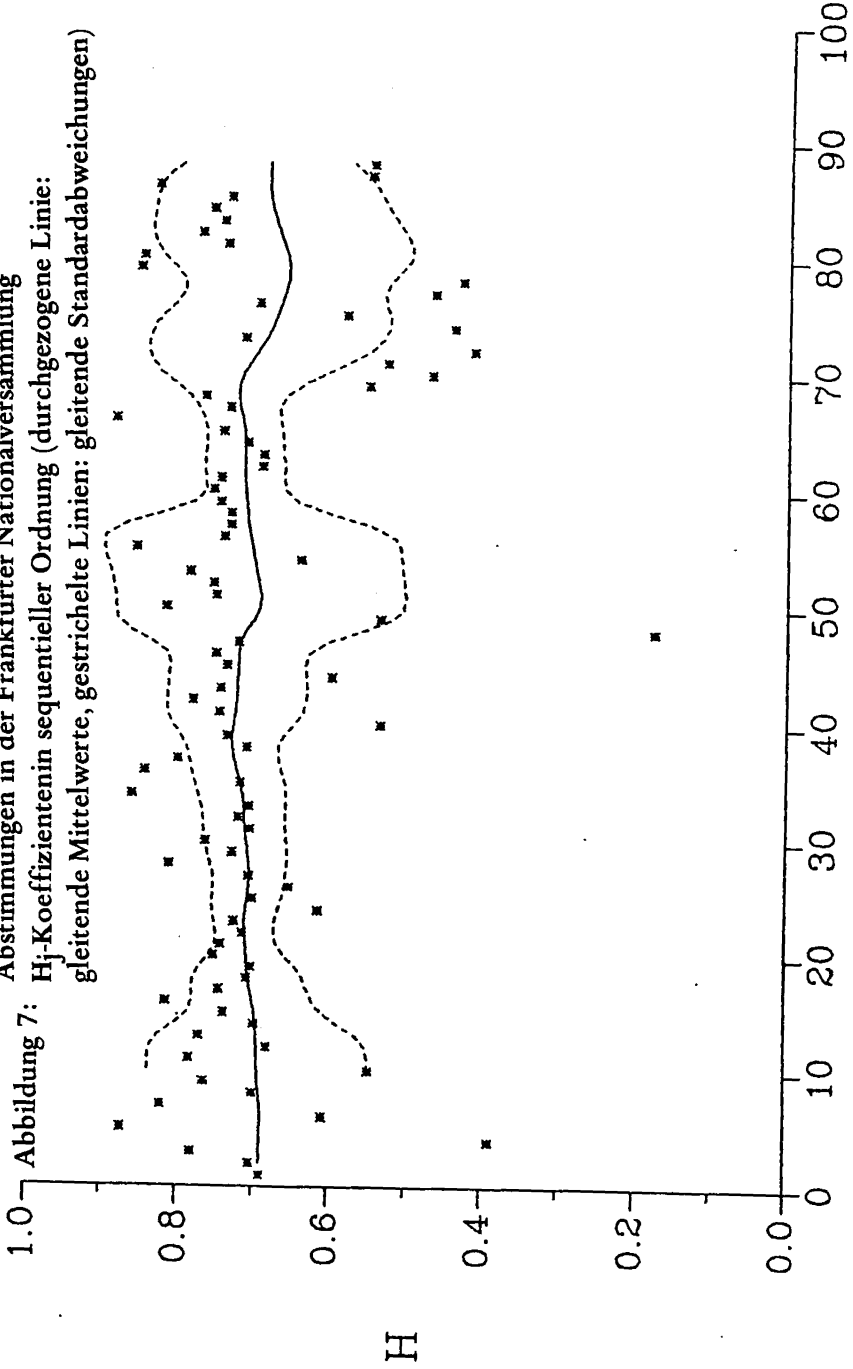
Die Differenzierung und Verstetigung politischer Konfliktgruppen wird

³⁶ Ronald A. Heiner, "The Origin of Predictable Behavior", in: *American Economic Review*, 73 (1983), S. 560-589.

³⁷ L.C. Dodd, "Party Coalitions in Multiparty Parliaments. A Gametheoretic Analysis", in: *American Political Science Review*, 68 (1974), S. 1093-1117.

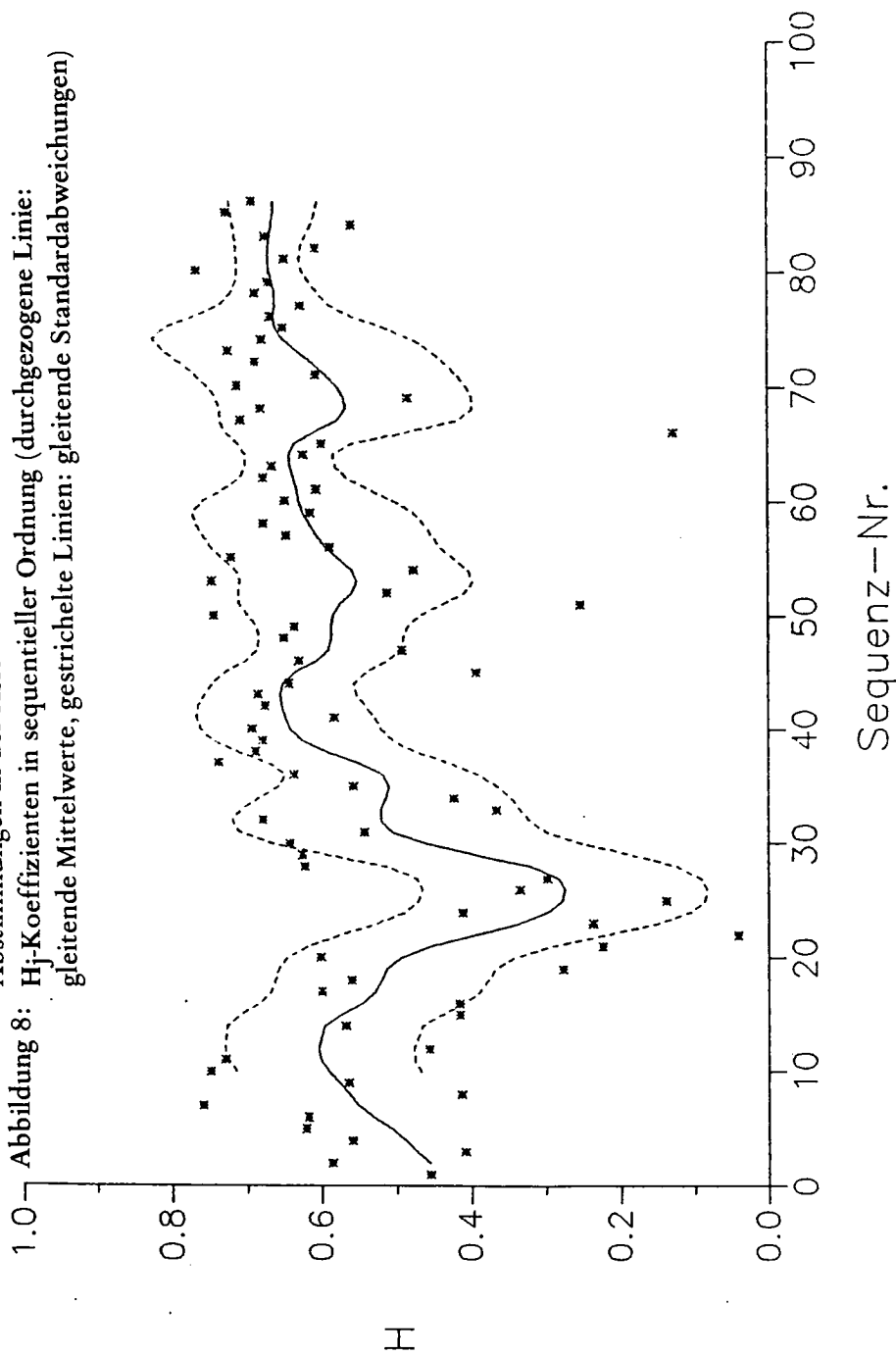
³⁸ Raymond Boudon, *Die Logik des gesellschaftlichen Handelns* (Neuwied u. Darmstadt 1980), S. 38; Michael Hechter, "A Theory of Group Solidarity", in: ders. (Hrsg.), *The Microfoundations of Macrosociology* (Philadelphia 1983), S. 16-57.

Abstimmungen in der Frankfurter Nationalversammlung
Abbildung 7: H_j-Koeffizienten in sequentieller Ordnung (durchgezogene Linie:
gleitende Mittelwerte, gestrichelte Linien: gleitende Standardabweichungen)



Sequenz-Nr.

Abstimmungen in der Assemblée Nationale Constituante:
 Abbildung 8: Hj-Koeffizienten in sequentieller Ordnung (durchgezogene Linie:
 gleitende Mittelwerte, gestrichelte Linien: gleitende Standardabweichungen)



hier als ein Lern- und Institutionalierungsprozeß, als das ungeplante Ergebnis des parlamentarischen Entscheidungshandelns der einzelnen Abgeordneten konzipiert. Tatsächlich widersprach die Formierung kohärenter und relativ disziplinierter Abstimmungsblöcke dem Prinzip der ungebundenen Stimmabgabe, das für die liberale Lehre von der freien Repräsentation konstitutiv ist.³⁹ Andererseits gehen wir davon aus, daß die parlamentarischen Konfliktstrukturen nicht einfach als ungestörte Abbilder gesellschaftlicher Spannungslinien aufgefaßt werden dürfen. Die verfassungsgebenden Versammlungen wurden vor der Entstehung organisierter Massenparteien und der Formierung sozioökonomischer Interessengruppen auf gesamtstaatlicher Ebene gewählt. Tatsächlich hatte der Prozeß der Partei- und Verbandsbildung einen seiner Ausgangspunkte in den verfassungsgebenden Versammlungen selbst. Wenn es also zutrifft, daß die Entstehung der parlamentarischen Konfliktgruppen das Ergebnis von autokatalytischen Lern- und Institutionalierungsprozessen war, die sich in den Parlamenten vollzogen, müßte sich dies in unseren Daten an der Variation der sequentiell geordneten H_i -Koeffizienten nachweisen lassen. Sie wurden als Homogenitätsmaße eingeführt, in denen die Übereinstimmung des Antwortmusters einer einzelnen Vorlage mit einem generellen Abstimmungsmuster zum Ausdruck kommt. Im Grenzfall einer eindimensional-homogenen Konfliktstruktur müßten alle Abstimmungen in einem Parlament eine Skala bilden. Strukturbildung als ein Prozeß, der von einer Situation relativer Strukturlosigkeit ausgeht und zu einem stabilen Konfliktmuster führt, müßte

- (a) in einem Anstieg der H_i -Koeffizienten bis zu einem Sättigungswert nahe 1 und nachfolgender Stabilität der Homogenitätskoeffizienten und/oder
- (b) in einer abnehmenden mittleren Varianz der Homogenitätskoeffizienten bei fortschreitender Sequenzposition resultieren.

Die Abbildungen 7 und 8 zeigen die Positionen der H_i -Koeffizienten als Sterne im Koordinatensystem. Die durchgezogenen Linien sind die Verläufe der exponentiell gewichteten Gleitmittelwerte der H_i -Koeffizienten (geglättet nach dem Modell 4235H).⁴⁰ Die gestrichelten Linien verbinden die Punkte

$$H_i \text{ gegl.} \pm SD(10) \text{ gegl.}$$

wobei: SD (10) Standardabweichungen der jeweils letzten 10 Punkte sind (gleitende Standardabweichung, Glättung auch hier nach 4235H). Als chronologische Referenz wählen wir die äquidistant geordneten Sequenzpositionen

³⁹ Eberhard Schütt-Wetschky, Grundtypen parlamentarischer Demokratie. Klassisch-altliberaler Typ und Gruppentyp (Freiburg und München 1984).

⁴⁰ Paul Velleman und D. Hoaglin, ABCs of EDA, Duxbury Press, 1981. John W. Tukey, Exploratory Data Analysis, Addison-Wesley, 1977.

nen der Abstimmungen und nicht die realen Zeitabstände, da wir - in Übereinstimmung mit Erkenntnissen amerikanischer Untersuchungen - davon ausgehen, daß Abgeordnete ihr Abstimmungsverhalten an ihren vorausgegangenen Voten, ihrer "voting history" orientieren, weitgehend unabhängig von der Dauer der Zeitintervalle zwischen den Abstimmungen.⁴¹

Die tatsächlichen Eigenschaften der Wertereihen weichen erheblich von den erwarteten Verläufen ab. Im Fall der deutschen Abgeordneten sind die gleitenden Mittelwerte der Homogenitätskoeffizienten annähernd stationär, in Frankreich beobachten wir dagegen zunächst den erwarteten Anstieg, der jedoch etwa mit der zwanzigsten Abstimmung abbricht und sich nach einem tiefen Einbruch erst ab etwa der vierzigsten Abstimmung - bei fortdauernden Schwankungen - im Wertebereich zwischen $H = 0.60$ und $H = 0.80$ einpendelt. Diese Unterschiede zeigen an, daß allgemeine Lerntheorien und der Verweis auf die Eigendynamik von Institutionalisierungsprozessen den tatsächlichen Verlauf der Strukturbildung in beiden Versammlungen nicht hinreichend aufzuklären vermögen. Sie unterschätzen die Bedeutung von Veränderungen des spezifischen Problemhaushalts in beiden Versammlungen im Prozeß der Verfassungsgebung, von Interventionen exogener Entwicklungen und Ereignisse, schließlich den Einfluß der Organisationsform parlamentarischer Konfliktgruppen.

So konnte in der Frankfurter Nationalversammlung bereits Anfang Juni 1848, also rund 14 Tage nach ihrem Zusammentritt von "fertigen, konstituierten und organisierten Parteien" die Rede sein, wobei "Partei" in diesem zeitgenössischen Zitat für den heutigen Begriff der Fraktion steht.⁴² Anfang Oktober 1848 (in der Abbildung 9 etwa Sequenzposition 13) war der Prozeß ihrer programmatischen Abgrenzung und organisatorischen Formierung abgeschlossen. In dieser Phase verteilten sich die Abgeordneten auf bis zu 9 Fraktionen, die ihr Verhalten in Abstimmungen jedoch eng, zum Teil durch fraktionsübergreifende "Kommissionen", koordinierten.⁴³ Ca. 20% der Abgeordneten waren zu diesem Zeitpunkt fraktionslos.⁴⁴ Die Mehrheitsposition hatten die Fraktionen des "Rechten Centrums", dem sich zumeist die Abgeordneten der äußeren Rechten anschlossen. Im Ergebnis bewirkte der Prozeß der Fraktions- und Koalitionsbildung in der Frankfurter Nationalversammlung die Institutionalisierung des rechts-links-Gegensatzes als dominante Konfliktfront. Das rechts-links-Schema bildete das "kognitive Konzept", mit

⁴¹ Aaron Wildavsky, *The Politics of the Budgetary Process* (Boston 1964).

⁴² Karl Biedermann, *Erinnerungen an die Paulskirche* (Leipzig 1849), S. 11.

⁴³ Georg Beseler, *Erlebtes und Erstrebtes 1809- 1859* (Berlin 1884), S. 76.

⁴⁴ Best, *Bildung und Besitz*, S. 326.

dem sich problemlos Abstimmungsvorlagen und Abgeordnete bewerten ließen. Auch die fraktionslosen Mitglieder der Nationalversammlung orientierten ihr Verhalten an diesem Raster und stimmten mehr oder weniger konsistent mit einem der Abstimmungsblöcke.

In unseren Daten kommt der hier skizzierte Prozeß der Institutionalisierung einer Konfliktstruktur in der abnehmenden Streuung der H_1 -Koeffizienten während der Konstituierungsphase der Nationalversammlung zum Ausdruck. Ab September 1848 ordneten sie sich in einem schmalen Wertebereich. Doch hatte diese Konsolidierung keinen Bestand, denn zunehmend geriet die Auseinandersetzung um rechte und linke Positionen, die sich vereinfachend als eine Kontroverse um die soziale Reichweite politischer und ökonomischer Partizipationsrechte deuten läßt, unter den Einfluß einer zweiten Konfliktfront: den großdeutsch-kleindeutschen Gegensatz, in dem es um die territoriale Reichweite einer deutschen Nationalstaatsbildung, die Frage nach der Einbeziehung des österreichischen Kaiserstaats ging.⁴⁵ Abstimmungen, die diesem Konfliktfeld zuzuordnen sind, bilden eine zweite Mokken-Skala, der bei einem Grenzwert von $H > 0.5033$ Abstimmungen zugeordnet werden. 28 dieser Abstimmungen sind auch Bestandteil der Hauptskala.

Hier wird deutlich, daß ein Item mehreren eindimensionalen Skalen zugeordnet werden kann. Dementsprechend wird Mokkens Verfahren auch als multiple eindimensionale Skalierung bezeichnet.⁴⁶ Obwohl es demnach zulässig ist, daß sich Skalen, die unterschiedliche Bewertungsdimensionen erfassen, weit überlappen, manifestiert sich in unseren Daten die Pluralisierung der Konfliktstruktur in einer zunehmenden Inhomogenität des Abstimmungsverhaltens: Die steigende Varianz der H_1 -Koeffizienten zwischen der 40. und 55., dann wieder nach der 70. Abstimmung ist dem Umstand zuzuschreiben, daß in diesen Phasen der großdeutsch-kleindeutsche Konflikt auf die Agenda der Nationalversammlung gesetzt wurde. Die Entscheidungen über das Reichsoberhaupt und die Ausdehnung des Reichsgebietes ließen die Frage des Verhältnisses zum österreichischen Kaiserstaat mit seinen umfangreichen nichtdeutschen Gebieten unabweisbar werden. In den Abstimmungen über dieses Thema durchmischten sich dynastische und Staatsloyalitäten mit verfassungspolitischen Grundsatzertwägungen, was erklärt, daß viele der betreffenden Abstimmungen sowohl der links-rechts wie der großdeutsch-klein-

⁴⁵ Ebd., S. 344-349.

⁴⁶ Kees Niemöller u. Wijbrandt van Schuur, "Stochastic Models for Unidimensional Scaling: Mokken and Rasch", in: David McKay et al., *Data Analysis and the Social Sciences* (London 1983), S. 145.

deutschen Skala zugewiesen werden. Erst in den Schlußabstimmungen über die Reichsverfassung gab es dann eine Serie von Voten, die exklusiv der großdeutsch-kleindeutschen Dimension zugeordnet werden, in denen diese zweite Konfliktfront also "rein" zum Ausdruck kommt. Auch in dieser Phase gab es jedoch Abstimmungen, in denen weiter nach dem Bewertungsschema des links-rechts-Gegensatzes entschieden wurde. Wir beobachten also die komplizierte Situation einer Konfliktstruktur, die kurzfristig zwischen zwei Zuständen wechselte, auf die gleichzeitig zwei unterschiedliche Mechanismen der Strukturbildung einwirkten. Die Kreuzung politischer Konfliktlinien überlastete die institutionellen Strukturen und informellen Prozeduren der Entscheidungsbildung, die sich bis Sommer und Herbst 1848 herausgebildet hatten. In diesem Spannungsfeld zermürbten und zerbrachen schließlich die Gesinnungs- und Handlungsgemeinschaften, die die Grundlage der Fraktionsbildung und der Bindung der Abgeordneten an ihre Mandate gewesen war. Die Spaltung der Nationalversammlung in sezessionierende Teilversammlungen und der vorzeitige Auszug vor allem österreichischer Abgeordneter aus dem Parlament sind Symptome dieses Zerfalls.⁴⁷

Während in Deutschland Fluktuationen der Homogenität des Abstimmungsverhaltens in den Streuungsmaßen zum Ausdruck kommen, wobei die gemittelten H-Koeffizienten annähernd stationär sind, beobachten wir in der Pariser Konstituante ausgeprägte Schwankungen dieser Wertereihe. Der Befund läßt sich plausibel mit der in Frankreich ephemeren Struktur politischer Konfliktgruppen erklären. Zwar waren auch die "Reunionen" in der Pariser Konstituante nach einem rechts-links-Schema differenziert, doch entwickelten sie im gesamten Untersuchungszeitraum keine stabilen Organisationsstrukturen, fluktuierte ihr Anhang, vermochten sie vor allem nicht das Prinzip der gebundenen Stimmabgabe bei ihren Teilnehmern durchzusetzen.⁴⁸ Der links-rechts Gegensatz war hier also nicht in Koalitionen und Kollisionen formal organisierter Konfliktgruppen institutionalisiert, sondern ein eher unscharfes Ordnungsschema, das vor allem durch die Orientierung auf die wechselnden Regierungen der zweiten Republik Kontur erhielt. So fällt der erste "Zyklus" in der Wertereihe der Homogenitätskoeffizienten mit der Etablierung und dem Niedergang der Regierung des Generals Cavaignac zusammen. Während der Phase des Rückgangs der Abstimmungshomogenität formierte sich in den Schlußabstimmungen über die Verfassung zudem noch einmal ein breiter "re-

⁴⁷ Best, *Bildung und Besitz*, S. 264.

⁴⁸ Paul Bastid, *Doctrines et institutions politiques de la Seconde République*, Bd. 1 (Paris 1945), S. 215 ff.

publikanischer" Konsens, der die im Gefolge des Juniaufstandes aufgebrochene Polarisierung zwischen einem linken und einem rechten Lager überbrückte. In der Abstimmung über die Verfassung in ihrer Gesamtheit, die hier mit der Sequenzposition 21 den Tiefpunkt der Wertereihe der Homogenitätskoeffizienten markiert, gab es nur 38 Gegenstimmen von insgesamt 776 Abstimmenden.⁴⁹ Es folgte die Wahl Louis-Napoléon Bonapartes zum Präsidenten und die Installierung des rechtsgerichteten Ministeriums Barrot-Faloux, das durch den "parti de l'ordre", ein aus den Anhängern der verschiedenen Zweige des bourbonischen Königshauses, Bonapartisten und rechten Republikanern zusammengesetztes Bündnis, unterstützt wurde. Während es General Cavaignac nicht gelungen war, einen stabilen parlamentarischen Rückhalt zu gewinnen, formierte sich nun ein gouvernementales Lager, dem eine linke Opposition gegenüberstand. Damit gewann der links-rechts-Gegensatz als handlungssteuernde Konfliktfront an Prägnanz und Stabilität. Der Fraktionszwang, der in der Nationalversammlung Konformität des Abstimmungsverhaltens bewirkte, hatte in Frankreich in den Interventionen der Regierung, vor allem durch Patronage und Klientelismus, ein Äquivalent.⁵⁰ Diese Beobachtung sollte jedoch nicht zu der Schlußfolgerung verleiten, es habe zugleich auch eine Stabilisierung der Mehrheiten gegeben: Das gouvernementale Lager und die linke Opposition blieben an ihren Rändern ephemere und inhomogene. Die immer noch beachtlichen Fluktuationen der Homogenitätskoeffizienten und ihr erheblicher Streubereich auch nach der Inauguration des Präsidenten Bonaparte am 20. Dezember 1848 (zwischen Sequenzposition 29 und 30) zeigen dies an.

Festzuhalten bleibt, daß der zeitliche Ablauf der Strukturbildung in beiden Parlamenten in der Tat als ein Sozialisations- und Institutionalisierungsprozeß beschrieben werden kann, in dessen Verlauf sich diffuse und inhomogene Orientierungen verstetigen und institutionell verfestigen. Offensichtlich begünstigt die Logik der Entscheidungsbildung in Parlamenten die Formierung "einfacher" Strukturen, in denen sich Issues in niederdimensionalen Konfliktträumen anordnen und sich Abgeordnete in polarisierten Konfliktgruppen gegenüber treten. Diese Deutung wird auch durch den Umstand gestützt, daß das in beiden Konstituanten beobachtete Konfliktmuster in sehr ähnlicher Weise in anderen "Parlamenten der ersten Stunde" auftrat: Im Kongress der Vereinigten Staaten, dem französischen Konvent von 1792 oder der mexikani-

⁴⁹ Abstimmungsergebnis in Heinrich Best, Die Abgeordneten der Assemblée Nationale Constituante. Sozialprofil und legislatives Verhalten (Köln 1983), S. 97.

⁵⁰ Roger Price, The French Second Republic. A Social History (London 1972), S. 225-245.

schen Konstituante von 1917.⁵¹ Andererseits hat aber die Rekonstruktion der Dynamik des Prozesses der Strukturbildung gezeigt, daß es eine unzulässige Verkürzung wäre, Parlamente ausschließlich als Interaktionssysteme zu modellieren, deren Strukturbildung "selbstreferentiell" dem Erfordernis folgt, die Komplexität von Problemlagen zu verringern und Konfliktgruppen zu stabilisieren. Gerade der Fall Deutschland liefert mit dem großdeutsch-kleindeutschen Konflikt das Beispiel dafür, daß auch äußere Problemlagen intervenieren und einen Grad an "Sperrigkeit" erreichen können, der sich jedem Versuch zur Realitätsvereinfachung widersetzt. Die Formierung einer zweiten dominanten Konfliktfront um die territoriale Reichweite der Nationalstaatsbildung, die man unter normativen Gesichtspunkten vielleicht als erwünschte Antwortbereitschaft eines Parlaments auf die Pluralität der vielfach parzellierten politischen Landschaft Deutschlands bewerten könnte, erwies sich tatsächlich als ein Desaster für den Parlamentarisierungsprozeß: Im Spannungsfeld sich kreuzender Konfliktlinien zermürbten und zerbrachen schließlich jene parlamentarischen Gesinnungs- und Handlungsgemeinschaften, die vielleicht in der Lage gewesen wären, das politische Programm der Reichsverfassung gegenüber den traditionellen Machträgern zu verwirklichen.

Die Möglichkeit, von den Schwankungen und der Streuung der Koeffizientenreihen auf bestimmte historische Ereignisse und institutionelle Entwicklungen rückzuschließen, bestätigt noch einmal die Angemessenheit der Mokken-Skalierung für die serielle Analyse namentlicher Abstimmungen: Mit ihr lassen sich nicht nur die Ergebnisse, sondern auch die Dynamiken von Strukturbildungsprozessen in Parlamenten und anderen Entscheidungskörperschaften abbilden. Für einen in der historischen Forschung weit verbreiteten Datentyp steht nunmehr ein mächtiges Verfahren bereit, das frei ist von den Nachteilen der sonst in diesem Forschungskontext eingesetzten Methoden.

⁵¹ Peter H. Smith, "The making of the Mexican Constitution", in: William O. Aydelotte (ed.), *The History of Parliamentary Behavior* (Princeton/N.J. 1977), S.186-224; Alison Patrick, *The Men of the First French Republic: Political Alignments in the National Convention of 1792* (Baltimore 1987); John Hoadley, "The Emergence of Political Parties in Congress, 1789-1803", in: *American Political Science Review*, 74 (1980), S.757-779.

Intergenerationale und intragenerationale Mobilität von Parlamentariern der Frankfurter Nationalversammlung. Eine Anwendung der Korrespondenzanalyse.

von Jörg Blasius

1. Einleitung

Bis Mitte der 60er Jahre wurde Mobilitätsforschung nahezu ausschließlich von Sozialwissenschaftlern betrieben, erst seit diesem Zeitpunkt sind auch Historiker dazu übergegangen "... eins der wichtigsten Erprobungsfelder für quantifizierende Methoden in der historischen Forschung ..." (Kaelble 1978)¹ anzuwenden. Das klassische Verfahren zur multivariaten Auswertung von Mobilitätsdaten ist die multidimensionale Skalierung (MDS)² (s. z.B. Blau und Duncan 1967; Best 1989, 1990),³ ein Verfahren, mit dem (Un)ähnlichkeiten im Mobilitätsverhalten zwischen allen Objekten (allen Berufsgruppen) durch Distanzen in einem (meistens) zweidimensionalen Raum dargestellt werden können. In diesem Beitrag wird die Korrespondenzanalyse als alternatives Verfahren zur multivariaten Auswertung von Mobilitätsdaten vorgestellt. Um die Vor- und Nachteile der Korrespondenzanalyse gegenüber der multidimensionalen Skalierung herauszustellen,

¹ vgl. Herbert K a e l b l e, "Einleitung," in: Herbert K a e l b l e, Hrsg., Geschichte der sozialen Mobilität seit der industriellen Revolution (Königstein/ Ts., 1978), S.3-11 (S.3).

² Die MDS ist ein exploratives Analyseverfahren, mit dem (Un)ähnlichkeiten in Distanzen umgerechnet werden. Je größer die räumliche Distanz im (i.d.R.) zweidimensionalen Darstellungsraum, desto größer die Unähnlichkeit zwischen den Objekten. Da nur relativ selten (Un)ähnlichkeiten direkt erhoben werden (ein klassisches Beispiel ist die Entfernung zwischen Städten; in der Umfrageforschung wären es Paarvergleiche, wobei jedes Objekt mit jedem Objekt verglichen werden muß), müssen die zu analysierenden Daten umgerechnet werden, z.B. in Dissimilaritätsindizes oder Korrelationskoeffizienten. Eine gute Einführung in die multidimensionale Skalierung geben A. P. M. C o x o n, "The User's Guide to Multidimensional Scaling" (London, 1982) und Ingwer B o r g, "Anwendungsorientierte multidimensionale Skalierung," (Heidelberg, 1981).

³ Peter M. B l a u und Otis D. D u n c a n, "The American Occupational Structure," (New York, 1967); Heinrich B e s t, "Soziale Morphologie und politische Orientierungen bildungsbürgerlicher Abgeordneter in der Frankfurter Nationalversammlung und in der Pariser Assemblée nationale constituante 1848/49," in: Jürgen K o c k a, Hrsg., Bildungsbürgertum im 19. Jahrhundert, Teil IV (Stuttgart, 1989), S. 53-94; Heinrich B e s t, "Die Männer von Bildung und Besitz", (Düsseldorf, 1990).

verwenden wir den gleichen Datensatz, der auch Bests Analysen zugrunde liegt.

Best untersuchte die inter- und die intragenerationale Mobilität der Mitglieder der Frankfurter Nationalversammlung (FNV) von 1848/49 und der zur gleichen Zeit tagenden Mitglieder der Pariser Assemblée nationale constituante. Wir beschränken uns in unserem Methodenvergleich auf die Mitglieder der FNV, also auf eine Elite von Bildungsbürgern (87% hatten ein akademisches Bildungspatent, als Äquivalente galten lediglich Besitz, Abstammungsprestige oder ein Offizierspatent⁴) in einem fest definierten Kontext.

Die FNV kann als erstes demokratisches Parlament in Deutschland angesehen werden. Sie tagte vom Mai 1848 bis zum Mai 1849. Die wichtigste Aufgabe der FNV war, eine einheitliche Verfassung für einen großen Flächenstaat zu entwerfen. Die Parlamentarier wurden in Wahlkreisen bestimmt, wobei in jedem dieser Wahlkreise 1 Sitz vergeben werden konnte⁵. Nach Langewiesche (1978) waren die Wahlen zur Nationalversammlung im wesentlichen Honorationswahlen⁶, an denen etwa 80% der männlichen Bevölkerung teilnehmen durften⁷. Neben unterschiedlichen Versuchen, Teile der unterprivilegierten Bevölkerung (insb. Landarbeiter, Tagelöhner, Fabrikarbeiter und Dienstboten) die Wahlberechtigung vorzuenthalten, gab es bei den Ungebildeten und Besitzlosen ein großes Desinteresse, an der Wahl teilzunehmen⁸.

2. Die Studie von Heinrich Best

Best hat in den letzten Jahren einen Datensatz zusammengestellt⁹, in dem 809 Parlamentarier der FNV verzeichnet sind. Diese 809 Personen übten ins-

⁴ vgl. Best (1989), S. 58.

⁵ Frank E y c k, "Deutschlands große Hoffnung - Die Frankfurter Nationalversammlung 1848/49," (München, 1973), S.77.

⁶ Dieter L a n g e w i e s c h e, "Die Anfänge der deutschen Parteien. Partei, Fraktion und Verein der Revolution von 1848/49," in: Wolfgang J. M o m m s e n, Hrsg., Liberalismus im aufsteigenden Industriestaat (Göttingen, 1978), S. 325.

⁷ siehe Dieter L a n g e w i e s c h e, "Die deutsche Revolution von 1848/49 und die vorrevolutionäre Gesellschaft: Forschungsstand und Forschungsperspektiven," in: Archiv für Sozialgeschichte 21 (1981), S. 458-498 (S. 469).

⁸ Vgl. Theodore S. H a m e r o w, "Die Wahlen zum Frankfurter Parlament," in: Ernst-Wolfgang B ö c k e n f ö r d e, Hrsg., Moderne deutsche Verfassungsgeschichte (1815-1918) (Köln, 1972), S. 215-236 (S. 232).

⁹ Heinrich B e s t unter Mitarbeit von Rainer K u z n i a, "Die Abgeordneten der Frankfurter Nationalversammlung 1848/49," Datenhandbuch, Institut für Angewandte

gesamt 927 verschiedene Berufe aus, die von Best in 10 Berufsgruppen zusammengefaßt wurden. Die Gruppen sind differenziert nach funktionalen Bereichen und innerhalb derer weiter unterschieden nach funktionalen Differenzierungen. So z.B. der Staatsdienst: dieser wird unterschieden nach Bildung (einschließlich der Wissenschaft), Justiz, Verwaltung und Militär. Die verbleibenden 6 Berufsgruppen enthalten die Angehörigen des Klerus, der Unternehmen, der Gutsbesitzer, der Advokaten, der Freiberufler (dieses sind hauptsächlich Publizisten, Journalisten, Literaten und Künstler) und die der Residualkategorie "Kleinbürger" (klein- und unterbürgerliche Positionen). Best untersuchte in seinen Arbeiten u.a. die intra- und intergenerationale Mobilität der Parlamentarier innerhalb und zwischen den 10 genannten Berufsgruppen. Als multivariates Auswertungsverfahren verwendete er die multidimensionale Skalierung.

Zur Beschreibung der intergenerationalen Mobilität verwendete Best eine Tabelle der Zuströme der 10 unterschiedlichen Berufsgruppen vom Beruf des Vaters zum Hauptberuf des Sohnes im Jahre 1848/49¹⁰. Diese Daten mußten im ersten Schritt, wie bereits erwähnt, in (Un)ähnlichkeiten umgerechnet werden. Best benutzt hierfür den relativ häufig verwendeten Dissimilaritätsindex¹¹. Die Matrix der resultierenden (Un)ähnlichkeiten der Mobilitätsströmungen zwischen den zehn Berufsbereichen wurde im zweiten Schritt der Analyse in Distanzen (zwischen den Objekten, also den Berufsbereichen) übersetzt und in einem zweidimensionalen Raum graphisch dargestellt. Zusätzlich zu dieser Umrechnung in Distanzen wurde die Matrix der Dissimilaritätsindizes für eine hierarische Clusteranalyse verwendet, mittels derer Best vier Cluster von Berufsbereichen unterscheiden konnte¹². Das

Sozialforschung, Universität zu Köln (Köln, 1986). Das Datenhandbuch ist Bestandteil des Projektes "Struktur und Wandel parlamentarischer Führungsgruppen in Deutschland 1848-1935".

¹⁰ Die Abgeordneten übten zum Zeitpunkt ihrer Tätigkeit in der FNV teilweise mehrere Berufe aus. Best beschränkt sich in seinen multivariaten Analysen auf den in den Jahren 1848/49 als Hauptberuf angegebenen.

¹¹ Der Dissimilaritätsindex ist ein Index, der auf einer Abweichung einer jeden Beobachtung von einem zentralen Wert basiert. Der Dissimilaritätsindex wird in der Arbeit von Otis D. Duncan und Beverly Duncan, "A Methodological Analysis of Segregation Indexes," in: American Sociological Review 20 (1955), S. 210-217, ausführlich beschrieben und u.a. von Peter M. Blau und Otis D. Duncan (1967) als Ähnlichkeitsmaß zur Analyse von Mobilitätsbewegungen mittels der MDS verwendet.

¹² Da mit Hilfe der MDS (lediglich) (Un)ähnlichkeiten in Distanzen umgerechnet werden können, wird häufig zusätzlich eine Clusteranalyse gerechnet, um ein numerisches Kriterium zu haben, anhand dessen die Objekte zu Clustern zusammengefaßt werden können.

erste Cluster bezeichnete er als das Cluster der Unternehmer und Kleinbürgerlichen, das zweite als Cluster der Kleriker, der Freiberuflichen und der Bildungsberufe, das dritte als Cluster der Verwaltungsbeamten, Richter und Rechtsanwälte und das vierte als Cluster der Militärs und Gutsbesitzer. Die beiden Berufsgruppen des vierten Clusters liegen jedoch in der Abbildung räumlich ziemlich stark getrennt voneinander, d.h. es gibt nur relativ wenig Mobilität von der einen zu der anderen Berufskategorie.

Die vier identifizierten Cluster entsprechen den sozialen Grundfunktionen der strukturell funktionalen Systemtheorie von Parsons (1985)¹³. Das Cluster der Unternehmer und Kleinbürgerlichen entspricht der durch das wirtschaftliche Subsystem wahrgenommenen Anpassungsfunktion (adaption). Das zweite Cluster umfaßt die im Bildungsbereich Tätigen, es entspricht der durch das kulturelle Subsystem wahrgenommenen Funktion der Normerhaltung und Spannungsbewältigung (pattern maintenance). Das dritte Cluster entspricht der Funktion der sozialen Kontrolle und Aufrechterhaltung von Gesetznormen (integration) und das vierte Cluster, die Gutsbesitzer und Militärs, den Exponenten der traditionellen Führungsgruppen mit der Funktion der Zielerreichung (goal attainment). Die Hauptachse seiner zweidimensionalen Darstellung interpretierte Best als Statusdimension, wo auf der einen Seite die Kleinbürgerlichen und die Unternehmer als weniger Privilegierte, auf der anderen Seite die Gutsbesitzer stehen.

Für die Untersuchung der intragenerationalen Mobilität führte Best mit dem Adel eine elfte Gruppe ein. Als Ähnlichkeitsmaß zur multidimensionalen Skalierung verwendete er für diese Analyse den Phi-Koeffizienten¹⁴, womit er auf eine Unterscheidung in Zu- und Abströme verzichtet hat. Auch in dieser Analyse sind die Cluster des Bildungssystems und des wirtschaftlichen Subsystems gut sichtbar. Die anderen Berufskategorien liegen verstreut im Darstellungsraum und sind nicht mehr eindeutig einem Cluster zuordbar. Eine eindeutige inhaltliche Interpretation der Achsen ist nicht möglich¹⁵. Best kommt u.a. zu dem Ergebnis, daß Status zwar intergenerational geerbt, aber nicht im Verlauf intragenerationaler Mobilität erworben werden kann.

Als Eingangsdaten werden bei der multidimensionalen Skalierung (Un)ähnlichkeiten zwischen den betrachteten Objekten verwendet, daher

¹³ Talcott P a r s o n s, "Das System moderner Gesellschaften," (Weinheim, 1985), S. 20-29.

¹⁴ Bei Vier-Felder-Tabellen (in diesem Beispiel gibt es zu jedem anderen Berufsbereich einen Zustrom und einen Abstrom, sowie die Berufskonstanz) entspricht der Phi-Koeffizient dem Produktmomentkorrelationskoeffizient.

¹⁵ Vgl. Heinrich B e s t (1990), S. 111.

auch die von Best vorgenommenen Umrechnungen zum Dissimilaritätsindex bzw. zum Phi-Koeffizienten. Diese (Un)-ähnlichkeiten werden derart in Distanzen umgerechnet, daß im (zweidimensionalen) Darstellungsraum die in den Daten vorhandenen Disparitäten optimal wiedergegeben werden. Ist die Anpassung von Distanzen an Disparitäten vollständig, ist der Streßwert, als Maßzahl zur Messung der Güte der Anpassung, Null.

Ebenso wie es verschiedene Möglichkeiten zur Berechnung des Streßwertes gibt und verschiedene Ähnlichkeitsmaße verwendet werden können, gibt es verschiedene Möglichkeiten zur Berechnung der Distanzen im Darstellungsraum. Auch wenn in der Regel die euklidische Metrik verwendet wird, so ist dies keinesfalls zwingend. Da die Ergebnisse mit den unterschiedlichen Berechnungsgrundlagen variieren, sind Begründungen für die eine oder andere Metrik und für das verwendete Ähnlichkeitsmaß oft willkürlich. Diese Beliebigkeit in der Wahl von Distanz- und Ähnlichkeitsmaßen gibt es bei der von uns verwendeten Korrespondenzanalyse nicht, hier ist die Metrik vorgegeben und kann nicht verändert werden¹⁶. Als Eingabedaten werden in der Regel Kontingenztabelle benutzt¹⁷. Die Zeilen und Spalten dieser Tabellen werden als Vektoren aufgefaßt und in einen niederdimensionalen Raum projiziert, in dem die euklidische Metrik angenommen werden kann.

3. Zur Korrespondenzanalyse

Bei der Korrespondenzanalyse handelt es sich primär um ein Verfahren zur simultanen graphischen Darstellung von Zeilen und Spalten einer oder mehrerer Kontingenztabelle. Ähnlich wie bei der Hauptkomponentenanalyse, der multidimensionalen Skalierung oder der Clusteranalyse handelt es sich um ein Datenreduktionsverfahren: Vorhandene Informationen sollen durch möglichst wenige, orthogonale Faktoren (Dimensionen) abgebildet werden. Analog zur Hauptkomponentenanalyse werden die Achsen sukzes-

¹⁶ Es sei angemerkt, daß es verschiedene Algorithmen zur Berechnung gibt, die unter dem Namen "Korrespondenzanalyse" verwendet werden. Es handelt sich hierbei um verwandte Verfahren wie das "optimal scaling", das "reciprocal averaging" oder das "Biplot-Verfahren". Wir beziehen uns in unserer Darstellung auf die Arbeit von Michael G r e e n a c r e, "Theory and Applications of Correspondence Analysis," (London, 1984).

¹⁷ Es ist jedoch keinesfalls zwingend, nominal- oder ordinalskalierte Daten als Eingangsmaterial zu verwenden. Es können auch metrisch skalierte Variablen benutzt werden. So z.B. von Jens D a n g s c h a t und Jörg B l a s i u s, "Social and Spatial Disparities in Warsaw in 1978. An Application of Correspondence Analysis to a Socialistic City," in: Urban Studies 24 (1987), S. 173-191.

sive derart eingeführt, daß die Summe der quadrierten Abstände der Vektorendpunkte zu den Achsen jeweils minimal ist.

Die Korrespondenzanalyse wurde Anfang der 60er Jahre in Frankreich entwickelt¹⁸ und kann auf Hirschfeld (1935)¹⁹ zurückgeführt werden. Jener gab als erster eine algebraische Formulierung der Korrelation von Zeilen und Spalten einer Kontingenztafel. Die Korrespondenzanalyse gilt in Frankreich als sozialwissenschaftliche Standardmethode: So verwendete sie z.B. Bourdieu schon 1979 in seinem Buch "La Distinction"²⁰. In seinem empirischen Teil beschreibt Bourdieu mit Hilfe der Korrespondenzanalyse "Feine Unterschiede" innerhalb dreier von ihm differenzierter Klassen (Oberschicht, Mittelschicht, Arbeiterklasse) der Gesellschaft.

Blasius (1987)²¹ verwendete die Korrespondenzanalyse zur Beschreibung der ersten Mitgliedschaft in sechs verschiedenen Organisationen zwischen 1933 und 1942 (NSDAP, SS, SA, NSDSTP, andere, keine Mitgliedschaft). Als beschreibende Merkmale benutzte er die Variablen "Alter" (kategorisiert), "Beruf des Vaters", "Geschlecht", "Studienfach", "Erhalt des ersten Darlehens" (1933-38, 1939-42) sowie "Heimatland des Vaters" (Deutscher, Ausländer). Best (1989)²² verwendete das Verfahren zur Beschreibung von Strukturproblemen des deutschen Parlaments zwischen 1867 und 1933. Hierbei beschrieb er die politischen Konfliktgruppen zu verschiedenen Zeitpunkten innerhalb der genannten Zeitspanne. Als beschreibende Variablen benutzte er den Adelstitel, den Beruf, die Herkunftsregion und die Konfession der Abgeordneten. Dangschat und Blasius (1987)²³ verwendeten das Verfahren zur Beschreibung von räumlichen und sozialstrukturellen Ungleichheiten in Warschau. Mit Hilfe des Zensusdatensatzes von 1978 beschrieben sie die Stadtteile der polnischen Hauptstadt anhand von Variablen zum Gebäude-

¹⁸ Jean-Paul B e n z é c r i, "Cours de Linguistique Mathématique". Publications de l'Institut de Statistique de l'Université de Paris (Paris, 1964), Brigitte E s c o f i e r - C o d i e r, "L'Analyse de Correspondance," (Rennes, 1965).

¹⁹ H. O. H i r s c h f e l d, "A Connection Between Correlation and Contingency," in: Proceedings of the Cambridge Philosophical Society 31 (1935), S. 520-524.

²⁰ Pierre B o u r d i e u, "La Distinction. Critique Social du Jugement," (Paris, 1979).

²¹ Jörg B l a s i u s, "Korrespondenzanalyse - Ein multivariates Verfahren zur Analyse qualitativer Daten," in: Historical Social Research - Historische Sozialforschung 42/43 (1987), S. 172-189.

²² Heinrich B e s t, "Mandat ohne Macht. Strukturprobleme des deutschen Parlamentarismus 1867-1933," erscheint in: Heinrich B e s t, Hrsg., Politik und Milieu. Wahl- und Elitenforschung im historischen und interkulturellen Vergleich, (St.Katharinen, 1989).

²³ Jens D a n g s c h a t und Jörg B l a s i u s (1987).

alter (kategorisiert), dem Bauträger, der Schulbildung, der Haushaltsgröße und dem Alter der Bewohner (kategorisiert). Dabei konnten u.a. Stadtteile benannt und mit Hilfe sozialstruktureller Merkmale beschrieben werden, in denen Personen mit höherer Schulbildung und in denen Personen mit einfacher Schulbildung überdurchschnittlich häufig (selten) lebten. Weitere Anwendungsbeispiele sowie ausführliche Darstellungen des Verfahrens sind den Lehrbüchern von Greenacre (1984) und Lebart et al. (1984) zu entnehmen²⁴.

Wir haben die Korrespondenzanalyse verwendet, um intragenerationale und intergenerationale Mobilität innerhalb einer Analyse zu untersuchen. Als Eingabeinformation verwenden wir zwei Kreuztabellen: Die erste (N=657) besteht aus der Kreuztabellierung von "Hauptberuf 1848/49" mit "Beruf des Vaters", die zweite (N=750) aus "Hauptberuf 1848/49" mit "Erstberuf". Beide Tabellen wurden untereinander geschrieben und bilden das Eingangsmaterial der Korrespondenzanalyse (vgl. Matrix INPUT DATA). Die Berechnungen wurden mit KORRES durchgeführt, einem von Blasius und Rohlinger (1988)²⁵ in der matrixorientierten Sprache SAS - PROC MATRIX geschriebenen Programm. Mit KORRES ist es möglich, alle Zwischenschritte zur Erlangung des Ergebnisses zu erhalten. Wir geben hier im Text jedoch nur die wichtigsten an und verweisen für die statistischen Grundlagen auf den Anhang²⁶.

4. Ergebnisse

Eine erste generelle Inspektion der Daten liefert uns die Chi-Quadrat-Statistik (siehe Matrix CELLIN). In der Matrix CELLIN sind die quadrierten Abstände der empirischen von den erwarteten Werten, dividiert durch den jeweiligen Erwartungswert, angegeben. Die Summe aller Werte ist der Chi-Quadrat-Wert. Dieser wird, nach Division durch das Gesamt-N der Eingabe-

²⁴ Michael Greenacre (1984); Ludovic Lebart, Alain Morineau, Kenneth M. Warwick, "Multivariate Descriptive Statistical Analysis. Correspondence Analysis and Related Techniques for Large Matrixes," (New York, 1984).

²⁵ Jörg Blasius und Harald Rohlinger, "KORRES - Ein multivariates Programm zur Auswertung von zweidimensionalen Kontingenztabellen," in: Frank Faulbaum und Hans-Martin Uehlinger, Hrsg., "Fortschritte der Statistik-Software 1, (Stuttgart, 1988), S. 387-397.

²⁶ Die Korrespondenzanalyse ist inzwischen auch in den großen Statistikpaketen integriert.

INPUT DATA

A	HJUSTZ	HVERWLTG	HBILDUNG	HMILIT	HKLERUS	HUNTERNH	HGUTSBES	HADVOKAT	HFREIBF	HKLEINB
VJUSTIZ	22.00	15.00	12.00	0.00	0.00	1.00	4.00	15.00	1.00	0.00
VVERWLTG	27.00	34.00	20.00	4.00	4.00	2.00	8.00	24.00	5.00	0.00
VBILDUNG	6.00	16.00	9.00	2.00	7.00	1.00	0.00	6.00	8.00	0.00
VMILIT	8.00	8.00	1.00	3.00	0.00	0.00	1.00	3.00	2.00	0.00
VKLERUS	8.00	6.00	13.00	0.00	3.00	1.00	4.00	6.00	5.00	0.00
VUNTERNH	15.00	9.00	13.00	2.00	3.00	31.00	1.00	7.00	9.00	2.00
VGUTSBES	11.00	12.00	6.00	3.00	1.00	4.00	35.00	5.00	5.00	0.00
VADVOKAT	4.00	10.00	2.00	0.00	0.00	1.00	2.00	10.00	2.00	0.00
VFREIBF	5.00	6.00	8.00	0.00	2.00	3.00	0.00	8.00	8.00	0.00
VKLEINB	13.00	14.00	28.00	1.00	5.00	12.00	0.00	20.00	19.00	7.00
EJUSTIZ	117.00	83.00	19.00	0.00	1.00	6.00	17.00	76.00	6.00	0.00
EVERWLTG	11.00	37.00	6.00	0.00	0.00	1.00	7.00	7.00	6.00	0.00
EBILDUNG	4.00	3.00	67.00	1.00	4.00	0.00	3.00	5.00	13.00	0.00
EMILIT	0.00	5.00	1.00	16.00	0.00	2.00	8.00	1.00	0.00	0.00
EKLERUS	0.00	1.00	11.00	0.00	26.00	0.00	0.00	1.00	0.00	0.00
EUNTERNH	1.00	0.00	0.00	0.00	0.00	14.00	0.00	0.00	0.00	1.00
EGUTSBES	0.00	4.00	0.00	0.00	0.00	0.00	19.00	1.00	0.00	0.00
EADVOKAT	8.00	2.00	1.00	0.00	0.00	0.00	0.00	22.00	1.00	0.00
EFREIBF	0.00	7.00	20.00	0.00	4.00	4.00	1.00	0.00	37.00	0.00
EKLEINB	0.00	1.00	1.00	1.00	0.00	19.00	0.00	0.00	3.00	6.00

	EJUSTIZ	Erstberuf:	Justiz	HJUSTIZ	Hauptberuf	1848/49:	Justiz
VJUSTIZ	Beruf des Vaters:	Justiz		HJUSTIZ	Hauptberuf	1848/49:	Justiz
VVERWLTG	Beruf des Vaters:	Verwaltung		HVERWLTG	Hauptberuf	1848/49:	Verwaltung
VBILDUNG	Beruf des Vaters:	Bildungsbereich		HBILDUNG	Hauptberuf	1848/49:	Bildungsbereich
VMILIT	Beruf des Vaters:	Militär		MMILIT	Hauptberuf	1848/49:	Militär
VKLERUS	Beruf des Vaters:	Klerus		MKLERUS	Hauptberuf	1848/49:	Klerus
VUNTERNH	Beruf des Vaters:	Unternehmer		HUNTERNH	Hauptberuf	1848/49:	Unternehmer
VGUTSBES	Beruf des Vaters:	Gutsbesitzer		HGUTSBES	Hauptberuf	1848/49:	Gutsbesitzer
VADVOKAT	Beruf des Vaters:	Advokat		HADVOKAT	Hauptberuf	1848/49:	Advokat
VFREIBF	Beruf des Vaters:	freie Berufe		HFREIBF	Hauptberuf	1848/49:	freie Berufe
VKLEINB	Beruf des Vaters:	Kleinbürger		MKLEINB	Hauptberuf	1848/49:	Kleinbürger

INERTIA IN EACH CELL

CELLIN	HJUSTIZ.	IVERWLTG	IBILDUNG	HMILIT	HKLERUS	HUNTERNH	HGUTSBES	HAADVOKAT	HFREIBF	HKLEINB
VJUSTIZ	7.00	0.24	0.02	1.60	2.91	3.15	0.44	1.91	4.46	0.78
VVERWLTG	0.68	3.97	0.06	0.39	0.33	5.48	0.47	1.18	3.69	1.42
VBILDUNG	1.54	3.02	0.00	0.44	9.73	2.14	4.38	0.62	1.88	0.61
VMILIT	2.35	1.94	2.52	9.74	1.08	1.84	0.55	0.21	0.05	0.29
VKLERUS	0.01	0.84	3.87	1.05	0.62	1.56	0.03	0.12	0.18	0.51
VUNTERNH	0.15	4.05	0.31	0.00	0.18	92.38	5.46	3.37	0.06	0.94
VGUTSBES	0.96	0.79	4.18	0.68	1.70	0.55	124.11	4.35	0.77	0.91
VADVOKAT	0.45	2.92	1.89	0.71	1.29	0.65	0.09	6.12	0.22	0.34
VFREIBF	0.67	0.32	0.30	0.91	0.07	0.01	3.19	0.66	5.37	0.44
VKLEINB	3.90	3.80	2.80	1.19	0.00	1.20	2.41	0.10	5.50	23.04
EJUSTIZ	58.45	7.56	22.31	7.43	11.58	12.53	3.05	15.10	18.49	3.60
EVERWLTG	0.46	36.73	3.27	1.71	3.12	3.49	0.18	1.62	0.08	0.83
EBILDUNG	10.89	13.38	154.84	0.72	0.01	7.06	3.09	6.69	1.77	1.11
EMILIT	5.94	0.25	3.62	308.21	1.37	0.05	10.98	3.16	2.97	0.37
EKLERUS	7.02	5.51	3.25	0.89	366.78	2.75	3.11	4.03	3.51	0.43
EUNTERNH	1.23	3.02	2.64	0.37	0.66	146.55	1.27	2.40	1.44	3.82
EGUTSBES	4.32	0.06	3.96	0.55	1.00	1.70	152.78	1.88	2.16	0.27
EADVOKAT	0.58	3.05	3.78	0.78	1.41	2.40	2.71	55.84	1.39	0.38
EFREIBF	13.14	3.35	5.28	1.67	0.31	0.26	3.99	10.97	140.88	0.81
EKLEINB	5.58	4.03	3.31	0.12	1.29	129.05	2.47	4.66	0.02	93.15

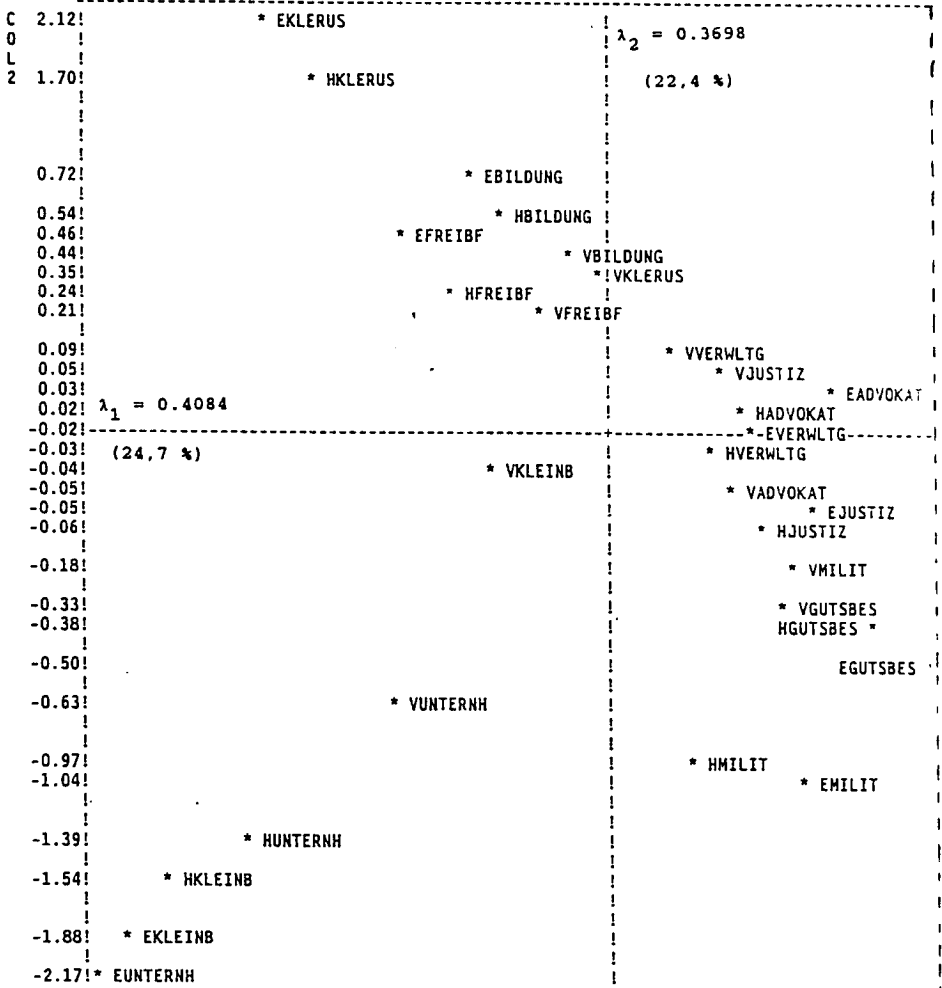
daten ($N=1407$), als "Gesamtträgheit" (total inertia) bezeichnet. Die "Gesamtträgheit" ist identisch mit der Summe der Eigenwerte; letztere ergeben sich zusammen mit den dazugehörigen Eigenvektoren bei der "Zerlegung" einer Matrix, in der alle Informationen der Eingabedaten enthalten sind (vgl. Anhang).

Die höchsten Werte der Matrix CELLIN, also die größten (quadierten und durch die jeweiligen Erwartungswerte dividierten) Abweichungen von empirischen und erwarteten Werten, stehen in der Hauptdiagonale der Kreuztabelle "Erstberuf" mit "Hauptberuf 1848/49". Somit hat die Berufskonstanz einen starken Einfluß auf den ausgeübten Hauptberuf. Wesentlich niedriger - aber in den meisten Fällen deutlich von Null verschieden - sind die Werte in der Hauptdiagonale von "Beruf des Vaters" mit "Hauptberuf 1848/49". Werden die beiden Mobilitätsarten miteinander verglichen, so hat die Berufsstabilität einen wesentlich stärkeren Einfluß auf den Hauptberuf 1848/49 als die Berufsvererbung.

Wie bereits erwähnt, wird der Darstellungsraum durch orthogonale Vektoren beschrieben. Anhand der zu jedem dieser Eigenvektoren gehörenden Eigenwerte können die erklärten Varianzen der einzelnen Faktoren bestimmt werden. Dabei erklärt die erste Achse 24,7% der Gesamtvariation, die zweite weitere 22,4% (vgl. EVAR). Somit werden durch die ersten beiden Achsen (lediglich) 47,1% der Gesamtvariation des Modells erklärt. Dies ist ein deutlicher Hinweis darauf, daß weitere relevante Dimensionen bestehen und beschrieben werden müssen.

EIGENVALUES		EXPLAINED VARIANCE	
LAMBOA	COL1	EVAR	COL1
1. AXIS	0.4084	1. AXIS	24.73
2. AXIS	0.3698	2. AXIS	22.40
3. AXIS	0.2969	3. AXIS	17.98
4. AXIS	0.2079	4. AXIS	12.59
5. AXIS	0.1758	5. AXIS	10.65
6. AXIS	0.0863	6. AXIS	5.23
7. AXIS	0.0497	7. AXIS	3.01
8. AXIS	0.0373	8. AXIS	2.26
9. AXIS	0.0191	9. AXIS	1.16
10. AXIS	-0.0000	10. AXIS	-0.00

MODIFIED RANK PLOT OF COL2 BY COL1



2	1	1	1	1	00	0	0	00	0	0	0	0	0	0	0	0	0	
0	9	7	3	3	1	77	5	5	44	2	1	0	2	3	444444	4455	5	7
0	1	5	7	4	4	76	6	0	21	4	5	6	9	6	000125	9902	7	1

COL1

Die geometrische Lage von Zeilen- und Spaltenvariablen der Ausgangstabelle ist für die ersten beiden Achsen in der nachfolgenden Abbildung dargestellt. Um eine übersichtliche graphische Darstellung zu erhalten, wurde ein (modifizierter) Rank-Plot²⁷ verwendet; die exakte Lage von Zeilen- und Spaltenvariablen ist an den Achsen angegeben.

Im linken oberen Quadranten ist ein Cluster²⁸ zu erkennen, dem die Ausprägungen der "Freien Berufe" (EFREIBF, HFREIBF, VFREIBF) und der "Bildungsberufe" (EBILDUNG, HBILDUNG, VBILDUNG) zuzurechnen sind. Ebenfalls zu diesem Cluster können die drei Kategorien des "Klerus" (EKLERUS, HKLERUS, VKLERUS) gezählt werden. Da zur Zeit der Frankfurter Nationalversammlung dem Klerus die Kontrolle der Bildung oblag und die freien Berufe insbesondere Journalisten, Publizisten, Künstler und Literaten umfaßte, entsprechen diesem Cluster jene Gruppen, deren gesellschaftliche Funktion Parsons (1985)²⁹ als "pattern maintenance" bezeichnete.

Im linken unteren Quadranten befinden sich die sechs Ausprägungen der "Unternehmer" (EUNTERNH, HUNTERNH, VUNTERNH) und der "Kleinbürger" (EKLEINB, HKLEINB, VKLEINB). Jene bilden das Cluster, welches nach Parsons dem wirtschaftlichen Subsystem entspricht, dessen Gruppenmitglieder durch Anpassung (adaption) gekennzeichnet sind. Auch die Berufsgruppen der beiden anderen Parsonschens Systeme clustern in der graphischen Darstellung. So sind die neun Kategorien, die zum System der gesellschaftlichen Gemeinschaft gehören und deren Hauptfunktion die soziale Kontrolle ist (integration) (-VERWLTG, -ADVOKAT, -JUSTIZ), im positiven Bereich der ersten Achse lokalisiert. Die sechs Kategorien der traditionellen Führungsgruppen (-GUTSBES, -MILIT) befinden sich im rechten

²⁷ Mit der SAS Prozedur PROC IDPLOT (vgl. Warren F. Kuhfeld, "Metric and Nonmetric Plotting Models," in: Psychometrika 51 (1986), S. 155-161) ist es möglich, Variablen kreuzweise zu plotten. Als Labels für die Variablen sind bei KORRES acht Stellen vorgesehen. Da es bei einer metrischen Ausgabe des Plots zu einer Überschreibung der Labels kommt, haben wir uns für die Ausgabe in Form eines (modifizierten) Rank-Plots entschieden.

²⁸ Für die Zuordnung eines Objektes zu einem Cluster ist bei der Korrespondenzanalyse der Winkel zu dem Achsenkreuz entscheidend, weniger die räumliche Nähe zwischen den Objekten. Hier ist ein wichtiger Unterschied in der Interpretation zwischen auf Vektoren basierenden Modellen (z.B. der Hauptkomponentenanalyse, der Korrespondenzanalyse) und auf Distanzen basierenden Modellen (z.B. der multidimensionalen Skalierung, der Clusteranalyse).

²⁹ Talcott Parsons (1985), S. 20-29.

unteren Quadranten. In allen Fällen dürfte der Einfluß der Berufsvererbung geringer als der der Berufsstabilität sein. Graphisch läßt sich dies daran erkennen, daß die jeweils gleiche Kategorie vom "Beruf des Vaters" näher beim Achsenkreuz liegt als die vom "Erstberuf"³⁰. Auf die Zusammenführung beider Mobilitätsarten und deren direkten Vergleich mußte Best bei Verwendung der multidimensionalen Skalierung verzichten.

Die Interpretationsmöglichkeiten der Korrespondenzanalyse sind jedoch nicht auf die graphische Ausgabe beschränkt, sondern es gibt auch eine numerische Darstellung. Wir haben die ersten acht der neun existierenden Achsen³¹ in die von uns als GENERAL STATISTIC bezeichnete numerische Ausgabe aufgenommen.

In den Spalten QCOR1 bis QCOR8 sind die (quadierten) Korrelationen - diese sind analog interpretierbar zu den (quadrierten) Ladungen bei der Hauptkomponentenanalyse - mit den ersten acht Achsen dargestellt. Die zeilenweisen Summen der quadrierten Korrelationen der berücksichtigten Faktoren sind in der Spalte SQCOR dargestellt. Die Komponenten von SQCOR sind den Kommunalitäten der Hauptkomponentenanalyse vergleichbar. In den Spalten LOC1 bis LOC8 ist dargestellt, auf welcher Seite und wie weit die einzelnen Variablenausprägungen vom Achsenkreuz (dem Nullpunkt) entfernt sind.

In den Spalten INR1 bis INR8 (inertia, Trägheitsgewicht) ist dargestellt, wie stark die Achsen durch die einzelnen Variablenausprägungen determiniert werden. Ihre Summen sind für Spalten- und Zeilenvariablen getrennt auf Eins normiert. Während mit den Koeffizienten QCOR1 bis QCOR8 beschrieben wird, welcher Anteil der Varianz der einzelnen Variablenausprägungen (sowohl der Zeilen- wie auch der Spaltenvariablen) durch die einzelnen Achsen determiniert wird, gilt bei den Koeffizienten INR1 bis INR8 die umgekehrte Logik. Hier wird die erklärte Varianz der Achsen durch die Variablenausprägungen angegeben. Die zeilenweisen Summen dieser Werte sind - gewichtet mit den Erklärungsanteilen der einzelnen Achsen, den Eigenwerten - in der Spalte INR aufgeführt³². In dieser Spalte ist abzulesen,

³⁰ Mit Hilfe der Korrespondenzanalyse kann neben der erklärten Varianz der Variablen(ausprägungen) durch die einzelnen Achsen auch die erklärte Varianz der Achsen durch die einzelnen Variablen(ausprägungen) - diese wird als Trägheitsgewicht (inertia) bezeichnet - angegeben werden. In die Berechnung dieser Trägheitsgewichte geht die (quadrierte) Distanz der Projektion der Objekte auf die Achsen zum Achsenkreuz ein (vgl. Anhang).

³¹ Aufgrund der verwendeten Metrik existiert immer eine Dimension weniger, als es dem Minimum von Zeilen/ Spalten entspricht (vgl. Anhang).

³² Berechnungsgrundlage für INR sind alle Achsen des (n-1)-dimensionalen Raumes.

wie stark die geometrische Ausrichtung des Gesamtmodells - wiederum getrennt für Zeilen und Spalten - von den unterschiedlichen Ausprägungen der Berufsvariablen determiniert wird. In der Spalte MASS ist die relative Häufigkeit der Variablenausprägungen aufgeführt, wiederum getrennt für Zeilen und Spalten auf Eins normiert. Die wichtigsten Interpretationshilfen sind die quadrierten Korrelationen der Variablenausprägungen mit den Achsen (QCOR1 bis QCOR8), da hier abzulesen ist, welche Variablenausprägungen auf welcher Achse wie stark laden.

Wählen wir als "Schwellenwert" für die Zuordnung einer Variablenausprägung zu einer Achse .25 (d.h. 25% ihrer Gesamtvarianz werden durch die jeweilige Achse determiniert; bei der Hauptkomponentenanalyse wäre die analoge Faktorladung .50), so laden auf der ersten Achse im positiven Bereich - die Vorzeichen sind der Spalte LOC1 zu entnehmen - die Variablenausprägungen "Hauptberuf Justiz", "Hauptberuf Verwaltung" und "Hauptberuf Advokaten"- Auf dem negativen Abschnitt dieser Achse laden die Variablenausprägungen "Hauptberuf Unternehmer" und "Hauptberuf Kleinbürgerlicher". Inhaltlich bedeutet dies, daß es einen hohen positiven Zusammenhang bezüglich inter- und intragenerationaler Mobilität zwischen den Hauptberufen "Justiz", "Verwaltung" und "Advokaten" gibt, ebenso wie zwischen den Berufsbereichen "Unternehmer" und "Kleinbürgerliche". Dementgegen sind die Mobilitätsraten zwischen diesen Gruppierungen sehr gering.

Von den Kategorien beim "Beruf des Vaters" laden auf dem positiven Abschnitt der ersten Achse "Justiz", "Verwaltung", "Militär" und "Advokat", auf dem negativen Abschnitt "Unternehmer" und "Kleinbürgerlicher". Bei den "Anfangsberufen" laden die Kategorien "Justiz" und "Verwaltung" auf dem positiven Abschnitt der ersten Achse, die des "Unternehmers" und des "Kleinbürgers" auf dem negativen Abschnitt. Entsprechend den Parsonischen Subsystemen kann die erste Achse als Gegensatz von "adaption" und "integration" interpretiert werden. Es gibt sowohl eine hohe Berufskonsistenz als eine hohe intergenerationale Mobilität zwischen den Berufsbereichen "Justiz", "Verwaltung" und "Advokaten" auf der einen und "Kleinbürgerlichen" und "Unternehmern" auf der anderen Seite. Während Berufswechsel vom Vater zum Sohn innerhalb der zwei Gruppierungen überdurchschnittlich häufig vorkommen, gibt es zwischen den beiden Gruppierungen nur wenige bzw. gar keine. Ebenso ist eine hohe Berufsstabilität und eine hohe intragenerationale Mobilität innerhalb, wiederum aber nicht zwischen den Berufen des Rechtssystems und den Berufen des wirtschaftlichen Subsystems zu verzeichnen. Eine Ausnahme ist der "Erstberuf: Advokat", der sich nicht in den Gegensatz von "adaption" und "integration" eingliedern läßt,- auch wenn

GENERAL STATISTIC

GENSTAT	MASS	SQCOR	INR	LOC1	QCOR1	INR1	LOC2	QCOR2	INR2	LOC3	QCOR3	INR3	LOC4	QCOR4	INR4
IJUSTIZ	0.180	0.885	0.053	0.445	0.411	0.087	-0.064	0.009	0.002	-0.398	0.329	0.096	-0.207	0.089	0.037
HVERMLTG	0.189	0.957	0.041	0.396	0.433	0.073	-0.025	0.002	0.000	-0.132	0.048	0.011	-0.083	0.019	0.006
HBILDUNG	0.165	0.999	0.093	-0.410	0.180	0.068	0.536	0.308	0.128	0.066	0.005	0.002	0.502	0.270	0.200
HNILIT	0.023	1.000	0.142	0.357	0.012	0.007	-0.966	0.091	0.058	2.342	0.534	0.422	-0.195	0.004	0.004
HKLERUS	0.042	1.000	0.170	-1.141	0.193	0.132	1.703	0.429	0.326	0.691	0.071	0.067	-1.407	0.293	0.395
HUNTERHH	0.071	0.998	0.174	-1.366	0.459	0.323	-1.392	0.476	0.370	-0.201	0.010	0.010	-0.317	0.025	0.034
HGUTSBES	0.080	1.000	0.136	0.707	0.177	0.097	-0.378	0.051	0.031	1.059	0.397	0.301	0.148	0.008	0.008
HADVOKAT	0.150	0.961	0.052	0.408	0.289	0.061	0.019	0.001	0.000	-0.415	0.299	0.087	-0.180	0.057	0.024
HFREIBF	0.090	0.997	0.082	-0.555	0.205	0.068	0.243	0.039	0.014	-0.008	0.000	0.000	-0.816	0.444	0.288
HKLEINB	0.011	0.983	0.056	-1.749	0.365	0.083	-1.539	0.283	0.071	-0.308	0.011	0.004	-0.236	0.007	0.003
VJUSTIZ	0.048	0.977	0.009	0.399	0.495	0.019	0.053	0.009	0.000	-0.319	0.316	0.017	-0.044	0.006	0.000
VVERMLTG	0.089	0.959	0.007	0.295	0.628	0.019	0.089	0.058	0.002	-0.053	0.021	0.001	-0.077	0.043	0.003
VBILDUNG	0.038	0.967	0.010	-0.151	0.052	0.002	0.439	0.436	0.020	0.095	0.020	0.001	-0.126	0.036	0.003
VHILIT	0.018	0.948	0.009	0.494	0.309	0.011	-0.184	0.043	0.002	0.187	0.044	0.002	-0.098	0.012	0.001
VKLERUS	0.032	0.985	0.004	0.057	0.017	0.000	0.352	0.647	0.011	0.018	0.002	0.000	0.163	0.139	0.004
VUNTERHH	0.064	1.000	0.045	-0.767	0.506	0.092	-0.631	0.343	0.069	-0.166	0.024	0.006	-0.143	0.018	0.006
VGUTSBES	0.057	0.996	0.058	0.490	0.141	0.033	-0.331	0.064	0.017	0.812	0.389	0.126	0.129	0.010	0.005
VADVOKAT	0.021	0.639	0.006	0.401	0.339	0.008	-0.048	0.005	0.000	-0.298	0.187	0.006	-0.060	0.008	0.000
VFREIBF	0.028	0.959	0.005	-0.244	0.199	0.004	0.211	0.150	0.003	-0.223	0.167	0.005	0.209	0.146	0.006
VKLEINB	0.086	0.994	0.018	-0.416	0.489	0.036	-0.040	0.004	0.000	-0.105	0.031	0.003	0.179	0.091	0.013
EJUSTIZ	0.225	0.969	0.067	0.518	0.544	0.148	-0.048	0.005	0.001	-0.398	0.321	0.120	-0.210	0.090	0.048
EVERMLTG	0.052	0.896	0.022	0.422	0.259	0.023	-0.019	0.001	0.000	-0.113	0.019	0.002	0.059	0.005	0.001
EBILDUNG	0.069	0.999	0.084	-0.497	0.124	0.042	0.716	0.257	0.096	0.157	0.012	0.006	0.809	0.328	0.218
EHILIT	0.023	1.000	0.141	0.503	0.176	0.014	-1.038	0.106	0.067	2.476	0.601	0.472	-0.177	0.003	0.003
EKLERUS	0.027	1.000	0.167	-1.339	0.176	0.119	2.115	0.439	0.327	0.854	0.072	0.066	-1.761	0.304	0.403
EUINTERHH	0.011	0.991	0.069	-1.998	0.391	0.108	-2.167	0.460	0.141	-0.404	0.016	0.006	-0.669	0.044	0.024
EGUTSBES	0.017	0.998	0.071	1.006	0.144	0.041	-0.498	0.035	0.011	1.467	0.306	0.120	0.210	0.006	0.004
EADVOKAT	0.024	0.944	0.030	0.569	0.152	0.019	0.031	0.000	0.000	-0.036	0.215	0.006	-0.289	0.039	0.009
EFREIBF	0.051	0.996	0.076	-0.756	0.231	0.071	0.460	0.085	0.029	0.078	0.002	0.001	-0.989	0.395	0.236
EKLEINB	0.021	0.997	0.102	-1.907	0.463	0.191	-1.878	0.449	0.205	-0.202	0.005	0.003	-0.338	0.015	0.012

GENERAL STATISTIC - CONTINUE

GENSTAT	LOC5	QCOR5	INR5	LOC6	QCOR6	INR6	LOC7	QCOR7	INR7	LOC8	QCOR8	INR8
HJUSTIZ	0.075	0.012	0.006	-0.056	0.006	0.006	0.070	0.010	0.018	-0.095	0.019	0.043
HVERWLTG	0.050	0.007	0.003	0.234	0.151	0.120	0.282	0.220	0.303	0.168	0.078	0.142
HBILDUNG	0.084	0.008	0.007	-0.450	0.217	0.387	0.105	0.012	0.037	0.025	0.001	0.003
HMILT	1.919	0.358	0.479	-0.012	0.000	0.000	-0.092	0.001	0.004	-0.025	0.000	0.000
HKLERUS	-0.227	0.008	0.012	0.213	0.007	0.022	-0.052	0.000	0.002	-0.017	0.000	0.000
HUNTERHH	-0.168	0.007	0.011	-0.060	0.001	0.003	0.135	0.004	0.026	-0.253	0.016	0.121
HGUTSBES	-1.012	0.363	0.464	-0.071	0.002	0.005	-0.087	0.003	0.012	-0.025	0.000	0.001
HADVOKAT	0.132	0.030	0.015	-0.100	0.017	0.017	-0.390	0.264	0.461	-0.040	0.003	0.006
HFREIBF	0.042	0.001	0.001	0.647	0.279	0.436	-0.180	0.022	0.059	-0.091	0.006	0.020
HKLEINB	-0.220	0.006	0.003	-0.173	0.004	0.004	-0.596	0.042	0.079	1.492	0.266	0.662
VJUSTIZ	0.042	0.005	0.000	-0.210	0.137	0.025	0.049	0.008	0.002	-0.023	0.002	0.001
VVERWLTG	0.133	0.128	0.009	-0.043	0.013	0.002	0.080	0.047	0.012	0.055	0.022	0.007
VBILDUNG	0.226	0.116	0.011	0.330	0.247	0.048	0.137	0.042	0.014	0.090	0.018	0.008
VHILT	0.579	0.424	0.034	0.244	0.075	0.012	0.177	0.039	0.011	0.041	0.002	0.001
VKLERUS	-0.099	0.051	0.002	-0.145	0.110	0.008	0.000	0.000	0.000	-0.059	0.018	0.003
VUNTERHH	0.012	0.000	0.000	-0.042	0.001	0.001	0.154	0.020	0.030	-0.319	0.088	0.174
VGUTSBES	-0.807	0.385	0.211	-0.013	0.000	0.000	-0.050	0.001	0.003	-0.096	0.005	0.014
VADVOKAT	0.014	0.000	0.000	0.143	0.043	0.005	-0.144	0.044	0.009	0.077	0.013	0.003
VFREIBF	0.106	0.038	0.002	0.182	0.111	0.011	-0.155	0.081	0.013	-0.143	0.068	0.015
VKLEINB	-0.007	0.000	0.000	-0.027	0.002	0.001	-0.245	0.170	0.104	0.270	0.206	0.168
EJUSTIZ	0.047	0.004	0.003	-0.008	0.000	0.000	0.029	0.002	0.004	-0.036	0.003	0.008
EVERWLTG	-0.092	0.012	0.002	0.361	0.190	0.079	0.451	0.297	0.213	0.280	0.114	0.109
EBILDUNG	0.125	0.008	0.006	-0.719	0.259	0.415	0.148	0.011	0.031	0.015	0.000	0.000
EMILT	1.643	0.265	0.351	-0.026	0.000	0.000	-0.105	0.001	0.005	-0.045	0.000	0.001
EKLERUS	-0.292	0.008	0.013	0.064	0.000	0.001	-0.034	0.000	0.001	-0.004	0.000	0.000
EUNTERHH	-0.372	0.014	0.009	-0.228	0.005	0.007	0.381	0.014	0.032	-0.695	0.047	0.143
EGUTSBES	-1.878	0.502	0.333	-0.072	0.001	0.001	-0.171	0.004	0.010	0.033	0.000	0.000
EADVOKAT	0.261	0.032	0.009	-0.199	0.019	0.011	-0.995	0.466	0.470	-0.208	0.020	0.027
EFREIBF	0.032	0.000	0.000	0.798	0.257	0.373	-0.144	0.008	0.021	-0.197	0.016	0.053
EKLEINB	-0.180	0.004	0.004	-0.051	0.000	0.001	-0.183	0.004	0.014	0.675	0.058	0.262

diese Variablenausprägung tendenziell auf dem positiven Abschnitt der ersten Achse lädt.

Wie bereits angemerkt, hat auch Best den Gegensatz von "adaption" und "integration" bei seiner Analyse der intergenerationalen Mobilität nachgewiesen. Ferner vermutete er ihn für die intragenerationale Mobilität, ohne dies aber mittels seiner multivariaten Analyse bestätigen zu können. Durch die Anwendung der MDS war es ihm jedoch nicht möglich anzugeben, welche (relative) Bedeutung die relativ starken Mobilitätsbewegungen innerhalb und die nicht bzw. kaum vorhandenen Mobilitätsbewegungen zwischen den beiden Subsystemen im Vergleich zu den anderen - noch zu beschreibenden - Mobilitätsströmungen haben. Mit Hilfe unserer Analyse werden durch den Gegensatz der beiden Subsysteme und durch deren gleichzeitig vorhandene hohe interne Zusammengehörigkeiten (jeweils bezogen auf die Mobilitätsbewegungen vom "Beruf des Vaters" und vom "Erstberuf" zum "Hauptberuf 1848/49") mit der ersten Achse 24,7% der Gesamtvariation erklärt.

Die zweite Achse erklärt 22,4% der Gesamtvariation des Modells. Sie kann als Gegensatz von "pattern maintenance" und "adaption" interpretiert werden. Im negativen Bereich der zweiten Achse laden auf der Ebene der Spaltenvariablen, also der "Hauptberufe 1848/49", die Berufskategorien "Unternehmer" und "Kleinbürger" und im positiven Bereich die Berufskategorien "Bildung" und "Klerus". Dies korrespondiert, mit Ausnahme bei der intergenerationalen Mobilität bei der Kategorie "Kleinbürger", mit den jeweiligen Kategorien beim "Erstberuf" und dem "Beruf des Vaters". Neben einer hohen Berufsstabilität gibt es somit auch innerhalb dieser zwei Gruppierungen eine hohe intergenerationale und eine hohe intragenerationale Mobilität, jedoch nicht zwischen ihnen. Im Gegensatz zur graphischen Darstellung fehlen die drei Kategorien der "Freien Berufe", - ihre quadrierten Korrelationen liegen unterhalb des von uns angegebenen "Schwellenwertes".

Nachdem "adaption" auch schon einen Gegensatzpol auf der ersten Achse bildete, dort zu "integration", sind die beiden genannten Berufskategorien dieses Subsystems auch für die Interpretation der zweiten Achse wichtig - hier als Gegensatz zu dem als "pattern maintenance" bezeichneten Subsystem. Da entsprechend dem Modell der Korrespondenzanalyse alle Achsen unkorreliert sind, gibt es keinen statistischen Zusammenhang zwischen den beiden Gegensätzen, zumindest nicht auf den ersten beiden Achsen. Inhaltlich bedeutet dies, daß es - zusätzlich zu einer hohen Berufsstabilität - sowohl interals auch intragenerationale Mobilität innerhalb des wirtschaftlichen Subsystems, des kulturellen Subsystems mit seinen Funktionen der latenten Normerhaltung der Spannungsbewältigung und des Subsystems der sozialen

Kontrolle mit seiner Funktion zur Aufrechterhaltung von Gesetznormen gibt. Ferner gibt es keine bzw. nur eine unterdurchschnittlich hohe intra- und intergenerationale Mobilität zwischen den Kategorien des wirtschaftlichen Subsystems und den anderen beiden Systemen. Die Mobilitätsbewegungen zwischen den Systemen der "integration" und der "pattern maintenance" können als durchschnittlich häufig angesehen werden, beide Systeme sind hinsichtlich der verwendeten Variablen unkorreliert.

Die dritte Achse erklärt 18,0% der Gesamtvariation. Hier sind als Gegensatz die Subsysteme "integration", dargestellt durch die Berufskategorien "Justiz" und "Advokaten", und des "goal attainment", dargestellt durch die Berufskategorien "Militär" und "Gutsbesitzer", nachzuweisen. Während die Berufsstabilität beim "Militär" mit quadrierten Korrelationen von über .50 (vgl. HMILIT und EMILIT in Spalte QCOR3) überdurchschnittlich hoch ist und eine hohe intergenerationalen Mobilität zum "Gutsbesitzer" sowie eine sehr geringe intergenerationalen Mobilität zum "Advokaten" bzw. in die "Justiz" vorhanden ist, ist das Ausmaß der Berufsvererbung beim "Militär" durchschnittlich. Die Söhne, deren Väter beim Militär das Offizierspatent hatten, gingen - relativ gesehen - ebenso häufig in die Justiz, die Verwaltung oder wurden Advokaten (vgl. 1. Achse, Spalte QCOR1).

Auf der vierten Achse, die 12,6% der Gesamtvariation erklärt, läßt von den Variablenausprägungen "Beruf des Vaters" keine. Einen Gegensatz in bezug auf die intragenerationale Mobilität bilden hier die Subsysteme des "pattern maintenance", dargestellt durch "Bildung" und "Freie Berufe", und des "Militärs" als Teil des "goal attainment".

Die fünfte Achse erklärt 10,7% der Gesamtvariation. Diese ist zu interpretieren als Gegensatz von "Militär" und "Gutsbesitzer", sowohl bei der inter- als auch bei der intragenerationalen Mobilität. Somit gibt es auch innerhalb des von Parsons als "goal attainment" bezeichneten Subsystems Unterschiede. Inhaltlich bedeutet dies, daß Berufsstabilität und Berufsvererbung stärker sind als die Mobilitätsbewegungen innerhalb des jeweiligen Subsystems. Gegensätze gibt es auch innerhalb der anderen Parsonischen Subsysteme; sie sind auf der sechsten und siebten Achse abgebildet. Zum einen sind dies die Berufskategorien "Bildung" versus "Freie Berufe" und zum anderen die Berufskategorien "Verwaltung" versus "Advokaten". Insgesamt gesehen sind jedoch die Unterschiede hinsichtlich der inter- und der intragenerationalen Mobilität zwischen den Gruppierungen wesentlich bedeutsamer als innerhalb der Gruppierungen. Auf die achte Achse wird wegen ihrer geringen Varianzerklärung nicht eingegangen.

Eine weitere wichtige Hilfe zur Interpretation der Ergebnisse ist das "Trägheitsgewicht" (inertia)³³. Hier ist abzulesen, wie stark der Einfluß der einzelnen Variablenausprägungen auf die geometrische Ausrichtung der Achsen ist. Somit kann mittels dieses Koeffizienten auch angegeben werden, welche Art von Mobilität einen stärkeren Einfluß auf den Darstellungsraum hat. Auf der Ebene der Spaltenvariablen, also des "Hauptberufes 1848/49", ist die Variablenausprägung "Unternehmer" mit einem Erklärungsanteil von 32,3% am dominantesten zur Bestimmung der ersten Achse (vgl. Spalte INR1). Auf der Ebene der Zeilenvariablen, also beim "Beruf des Vaters" und beim "Erstberuf", hat bei der intergenerationalen Mobilität lediglich der Beruf "Unternehmer" mit 9,2% erklärter Varianz einen relativ hohen Anteil an der Determination der geometrischen Ausrichtung der ersten Achse; bei der intragenerationalen Mobilität haben vier Variablenausprägungen einen Anteil von über 10% Erklärungskraft: es sind dies die Erstberufe in "Justiz", als "Unternehmer", beim "Klerus" und als "Kleinbürger". Werden die Trägheitsgewichte für inter- und intragenerationale Mobilität (eingeschlossen der Berufsstabilität auf seiten der intragenerationalen Mobilität und der Berufsvererbung auf seiten der intergenerationalen Mobilität) insgesamt betrachtet, so haben die "Erstberufe" einen Anteil von 77,6% an der geometrischen Ausrichtung der ersten Achse. Die "Berufe der Väter" die restlichen 22,4%³⁴.

Ähnliches ist auch für die anderen Achsen festzustellen. So hat für die geometrische Ausrichtung der zweiten Achse beim "Beruf des Vaters" lediglich die Berufskategorie "Unternehmer" einen deutlichen Einfluß (INR2=.069). Das Verhältnis der Erklärungsanteile von intra- zu intergenerationaler Mobilität (inbegriffen auch hier Berufsstabilität und Berufsvererbung) beträgt für diese Achse 87,7:12,2.

³³ Das Trägheitsgewicht einer Achse wird aus dem Quadrat der Entfernung zum Nullpunkt (dem Achsenkreuz), multipliziert mit der relativen Häufigkeit der Variablenausprägung, seiner Masse (vgl. Spalte MASS) berechnet. Um zu einer vergleichenden Darstellung zu gelangen, werden Zeilen- und Spaltenträgheitsgewichte auf Eins normiert.

³⁴ Die Prozentwerte ergeben sich durch Addition der einzelnen Werte in der Spalte INR1.

5. Zusammenfassung

Werden die Ergebnisse insgesamt betrachtet, so gibt es zwar in nahezu allen der zehn verwendeten Kategorien eine Berufsvererbung- eine Ausnahme bildet bei den "Berufen der Väter" lediglich die Kategorie "Freie Berufe", diese konnten keiner gemeinsamen Achse mit den "Hauptberufen" der Söhne von 1848/49 eindeutig zugeordnet werden. Mittels unserer Analyse war es möglich, die schon von Best nachgewiesenen vier Parsonschen Subsysteme zu bestätigen, weiterhin konnte aber mit der Korrespondenzanalyse möglich, die "Wichtigkeit" der einzelnen Mobilitätsströme, ebenso wie Unterschiede innerhalb der Parsonschen Subsysteme differenziert werden. Ferner konnten als weiterführende Ergebnisse zum einen zusätzliche Dimensionen aufgespürt und beschrieben werden, zum anderen konnten inter- und intra-generationale Mobilität innerhalb einer einzigen Analyse beschrieben und deren Unterschiedlichkeiten angegeben werden. Hierbei wurde nachgewiesen, daß der Erstberuf einen wesentlich stärkeren Einfluß auf den Hauptberuf im Jahre 1848/49 als die Berufsvererbung hat.

Die Anwendung der Korrespondenzanalyse in der Mobilitätsforschung hat ferner den Vorteil, daß die Originaldaten direkt in die Analyse eingehen können und nicht erst in (Un)ähnlichkeitsmaße umgerechnet werden müssen. Durch diese direkte Verwendungsmöglichkeit entfällt zum einen die vorhandene Willkürlichkeit bei der Entscheidung für oder gegen die Wahl eines bestimmten Maßes (Dissimilaritätsindex, Phi-Koeffizient, ...), und zum anderen wird die Häufigkeit von Zeilen und Spalten (im Beispiel der Mobilitätsdaten also die unterschiedliche Vorkommenshäufigkeit der einzelnen Berufsbereiche, sowohl die des Vaters, des Erstberufes als auch des Hauptberufes 1848/49) berücksichtigt (vgl. Anhang). Schließlich entfällt auch die Wahl einer geeigneten Metrik zur Berechnung der Distanzen zwischen den Objekten (City-Block-Metrik, euklidische Metrik, ...).

Auch wenn bei Anwendung der MDS in der Regel die euklidische Metrik verwendet wird, so ist dies keineswegs zwingend. Der größte Vorteil der MDS ist die graphisch direkt übersetzbare Darstellung von (Un)ähnlichkeiten von Objekten als Distanzen; dieses ist bei der Korrespondenzanalyse umständlicher, da hier die (Un)ähnlichkeit von Objekten nur anhand ihrer Winkel zum Achsenkreuz erkennbar wird. Die Vorteile der Korrespondenzanalyse liegen in den breiten Anwendungsmöglichkeiten insbesondere bei nominal skalierten Daten und in den guten Interpretationsmöglichkeiten der numerischen Ergebnisse.

Anhang

Verbindung von Zeilen- und Spaltendarstellung

In diesem Anhang soll gezeigt werden, daß sich Zeilen und Spalten in einem Raum abbilden lassen. Der hier angeführte Beweis geht auf die Eckart-Young-Zerlegung (Eckart & Young, 1936)³⁵ zurück. Vektoren und Matrizen wurden zur besseren Übersicht unterstrichen.

$\underline{A} = (a_{ij})$ ist die Ausgangsmatrix mit n-Zeilen und m-Spalten.

$$\sum_{i=1}^n \sum_{j=1}^m a_{ij} = N \quad \text{mit } a_{ij} \geq 0$$

$\underline{1}'_i$ und $\underline{1}_j$ sind Zeilenvektoren mit n bzw. m Komponenten, die alle Eins sind.

$\underline{P} = (p_{ij}) = 1/N \underline{A}$ ist die Korrespondenzmatrix. In dieser Matrix stehen die relativen Zellenhäufigkeiten von \underline{A} bezogen auf N.

$$\text{Es gilt} \quad \sum_{i=1}^n \sum_{j=1}^m p_{ij} = 1$$

Die relativen Häufigkeiten von Zeilen und Spalten sollen in die Analyse einbezogen werden, um damit eine Gewichtung der Variablen entsprechend ihrer Ausprägungshäufigkeit vorzunehmen. In dem Vektor \underline{z} sollen die relativen Häufigkeiten der Spalten der Ausgangsmatrix \underline{A} stehen, wir bezeichnen ihn als Schwerpunkt bei der Zeilendarstellung. Für die Berechnung von \underline{z} werden die Elemente von \underline{P} spaltenweise addiert.

$$(1) \quad \underline{z} = \underline{1}'_i \underline{P}$$

Das, was für die relativen Häufigkeiten der Spalten gilt, gilt auch für die relativen Häufigkeiten der Zeilen. Den resultierenden n-elementigen Spaltenvektor nennen wir \underline{s} . Bei der Zeilendarstellung entsprechen die Komponenten von \underline{s} den Massen (den relativen Häufigkeiten) der Zeilen.

$$(2) \quad \underline{s} = \underline{P} \underline{1}_j$$

Durch Diagonalisierung von \underline{z} und \underline{s} ergeben sich:

³⁵ Carl Eckart und Gale Young, "The Approximation of one Matrix by another of Lower Rang," in: Psychometrika 1 (1936), S.211-218.

$$D_z = \begin{pmatrix} z_1 & & 0 \\ & z_j & \\ 0 & & z_m \end{pmatrix} \quad D_s = \begin{pmatrix} s_1 & & 0 \\ & s_i & \\ 0 & & s_n \end{pmatrix}$$

Wie dargestellt, werden sowohl Zeilen- als auch Spaltenvektoren in einem niederdimensionalen Raum projiziert.

Wir beginnen mit der Zeilendarstellung. Hierbei gewichten wir die Korrespondenzmatrix (\underline{P}) im ersten Schritt mit den relativen Häufigkeiten der Spalten, also mit dem Vektor \underline{z} . Als resultierende erhalten wir die Matrix der Zeilenprofile \underline{Z} . Bei der Spaltendarstellung wird die Korrespondenzmatrix mit den relativen Häufigkeiten der Zeilen gewichtet. Als Resultat erhalten wir die Matrix der Spaltenprofile \underline{S} .

$$(3) \quad \underline{Z} = \underline{D}_S^{-1} \underline{P}$$

$$(4) \quad \underline{S}' = \underline{D}_Z^{-1} \underline{P}'$$

Geometrische Überlegungen

Bei der Darstellung der Zeilen summieren sich die Komponenten jedes Zeilenvektors zu Eins. Die n -Punkte innerhalb des m -dimensionalen Raumes spannen einen höchsten $(m-1)$ -dimensionalen Raum auf, der orthogonal zum Vektor $\underline{1}_j$ ist. Für diesen Raum gilt die euklidische Metrik nicht. Der quadrierte Abstand zwischen zwei Punkten innerhalb dieses Raumes ist definiert als:

$$(5) \quad d^2(\underline{z}_i, \underline{z}_k) = (\underline{z}_i - \underline{z}_k)' \underline{D}_Z^{-1} (\underline{z}_i - \underline{z}_k)$$

Der Ursprung des Raumes wird als Schwerpunkt gewählt und ein orthogonales Koordinatensystem so eingeführt, daß die Summen der quadrierten Abstände der \underline{z}_i von den Achsen sukzessive minimal sind. Zur Lösung dieses Problems wird durch Umdefinition der Achseneinheiten die Gültigkeit der euklidischen Metrik hergestellt.

$$(6) \quad \underline{z}_i^* = \underline{D}_Z^{-1/2} \underline{z}_i$$

Die Transformation des Schwerpunktes in den neuen Raum erfolgt analog den Punkten.

$$(7) \quad \underline{z}^* = \underline{D}_Z^{-1/2} \underline{z}$$

Der neuen Metrik entsprechend ist:

$$(8) \quad d^{2*}(\underline{z}_i^*, \underline{z}_k^*) = (\underline{z}_i^* - \underline{z}_k^*)' (\underline{z}_i^* - \underline{z}_k^*) = \\ (\underline{z}_i - \underline{z}_k)' \underline{D}_z^{-1} (\underline{z}_i - \underline{z}_k) = d^2(\underline{z}_i, \underline{z}_k)$$

Berechnung der Achsen

Im folgenden sollen auch die relativen Häufigkeiten der Zeilen (s_i) berücksichtigt werden. Die Bestimmung der Achseneinheitsvektoren \underline{e} erfolgt im modifizierten Raum durch sukzessive Maximierung von:

$$(9) \quad \underline{e}' \left[\sum_{i=1}^n s_i (\underline{z}_i^* - \underline{z}^*) (\underline{z}_i^* - \underline{z}^*)' \right] \underline{e}$$

Die Bestimmung von Eigenwerten und Eigenvektoren geschieht durch:

$$(10) \quad \underline{U} = \sum_{i=1}^n s_i (\underline{z}_i^* - \underline{z}^*) (\underline{z}_i^* - \underline{z}^*)' = \\ \sum_{i=1}^n s_i \underline{D}_z^{-1/2} (\underline{z}_i - \underline{z}) (\underline{z}_i - \underline{z})' \underline{D}_z^{-1/2} = \\ \underline{D}_z^{-1/2} \left[\sum_{i=1}^n s_i (\underline{z}_i - \underline{z}) (\underline{z}_i - \underline{z})' \right] \underline{D}_z^{-1/2}$$

Aus dem letzten Teil von Gleichung (10) soll die "Summation" entfallen, wir betrachten statt Vektoren nun Matrixen.

$$(11) \quad \underline{U} = \underline{D}_z^{-1/2} (\underline{Z}' - \underline{z} \underline{1}_i') \underline{D}_s (\underline{Z} - \underline{1}_i \underline{z}') \underline{D}_z^{-1/2} = \\ (\underline{m} \times \underline{m}) \quad \underline{D}_z^{-1/2} (\underline{Z} - \underline{1}_i \underline{z}')' \underline{D}_s (\underline{Z} - \underline{1}_i \underline{z}') \underline{D}_z^{-1/2} \\ (\underline{m} \times \underline{m}) \quad (\underline{m} \times \underline{n}) \quad (\underline{n} \times \underline{n}) \quad (\underline{n} \times \underline{m}) \quad (\underline{m} \times \underline{m})$$

Bei der ($\underline{m} \times \underline{m}$) Matrix \underline{U} handelt es sich um eine symmetrische Matrix, deren Spur gleich der Summe der Eigenwerte ist. Diese Summe wurde "Gesamtträgheit" bezeichnet. Der Wert dieser Gesamtträgheit ist identisch mit dem durch N (total N) dividierten Chi-Quadrat-Wert, also der Summe der Werte der Matrix CELLIN.

Aufgrund der Analogien von Zeilen- und Spaltendarstellung läßt sich die ($\underline{n} \times \underline{n}$)-Matrix \underline{V} darstellen als:

$$(12) \quad \underline{V} = \underline{D}_S^{-1/2} (\underline{S} - \underline{1}_j \underline{s}')' \underline{D}_Z (\underline{S} - \underline{1}_j \underline{s}') \underline{D}_S^{-1/2}$$

Die Spur von \underline{V} ist ebenfalls gleich der Summe der Eigenwerte und entspricht ebenfalls der "Gesamttr agheit".

Nun soll \underline{V} derart umgeformt werden, da  die Verbindung zu \underline{U} ersichtlich wird. Es galt nach den Gleichungen (3) und (4):

$$\underline{Z} = \underline{D}_S^{-1} \underline{P} \quad \text{und} \quad \underline{S} = \underline{D}_Z^{-1} \underline{P}'$$

Aus (3) folgt:

$$(13) \quad \underline{P} = \underline{D}_S \underline{Z}$$

$$(14) \quad \underline{P}' = \underline{Z}' \underline{D}_S$$

$$(15) \quad \underline{S} = \underline{D}_Z^{-1} \underline{Z}' \underline{D}_S$$

Dies setzen wir in Gleichung (12) ein und erhalten:

$$(16) \quad \underline{V} = \underline{D}_S^{-1/2} (\underline{D}_Z^{-1} \underline{Z}' \underline{D}_S - \underline{1}_j \underline{s}')' \underline{D}_Z (\underline{D}_Z^{-1} \underline{Z}' \underline{D}_S - \underline{1}_j \underline{s}') \underline{D}_S^{-1/2}$$

Erweitern wir die rechte Seite von Gleichung (16) innerhalb der beiden Klammerausdr ucke:

$$(17) \quad \underline{V} = \underline{D}_S^{-1/2} (\underline{D}_Z^{-1} \underline{Z}' \underline{D}_S - \underline{D}_Z^{-1} \underline{D}_Z \underline{1}_j \underline{s}' \underline{D}_S^{-1} \underline{D}_S)' \quad *$$

$$\underline{D}_Z (\underline{D}_Z^{-1} \underline{Z}' \underline{D}_S - \underline{D}_Z^{-1} \underline{D}_Z \underline{1}_j \underline{s}' \underline{D}_S^{-1} \underline{D}_S) \underline{D}_S^{-1/2}$$

$$(18) \quad \underline{V} = \underline{D}_S^{-1/2} (\underline{D}_S \underline{Z} \underline{D}_Z^{-1} - \underline{D}_S \underline{D}_S^{-1} \underline{s} \underline{1}_j' \underline{D}_Z \underline{D}_Z^{-1}) \quad *$$

$$\underline{D}_Z (\underline{D}_Z^{-1} \underline{Z}' \underline{D}_S - \underline{D}_Z^{-1} \underline{D}_Z \underline{1}_j \underline{s}' \underline{D}_S^{-1} \underline{D}_S) \underline{D}_S^{-1/2}$$

\underline{D}_Z^{-1} und \underline{D}_S k onnen aus den Klammerausdr ucken herausgezogen werden.

$$(19) \quad \underline{V} = \underline{D}_S^{1/2} (\underline{Z} - \underline{D}_S^{-1} \underline{s} \underline{1}_j' \underline{D}_Z) \underline{D}_Z^{-1} (\underline{Z}' - \underline{D}_Z \underline{1}_j \underline{s}' \underline{D}_S^{-1}) \underline{D}_S^{1/2}$$

$\underline{D}_S^{-1} \underline{s}$ ist ein Spaltenvektor, der aus n Einsen besteht. Multiplizieren wir diesen mit einem m-elementigen Zeilenvektor, so erhalten wir eine Matrix mit n Zeilen und m Spalten, die als Elemente nur Einsen hat. Multiplizieren wir diese mit der Diagonalmatrix \underline{D}_Z , so ergibt sich eine (nxn)-Matrix in der

spaltenweise die jeweils identischen Elemente von \underline{D}_z stehen. Somit ist:

$$(20) \quad \underline{D}_s^{-1} \underline{e} \underline{1}'_j \underline{D}_z = \underline{1}_i \underline{z}'$$

$$(21) \quad \underline{V} = \underline{D}_s^{1/2} (\underline{Z} - \underline{1}_i \underline{z}') \underline{D}_z^{-1} (\underline{Z}' - \underline{z} \underline{1}'_i) \underline{D}_s^{1/2}$$

$$(22) \quad \underline{V} = \underline{D}_s^{1/2} (\underline{Z} - \underline{1}_i \underline{z}') \underline{D}_z^{-1} (\underline{Z} - \underline{1}_i \underline{z}')' \underline{D}_s^{1/2}$$

Wird

$$(23) \quad \underline{W} = \underline{D}_s^{1/2} (\underline{Z} - \underline{1}_i \underline{z}') \underline{D}_z^{-1/2} \quad \text{gesetzt, so ist}$$

$$(24) \quad \underline{V} = \underline{W} \underline{W}' \quad \text{und sich ableitend aus (11)}$$

$$\underline{U} = \underline{W}' \underline{W}$$

Aus diesem Ergebnis folgt, daß \underline{U} und \underline{V} die gleichen von Null verschiedenen Eigenwerte haben. Ist $\lambda = 0$ ein Eigenwert von \underline{U} mit dem Eigenvektor \underline{e} , so ist

$$(25) \quad \underline{W}' \underline{W} \underline{e} = \lambda \underline{e} \quad \text{und}$$

$$(26) \quad \underline{W} \underline{W}' (\underline{W} \underline{e}) = \lambda (\underline{W} \underline{e})$$

Ist $\lambda = 0$ ein Eigenwert von $\underline{W}' \underline{W}$, dann ist er es auch von $\underline{W} \underline{W}'$. Ist \underline{e} der zugehörige Eigenvektor von $\underline{W}' \underline{W} = \underline{U}$, dann ist $\underline{W} \underline{e}$ - nach Normierung auf die Länge Eins - der zugehörige Eigenvektor von $\underline{V} = \underline{W} \underline{W}'$. Dieser normierte Eigenvektor $\underline{W} \underline{e}$ soll \underline{f} heißen. Ist \underline{e} bekannt, so kann \underline{f} bestimmt werden.

Wurde \underline{f} wie hier gezeigt berechnet, so sind sowohl die Längen der Achsen als auch deren geometrische Ausrichtung im Raum für Zeilen- und Spalten-darstellung identisch. Diese Gleichheit ist Voraussetzung für die im Textteil gemachten Interpretationen.

Realisierung von Missing-Data-Ersetzungstechniken innerhalb statistischer Programmpakete und ihre Leistungsfähigkeit

von Rainer Schnell

1. Einleitung

In den meisten Datensätzen der empirischen Sozialforschung fehlen einzelne Angaben für Untersuchungsobjekte. Das Ausmaß fehlender Werte (missing data: MD) variiert abhängig vom Untersuchungsgegenstand, der Untersuchungspopulation, der Datenerhebungsmethode, der Datenbereinigung usw. In den meisten Fällen liegt der Prozentsatz der MD zwischen 1% und 10%. Trotz dieser relativ großen Anteile werden in den meisten Analysen die fehlende Werte einfach ignoriert. Santos (1981:22) hat diese Methode als die "Do Nothing Technique" bezeichnet. Um unverzerrte Ergebnisse zu erhalten, muß man bei Anwendung dieser Methode annehmen, daß die Daten zufällig fehlen ("missing at random": MAR).

Obwohl exakte mathematische Definitionen von MAR vorliegen (Rubin 1976, Little/Rubin 1983:216, Anderson/Basilevsky/Hum 1983:416-417), reicht für die meisten praktischen Zwecke die einfache Definition von Glasser (1964:834) aus: Es sei B_{ij} eine $(n \times p)$ -Indikatormatrix, wobei $B_{ij} = 0$ falls x_{ij} fehlt und $B_{ij} = 1$ andernfalls. Falls

$$E(B_{i1}, B_{i2}, \dots, B_{ip}) = E(B_{i1}) E(B_{i2}) \dots E(B_{ip})$$

dann werden die Daten als zufällig fehlend betrachtet. MAR in diesem Sinn ist die stochastische Unabhängigkeit des Fehlens einer Variablen vom Fehlen jeder anderen Variablen im Datensatz.

Falls die MAR-Annahme korrekt ist, dann ist die Beschränkung der Analyse auf diejenigen Fälle, für die alle Variablen gültige Werte besitzen ("listwise deletion") nur ein weiterer Schritt bei der Ziehung einer Zufallsstichprobe. In diesem Fall sind die Schätzungen unverzerrt, lediglich die Mächtigkeit der statistischen Tests ist (auf Grund der kleineren Stichprobe) geringer (vgl. Little/Rubin 1987). Um die MAR-Annahme zu testen stehen eine Reihe von Verfahren zu Verfügung (Cohen/Cohen 1975, Frane 1978, Hill/Dixon 1981, Möntmann/Bollinger/Herrman 1983). Obwohl diese prinzipiell nicht in der Lage sind, jeden möglichen systematischen Prozeß zu entdecken der fehlende Werte hervorbringen kann ("missing data generating

process", vgl. hierzu Rubin 1976), reichen in der Praxis meist die einfachsten Testverfahren schon aus, um jede Hoffnung auf zufällig fehlende Werte zu zerstören (Frane 1978:27). Bemerkenswerterweise existieren nur sehr wenige Ansätze zur Analyse von Datensätzen mit systematisch fehlenden Werten, die dann aber auch immer nur in sehr speziellen Fällen angewendet werden können (vgl. Little/Rubin 1987:218-243). Wenn MD-Probleme in der Praxis überhaupt angegangen werden, dann werden nahezu ausschließlich sehr einfache Schätz- und Ersetzungsverfahren ("imputation methods") verwendet.

2. Missing-Data-Techniken

Da es hier unmöglich ist, einen vollständigen Überblick über die Literatur zu MD-Techniken zu geben ¹, sollen nur die wichtigsten Methoden für metrische Daten ² erwähnt werden.

Grob kann zwischen Methoden zur Schätzung von Parametern auf der Basis von Datensätzen mit fehlenden Werten und Verfahren zur Ersetzung fehlender Werte durch ihre Schätzung ("imputation methods") unterschieden werden.

Die allgemein bekanntesten Verfahren der ersten Gruppe sind die verschiedenen "missing data correlations", z.B. "pairwise", "listwise" or "allvalue". Zu den Parameterschätzverfahren gehören aber auch zwei vergleichsweise neuere Verfahren: Der sogenannte "EM-Algorithmus" und eine spezielle Klasse ökonometrischer Modelle, die "sample selection-" bzw. "Nelson-Heckman"-Modelle (Heckman 1976, Nelson 1977).

Die "sample selection"-Modelle (Berk/Ray 1982; Berk 1983; Little 1983) erlauben unverzerrte Schätzungen von Regressionsparametern trotz systematisch fehlender Werte. Diese Modelle basieren alle auf der Annahme eines fehlende Werte generierenden Prozesses, bei dem ein Wert dann fehlt, wenn eine latente Variable einen unbeobachteten stochastischen Schwellenwert überschreitet. Neben einigen Verteilungsannahmen ist das Hauptproblem ³ dieser Modelle die Spezifikation eines theoretischen Modells des "Missing

¹ Eine knappe Einführung geben Anderson/Basilevsky/Hum (1983). Eine ausführliche Darstellung dieser und anderer, eher selten verwendeter Methoden gibt Schnell (1986:83-137).

² Für Methoden zur Behandlung (auch systematisch) fehlender Werte bei qualitativen Daten vgl. vor allem Fuchs (1982), Fay (1986), Little/Rubin (1987) und Haagenars (1988).

³ Die Berechnung solcher Modelle ist mit speziellen Programmen sehr einfach geworden. Ein solches auch für IBM-PC's verfügbares Programm ist LIMDEP (Greene 1986).

Data generierenden Prozesses". Unglücklicherweise existieren solche Modelle außerhalb der Ökonometrie in den Sozialwissenschaften kaum, so daß nur selten Anwendungen gefunden werden können.

Die neuere MD-Literatur wird durch den "EM-Algorithmus" als Parameterschätzverfahren dominiert. Der "expectation maximization" (EM)-Algorithmus ist eine allgemeine Methode⁴ zur Berechnung von Maximum-Likelihood-Schätzern, die (im Prinzip) für jedes MD-Problem verwendet werden könnte (Orchard/Woodbury 1972; Dempster/Laird/Rubin 1977). Der EM-Algorithmus ist eine Iterationsmethode, bei der jede Iteration aus einem E-Schritt, in dem die Log-Likelihood-Funktion berechnet wird, und einem M-Schritt, in dem diejenigen Parameter berechnet werden, die die Log-Likelihood-Funktion maximieren, besteht (vgl. Little 1983:369-371). In diesem Zusammenhang kann der EM-Algorithmus folgendermaßen dargestellt werden:

1. Berechnung eines Ausgangsmittelwertvektors und einer Ausgangskovarianzmatrix (listwise);
2. Für jeden Fall mit MD: Partitionieren des Mittelwertvektors und der Kovarianzmatrix in Matrizen für die vollständigen und unvollständigen Variablen;
3. Schätzung der MD durch multiple Regression, wobei alle vorhandenen Variablen dieses Falles für die Schätzung verwendet werden.
4. Berechnung eines neuen Mittelwertvektors und einer neuen Kovarianzmatrix;
5. Korrektur der Kovarianzmatrix: Für jeden Fall mit MD wird die residuale Kovarianz der Prädiktorvariablen zum entsprechenden Element der Kovarianzmatrix addiert;
6. Berechnung eines Konvergenzkriteriums; falls keine Konvergenz: Wiederholung der Schritte 2-6; sonst Stopp.

Da diese Form des EM-Algorithmus also auch die fehlenden Werte ersetzt, gehört es zu beiden Gruppen von Verfahren. Eine Implementation dieser Methode (mit der Ausnahme von ALLVALUE in Schritt 1) findet sich in BMDPAM mit der "ML"-Option im "ESTIMATE TYPE="-Kommando.

Natürlich ist es möglich fehlende Werte direkt mit anderen Methoden zu schätzen. Die Ersetzung fehlender Werte durch Schätzungen wird "imputation" genannt. Imputation besitzt zwei Vorteile:

1. Nach der einmaligen Ersetzung liegt ein vollständiger "Allzweck"-Datensatz vor, der mit traditionellen Methoden (und Programmen) für vollständige Daten analysiert werden kann;

⁴ Die Bezeichnung als 'Algorithmus' (durch Dempster/Laird/Rubin 1977) ist daher äußerst irreführend.

2. Durch die Imputation kann möglicherweise die Verzerrung durch systematisch fehlende Werte ausgeglichen werden, vor allem dann, wenn die Ersetzung durch die Produzenten des Datensatzes vorgenommen wird, die ihre möglicherweise vorhandenen zusätzlichen Informationen über den MD generierenden Prozeß bei der Schätzung verwenden können.

Eine nützliche Unterscheidung zwischen "informativen" und "nicht informativen" Imputationsmethoden stammt von Santos (1981). Informative Imputationen wurden von Santos (1981:30) als Methoden definiert, die zusätzliche Informationen bei der Zuweisung von Werten zu unvollständigen Fällen verwenden.

Die einfachste uninformative Methode ist das pure Raten eines fehlenden Wertes, die nächst einfache Methode ist die Ersetzung eines MD mit dem Mittelwert der vorhandenen Werte (im Folgenden: MEAN). Bei der Schätzung von Korrelationen wird diese Methode häufig als "Wilk's Methode" bezeichnet. Eine Reihe von Statistikprogramm Paketen (z.B. SPSS-X) bieten diese Methode als "mean substitution" bei der Schätzung der Kovarianzmatrix für Regressionen an.

Ein sehr einfaches Beispiel einer informativen Ersetzung ist die "Zellenmittelwertersetzung" (im Folgenden: CELL). Hier werden die MD durch den Mittelwert einer Subgruppe ersetzt, zu der der unvollständige Fall gehört. Das Problem dieser und anderer subgruppenbasierter Techniken liegt in der Bestimmung der Subgruppen, die in diesem Zusammenhang als "Imputationsklassen" bezeichnet werden. Imputationsklassen sollten nur sehr geringe Binnengruppenvarianzen besitzen, aber soviel Zwischengruppenvarianz wie möglich. Weiterhin sollte die Zahl der Klassen gering und die Klassifikationsvariablen weitgehend unabhängig voneinander sein (Rizvi 1983:306). Damit eine Variable überhaupt als Klassifikationsvariable verwendet werden kann, muß der Anteil fehlender Werte für diese Variable sehr klein sein. Häufig wird eine Form der bekannten "Automatic Interaction Detector (AID)"-Technik⁵ (oder eine andere Clusteranalysetechnik⁶) verwendet um die Klassifikationsvariablen zu bestimmen⁷.

⁵ Die klassische Arbeit ist Sonquist/Baker/Morgan (1971). Dillon/Goldstein (1984:183-186) geben eine kurze Einführung. Die beste technische Dokumentation von AID findet sich bei Hartigan (1975). Eine allgemeine Kritik von AID gibt Fielding (1979).

⁶ Eine ausführliche Behandlung von MD-Problemen bei Clusteranalysen findet sich in zwei exzellenten Arbeiten von Wishart (1978, 1985).

⁷ Vgl. Rizvi (1983). Für AID-Techniken in diesem Zusammenhang, vgl. Kalton/Santos (1983:84-87, 93-96) und Chapman (1983). Vgl. ferner Oh/Scheuren (1983:157-158). Für eine Kritik der theoretischen Annahmen, vgl. Schnell (1986:104-106).

Eine etwas interessantere Menge von Techniken bilden die sogenannten "Hot-Deck"-Methoden, die vor allem in der amtlichen Statistik, insbesondere bei Volkszählungen, verwendet werden. Obwohl es keine einheitliche Definition von Hot-Decks gibt (Ford 1983:185-186), ist allen Definitionen die Erwähnung des stochastischen Charakters, die Verwendung von Daten des unvollständigen Datensatzes und des "Dopplungscharakters" des Verfahrens gemein. Dies wird klarer, wenn wir zwischen zwei Arten von Hot-Deck-Methoden unterscheiden: Sequentiellen und simultanen Hot-Decks.

Sequentielle Hot-Decks (auch als "traditionelle Hot-Decks" bezeichnet, vgl. Kalton/Kasprzyk 1982:23) beginnen mit der Bestimmung der Imputationsklassen und der Startwerte für die fehlenden Variablen jeder Imputationsklasse. Die Startwerte sind häufig die Zellenmittelwerte der Variablen in früheren Untersuchungen. Der unvollständige Datensatz wird dann sequentiell abgearbeitet. Für jeden Fall des unvollständigen Datensatzes wird seine Imputationsklasse bestimmt. Trifft das Programm auf einen gültigen Wert des aktuellen Falles, so wird dieser Wert in die entsprechende Imputationsklasse kopiert (der Fall wird ein "Donor"). Trifft das Programm auf einen fehlenden Wert, so wird dieser durch den derzeitigen Inhalt der Imputationsklasse ersetzt (der Fall wird ein "Rezipient"). Das Verfahren wird bis zum Ende des Datensatzes fortgesetzt, anschließend besitzt jeder Fall für die vervollständigte Variable einen gültigen Wert. Falls mehr als eine Variable ersetzt wird, ist es möglich, daß bestehende Editierungsregeln verletzt werden, wenn die Daten eines bestimmten Falles von verschiedenen "Donors" stammen: So z.B. wenn einer der "Donors" weiblich und ein anderer "Donor" männlich ist und das Geschlecht des zweiten für die Imputation verwendet wird, aber die Anzahl der Schwangerschaften des ersten. In tatsächlichen Hot-Deck-Systemen sind die Editierungsregeln weit komplizierter als die des Beispiels. Die einfachste Lösung dieses Problems besteht darin, zunächst den vollständigen Datensatz gemäß den Editregeln zu bereinigen und dann den bereinigten vollständigen Datensatz so zu verwenden, daß die Daten jedes unvollständigen Falles für alle durch mögliche Editregeln betroffenen Variablen durch einen Vektor von Variablen eines vollständigen Falles ersetzt werden. Dabei werden auch vorhandene gültige Werte durch Werte des Donors ersetzt. Sequentielle Hot-Decks werfen eine Reihe weiterer Probleme auf, so z.B. die Bestimmung der Imputationsklassen, ungleichmäßige Benutzung von "Donoren" (d.h. manche Fälle werden extrem häufig zur Imputation herangezogen), und Abfolgeeffekte⁸.

⁸ Detaillierte und praxisbezogene Darstellungen dieser Probleme geben Sande (1982) und Ford (1983).

Die seltener verwendeten ⁹ simultanen Hot-Decks verwenden sogenannte "matching variables" (entsprechend den Klassifikationsvariablen), wobei aber jeder unvollständige Fall durch die Daten desjenigen vollständigen Falles vervollständigt wird, zu dem er die kleinste Distanz auf den Matchvariablen aufweist. Zumeist wird hierbei die euklidische Distanz

$$d_{i,j} = \left[\sum_{k=1}^p (x_{i,k} - x_{j,k})^2 \right]^{1/2}$$

verwendet, gelegentlich auch die Mahalanobis-Distanz:

$$d_{i,j} = (X_i - X_j)' S^{-1} (X_i - X_j),$$

wobei X_i ein $(1 \times p)$ -Vektor der Variablen des Falles i , X_j ein $(1 \times p)$ -Vektor der Variablen des Falles j , und S^{-1} die Inverse der $(p \times p)$ -Kovarianzmatrix der Variablen ist.

Sande (1979:248) betrachtet die Wahl der Distanzfunktion als unkritisch. Methoden mit solchen Ersetzungsregeln werden gelegentlich als "nearest-neighbor"-Methoden ¹⁰ bezeichnet.

Selbstverständlich lassen sich viele multivariate Verfahren zur Schätzung fehlender Werte verwenden. Eine der frühesten Methoden stammt von Dear (1959). Dear berechnet die erste Hauptkomponente einer über Mittelwerte vervollständigten Datenmatrix. Der zum größten Eigenwert der Korrelationsmatrix gehörende normalisierte Eigenvektor g wird zusammen mit der z -transformierten Datenmatrix ($z_{i,j} = 0$, falls $x_{i,j}$ fehlt) zur Schätzung verwendet. Ein fehlender Wert $z_{i,j}$ wird geschätzt durch:

$$z_{i,j} = g_j \sum_{k=1}^p z_{i,k} g_k$$

und dann wieder in die ursprünglichen Einheiten transformiert ¹¹.

⁹ Selbst mit spezieller Software sind diese Techniken weit aufwendiger als traditionelle Hot-Decks. Bei großen Datensätzen (Mikrozensus, Volkszählungen etc.) dürften solche Verfahren unbezahlbar werden.

¹⁰ Einzelheiten zu simultanen Hot-Decks finden sich bei Vacek/Ashikaga (1980) and Sande (1979).

¹¹ Diese Beschreibung basiert auf Hamilton (1975:27). Die Originalarbeit war mir nicht zugänglich.

Offensichtlich ist es möglich fehlende Werte durch Regressionsverfahren zu schätzen. Dies wurde schon 1960 von Buck vorgeschlagen. Buck's Methode beginnt mit der $(p \times p)$ Korrelationsmatrix der vollständigen Fälle. Für jeden Fall i und jede unvollständige Variable j ($j=1,2,\dots,m$) wird der fehlende Wert durch eine multiple Regression mit allen vollständigen $(p-m_j)$ Variablen des Falls i als unabhängige Variablen geschätzt. Falls jede Kombination fehlender Variablen möglich ist und k Variablen fehlen, müssen maximal

$$p \binom{p-1}{k-1}$$

verschiedene Regressionsgleichungen berechnet werden (Buck 1960:303). Da jeder geschätzte fehlende Wert exakt auf der Regressionsgeraden liegt, wäre die resultierende Kovarianzmatrix des vervollständigten Datensatzes verzerrt. Buck (1960:304) schlägt daher eine Korrektur der Kovarianzmatrix (A) des vervollständigten Datensatzes vor. Falls die Variable j für einen Anteil I_j des Samples fehlt, wird die korrigierte Varianz a_{jj} berechnet als

$$a_{jj} = a_{jj} + I_j/c_{jj}$$

wobei c_{jj} das Diagonalelement von A^{-1} ist. Diese Korrektur wird nicht immer in der Literatur erwähnt und jede Art von Regressionsmethode als "Buck's method" bezeichnet.

Obwohl eine große Zahl verschiedener Abwandlungen von Regressions-techniken vorgeschlagen wurde (vgl. Schnell 1986:117-119) werden diese aber nur selten verwendet. Lediglich iterierte Formen von multiplen Regressionen sind üblich. Eine iterierte multiple Regressionsmethode zur Schätzung fehlender Werte (ohne Korrektur der Kovarianzmatrix) wurde 1959 von Federspiel, Monroe & Greenberg vorgeschlagen. Versionen solcher Subroutinen finden sich z.B. in den speziellen MD-Programmen von Berger (1979) und O'Grady (1982). Schmee/Hahn (1979) schlugen eine "iterated least square"-Technik für die Lösung von Censoring-Problemen vor. Interessanterweise ist eine iterative Buck-Methode (mit Korrektur) mit dem oben erwähnten EM-Algorithmus identisch, wie Beale/Little (1975:134-137) gezeigt haben.

3. Realisierung mit Standardpaketen

Nur wenige der erwähnten Techniken sind als Prozeduren in den Standardstatistikprogramm Paketen wie SPSS-X oder SAS enthalten (im Gegensatz zu BMDP, welches mit dem Programm AM (Autor: James Frane) die

meisten MD-Ersetzungstechniken, z.B. MEAN, CELL, multiple Regression und iterierte multiple Regression zur Verfügung stellt). Trotzdem ist es möglich, die meisten Techniken innerhalb der Standardpakete mit deren Daten-transformations- und Fallselektionsmöglichkeiten zu realisieren.

Im Folgenden sollen einige Möglichkeiten gezeigt werden, wie mit einem einfachen derzeitigen Statistiksystem (SPSS-PC+) Ersetzungen vorgenommen werden können. Mit nur kleinen Änderungen funktionieren diese Methoden auch mit SPSS-9 und SPSS-X. Andere Pakete (z.B. SAS, BMDP oder SYSTAT) erfordern etwas größere Änderungen, aber die Logik der Prozeduren bleibt erhalten. Selbstverständlich können auch die kompliziertesten Techniken mit Hochsprachen wie z.B. PASCAL leicht realisiert werden. Für GAUSS liegt das Programm MISS vor, das viele der erwähnten Techniken zur Verfügung stellt.

Die Beispiele basieren auf einem kleinen Datensatz (MISSING.SYS) mit 5 Variablen (V1 to V5) mit fehlenden Werten. Der Datensatz enthält darüber hinaus drei zusätzliche Variablen AUX1 bis AUX3 ohne fehlende Werte. Die letzte Variable im Datensatz ist eine vollständige Gruppierungsvariable GROUP. Es muß angenommen werden, daß zumindest ein Fall innerhalb jeder Gruppe (GROUP) vollständige Daten für alle Variablen besitzt.

Eine der einfachsten Imputationstechniken ist die Ersetzung durch den Mittelwert. Diese kann z.B. durch den folgenden SPSS-PC-Job erfolgen:

```
GET FILE='MISSING.SYS' .
COMPUTE DUMMY=1.
SAVE OUTFILE='SCRATCH' .
AGGREGATE OUTFILE= *
/BREAK=DUMMY
/M1 TO M5 = MEAN (V1 TO V5).
JOIN MATCH /TABLE=* /FILE='SCRATCH'/BY DUMMY.
IF (MISSING(V1)) V1=M1.
IF (MISSING(V2)) V2=M2.
IF (MISSING(V3)) V3=M3.
IF (MISSING(V4)) V4=M4.
IF (MISSING(V5)) V5=M5.
```

Die Variable DUMMY wird lediglich durch die Syntax des AGGREGATE-Kommandos erzwungen.

Eine Ersetzung durch Zellenmittelwerte (alle fehlenden Werte werden durch den Mittelwert einer Variablen in einer Zelle, zu der der Fall gehört, ersetzt) mit der Variablen GROUP als Gruppierungsvariable benötigt nur eine kleine Modifikation dieses Jobs:

```

GET FILE='MISSING.SYS'.
COMPUTE SEQUENCE=$CASENUM.
SORT CASES BY GROUP.
SAVE OUTFILE='SCRATCH'.
AGGREGATE OUTFILE= *
/PRESORTED
/BREAK=GROUP
/M1 TO M5 = MEAN (V1 TO V5).
JOIN MATCH /TABLE=* /FILE='SCRATCH'/BY GROUP.
IF (MISSING(V1)) V1=M1.
IF (MISSING(V2)) V2=M2.
IF (MISSING(V3)) V3=M3.
IF (MISSING(V4)) V4=M4.
IF (MISSING(V5)) V5=M5.
SORT CASES BY SEQUENCE.

```

Die erste Sortierung ist notwendig für das JOIN MATCH-Kommando, die letzte Sortierung stellt die ursprüngliche Abfolge der Fälle im Datensatz wieder her.

Hot-Deck-Verfahren können in vielen Varianten realisiert werden. Ein einfache Methode für eine variablenweise Ersetzung ist der folgende Job:

```

GET FILE='MISSING.SYS'.
COMPUTE SEQUENCE=$CASENUM.
COUNT NMISS=V1 TO V5 (MISSING).
SORT CASES BY GROUP NMISS.
IF (MISSING(V1)) V1=LAG(V1).
IF (MISSING(V2)) V2=LAG(V2).
IF (MISSING(V3)) V3=LAG(V3).
IF (MISSING(V4)) V4=LAG(V4).
IF (MISSING(V5)) V5=LAG(V5).
SORT CASES BY SEQUENCE.

```

Die Berechnung von NMISS und das Sortieren nach GROUP und NMISS garantieren, daß der erste Fall jeder Gruppe für jede Variable gültige Werte besitzt (falls die oben genannte Voraussetzung über die Struktur des Datensatzes erfüllt ist).

Diese Art des Hot-Decks kann bestehende Edit-Regeln verletzen, falls die zugeschrieben Werte von verschiedenen "Donors" stammen. Um dies zu vermeiden, führt eine leichte Änderung zu einer fallweisen Ersetzung der fehlenden Werte:

```

GET FILE='MISSING.SYS'.
COMPUTE SEQUENCE=$CASENUM.
COUNT NMISS=V1 TO V5 (MISSING).
SORT CASES BY GROUP NMISS.
IF (NMISS NE 0) V1=LAG(V1).
IF (NMISS NE 0) V2=LAG(V2).
IF (NMISS NE 0) V3=LAG(V3).
IF (NMISS NE 0) V4=LAG(V4).
IF (NMISS NE 0) V5=LAG(V5).
SORT CASES BY SEQUENCE.

```

Obwohl die Resultate zweier Simulationen (Kaiser 1983, Schnell 1986; abweichende Ergebnisse werden von Vacek/Ashikaga 1980 berichtet) darauf hinweisen, daß Verfahren mit "nearest-neighbor rule" zur Ersetzung zu stark verzerrten Kovarianzschätzungen führen, soll eine Version dieser Technik hier illustriert werden. Da keine SPSS-Version die Verwendung eines berechneten Resultates als Parameter eines Kommandos erlaubt, muß dies in zwei Schritten geschehen. Der erste Job berechnet die Z-Werte der zusätzlichen Variablen AUX1 bis AUX3. Diese Z-Werte werden nur für die QUICK CLUSTER-Prozedur verwendet. Weiterhin sortiert der erste Job die Fälle für eine Hot-Deck-ähnliche Ersetzung. Schließlich berechnet der Job noch die Anzahl der vollständigen Fälle auf den Variablen V1 bis V5 über die (Listwise-) Berechnung der Korrelationen:

```

* JOB1.
GET FILE='MISSING2.SYS'.
COUNT NMISS= V1 TO V5 (MISSING).
COMPUTE SEQUENCE=$CASENUM.
SORT CASES BY NMISS.
DESCRIPTIVES AUX1 AUX2 AUX3 /OPTIONS 3.
CORRELATION V1 TO V5.
SAVE OUTFILE='SCRATCH'.

```

Die Anzahl vollständiger Beobachtungen auf den Variablen V1 to V5 ("6") muß manuell in das QUICK CLUSTER Kommando im zweiten Job eingefügt werden.

```

* JOB2.
GET FILE='SCRATCH'.
QUICK CLUSTER ZAUX1 TO ZAUX3
  /CRITERIA=CLUSTERS(6) NOUPDATE
  /INITIAL=FIRST
  /SAVE CLUSTER(CLUS).
SORT CASES BY CLUS NMISS.

```

```

IF (NMISS NE 0) V1=LAG(V1).
IF (NMISS NE 0) V2=LAG(V2).
IF (NMISS NE 0) V3=LAG(V3).
IF (NMISS NE 0) V4=LAG(V4).
IF (NMISS NE 0) V5=LAG(V5).
SORT CASES BY SEQUENCE.

```

QUICK CLUSTER sucht den nächsten (euklidische Distanz) Nachbarn auf den Variablen ZAUX1 bis ZAUX3. Die Clusterzugehörigkeit (CLUS) wird für jeden Fall gespeichert. Die Distanz jedes vollständigen Falles zum nächsten vollständigen Fall ist null. Das Sortieren nach CLUS und NMISS garantiert, daß ein vollständiger Fall zu Beginn des Datenfiles liegt. Fehlende Werte werden durch die Daten des nächsten (im Sinne der euklidischen Distanz) vollständigen Falles ersetzt. In diesem Job wurde ein variablenweises Hot-Deck verwendet, es kann aber durch die oben beschriebene fallweise Imputation ersetzt werden. Für einen Datensatz mit 1800 vollständigen Fällen benötigt ein 20Mhz-80386/10MHz-80287-System für diesen Job mehr als 70 Minuten.

Eine Ersetzungsmethode, die der von Dear (1959) vorgeschlagenen Verwendung der ersten Hauptkomponente sehr ähnelt, kann ebenso mit SPSS-PC+ realisiert werden. Da hierfür das FACTOR /SAVE Kommando verwendet werden muß, müssen die fehlenden Werte durch Mittelwerte ersetzt werden. Um die resultierenden Factorscores (die mit der Anderson/Rubin-Methode berechnet werden) wieder auf die ursprünglichen Einheiten zu bringen, werden die Mittelwerte und Standardabweichungen der Variablen benötigt. Diese können mit dem AGGREGATE-Kommando gespeichert werden. Der letzte Schritt besteht aus der Ersetzung durch die geschätzten Werte:

```

* DEAR-LIKE-SUBSTITUTION.
GET FILE='MISSING.SYS'.
* 1. REPLACE MD WITH MEANS (NECESSARY FOR
* FACTOR-SAVE COMMAND.
COMPUTE DUMMY=1.
SAVE OUTFILE='SCRATCH'.
AGGREGATE OUTFILE= *
/BREAK=DUMMY
/M1 TO M5 = MEAN (V1 TO V5)
/S1 TO S5 = SD (V1 TO V5).
* SAVE STANDARD DEVIATIONS FOR RESCALING OF 'Z-SCORES'.
JOIN MATCH /TABLE=* /FILE='SCRATCH'/BY DUMMY.

```

```

COMPUTE T1=V1.
COMPUTE T2=V2.
COMPUTE T3=V3.
COMPUTE T4=V4.
COMPUTE T5=V5.
IF (MISSING(V1)) T1=M1.
IF (MISSING(V2)) T2=M2.
IF (MISSING(V3)) T3=M3.
IF (MISSING(V4)) T4=M4.
IF (MISSING(V5)) T5=M5.
* 2. COMPUTE 1. PRINCIPAL COMPONENT.
FACTOR VARIABLES=T1 TO T5
  /ANALYSIS=ALL
  /CRITERIA=FACTORS(1)
  /EXTRACTION=PC
  /ROTATION=NOROTATE
  /SAVE=AR (1 DEAR).
* 3. RESCALE ESTIMATES.
COMPUTE T1=DEAR1*S1+M1.
COMPUTE T2=DEAR1*S2+M2.
COMPUTE T3=DEAR1*S3+M3.
COMPUTE T4=DEAR1*S4+M4.
COMPUTE T5=DEAR1*S5+M5.
* 4. REPLACE MISSING VALUES WITH ESTIMATES.
IF (MISSING(V1)) V1=T1.
IF (MISSING(V2)) V2=T2.
IF (MISSING(V3)) V3=T3.
IF (MISSING(V4)) V4=T4.
IF (MISSING(V5)) V5=T5.

```

Eine multiple Regressionsimputation ist mit SPSS-PC+ sehr einfach. Der einzige kleine Trick des Jobs besteht in der Beschränkung des Samples auf die vollständigen Fällen mit dem SELECT-Subkommando der Regressionsprozedur. Zusammen mit dem /MISSING= MEANSUBSTITUTION Subkommando zwingt dies SPSS-PC+ die Korrelationsmatrix "listwise" zu berechnen und die über Mittelwerte vervollständigten Daten für die Schätzung der vorhergesagten Werte zu verwenden:

```

GET FILE='MISSING.SYS'.
COUNT NMISS=V1 TO V5(MISSING).
REGRESSION VARIABLES=V1 TO V5
  /MISSING=MEANSUBSTITUTION
  /SELECT= NMISS EQ 0

```

```

/DEPENDENT=V1/ENTER/SAVE PRED(M1)
/DEPENDENT=V2/ENTER/SAVE PRED(M2)
/DEPENDENT=V3/ENTER/SAVE PRED(M3)
/DEPENDENT=V4/ENTER/SAVE PRED(M4)
/DEPENDENT=V5/ENTER/SAVE PRED(M5).
IF (MISSING(V1)) V1=M1.
IF (MISSING(V2)) V2=M2.
IF (MISSING(V3)) V3=M3.
IF (MISSING(V4)) V4=M4.
IF (MISSING(V5)) V5=M5.

```

Eine iterierte multiple Regressionsimputationsmethode, die dem EM-Algorithmus für dieses Problem entsprechen würde (mit der Ausnahme der zwar notwendigen, aber vernachlässigten Korrektur der Kovarianzmatrix) ist innerhalb von SPSS-PC+ ohne großen manuellen Aufwand nicht möglich. Dies würde getrennte multiple Regressionen für jede Menge von unvollständigen Beobachtungen mit einem anderen Muster der fehlenden Werte erfordern¹².

4. Die Leistungsfähigkeit der MD-Techniken

Selbstverständlich gilt das Interesse zunächst der Leistungsfähigkeit der MD-Techniken. Ihre Leistung hängt von sehr vielen Parametern ab, so z.B. dem fehlende Werte erzeugenden Mechanismus, dem Anteil fehlender Werte, der Anzahl der Variablen, der Anzahl der Fälle, dem Ausmaß der

¹² Eine einfache, aber sehr informative Analyse der fehlenden Werte ermöglicht ein "MD-pattern plot". Solch ein Plot ist in BMDPAM verfügbar. In SPSS-PC läßt sich dies so erreichen:

```

GET FILE='MISSING.SYS'.
COMPUTE PATTERN=0.
IF (MISSING(V1)) PATTERN=PATTERN+10000.
IF (MISSING(V2)) PATTERN=PATTERN+1000.
IF (MISSING(V3)) PATTERN=PATTERN+100.
IF (MISSING(V4)) PATTERN=PATTERN+10.
IF (MISSING(V5)) PATTERN=PATTERN+1.
RECODE V1 TO V5 (LOWEST THRU HIGHEST=1).
AGGREGATE OUTFILE=*
  /BREAK=PATTERN
  /NOFMDP=NU
  /M1 TO M5=LAST(V1 TO V5).
LIST PATTERN NOFMDP M1 TO M5.

```

Multikollinearität der Variablen, der Verteilungsform der Variablen und nicht zuletzt von den verwendeten Kriterien zur Beurteilung der Verfahren. Daher können nur für sehr einfache Methoden und spezielle Situationen (z.B. exakt ein fehlender Wert pro Variable) und nur unter der Annahme zufällig fehlender Werte analytische Aussagen über die Leistungsfähigkeit der Verfahren gemacht werden. Für den Fall systematisch fehlender Werte liegen keinerlei analytische Ergebnisse vor¹³. Der einzige Weg zur Beurteilung numerischer Verfahren, wenn keine analytischen Aussagen möglich sind, führt über Monte-Carlo-Simulationsstudien. Die MD-Literatur enthält zwar eine große Zahl von Simulationsstudien, von diesen werden aber nur ungefähr 10 häufiger zitiert¹⁴. Da es keinerlei Standardisierung in diesem Forschungsfeld gibt, können die meisten Studien nicht direkt verglichen werden. So verwenden die älteren Arbeiten nur selten ein experimentelles Design, die Ergebnisse können daher nicht auf bestimmte Faktorkombinationen zurückgeführt werden. Weiterhin verwenden die Arbeiten verschiedene Schätzmethoden und unterschiedliche Kriterien der Leistungsfähigkeit¹⁵. Trotzdem existiert in der Literatur weitgehend Konsens in bezug auf die prinzipielle Leistungsfähigkeit der Verfahren¹⁶. Die einfachen Ersetzungsverfahren (MEAN, CELL und Hot-Decks) werden in bezug auf die Schätzung von Kovarianzmatrizen als nicht so leistungsfähig betrachtet wie die multivariaten Verfahren. DEAR scheint nur bei hoher Multikollinearität brauchbar zu sein. Iterierte multiple Regressionen (insbesondere der EM-Algorithmus) werden als die

¹³ Allerdings haben Santos (1981) für Regressionparameter und Brown (1983) für die Schätzung von Faktorladungen erste analytische Versuche unternommen.

¹⁴ Die häufigst zitierten Arbeiten stammen von Haitovsky (1969), Timm (1970), Gleason/Staelin (1975), Beale/Little (1975), Kim/Curry (1977) und Little (1979). Die neueren Arbeiten von Vanguilder/Azen (1981), Kaiser (1984) und Basilevsky/Sabourin/Hum/Anderson (1985) sind wesentlich sorgfältiger entworfen, mit nur wenigen Ausnahmen führen sie aber kaum zu anderen Ergebnissen.

¹⁵ Eine interessante Konsequenz dieser Situation ist die Ablehnung der "pairwise"-Korrelationsschätzung, hauptsächlich auf der Basis der Simulationsergebnisse von Haitovsky (1969). Viele Simulationsstudien schlossen die "pairwise"-Schätzung daher schon im Design aus; mathematische Statistiker lehnten die Methode aus theoretischen Gründen ab (z.B. Heiberger 1977). Es ist bemerkenswert, daß in den letzten Jahren der Pairwise-Schätzer sowohl zunehmende theoretische Unterstützung (Brown 1983) und vor allem noch wesentlich bessere Simulationsergebnisse (Finkbeiner 1979, Basilevsky/Sabourin/Hum/Anderson 1985, Schnell 1985) erhalten hat als zuvor.

¹⁶ Eine relativ aktuelle, wenn auch unvollständige Übersicht findet sich bei Anderson/Basilevsky/Hum (1983). Eine lesbare, wenn auch etwas veraltete Übersicht gibt Hamilton (1975). Eine kritische Übersicht über fast alle publizierten Arbeiten bis 1984 gibt Schnell (1986:138-157).

besten Verfahren angesehen. In Hinsicht auf die Schätzung einzelner fehlender Werte weichen die Ergebnisse kaum ab; dieses Problem wird aber in nur wenigen Arbeiten aufgegriffen. Die meisten Arbeiten konzentrieren sich auf die Schätzung von Summen und Mittelwerten.

Die wichtigste Beschränkung der Simulationsstudien liegt in der Verwendung nur eines fehlende Werte erzeugenden Prozesses: Mit der Ausnahme von Van Guilder/Azen (1981) basiert jede Simulationsstudie auf der Annahme zufällig fehlender Werte ("missing at random": MAR). Diese Annahme ist aber bei den meisten praktischen Problemen nicht erfüllt. Falls Empfehlungen auf der Basis der Ergebnisse von Simulationsstudien, die alle die MAR-Annahme verwenden, gegeben werden sollen, muß angenommen werden, daß es zwischen dem Ausfallprozeß und der Leistungsfähigkeit der Verfahren keinen Interaktionseffekt geben darf. Mit anderen Worten: Die Verfahren müssen unter jedem möglichen Ausfallmechanismus gleichmäßig leistungsfähig sein. Um diese kühne Annahme zu testen, wurde eine neue Simulationsstudie entworfen.

5. Entwurf der Simulationsstudie

Die Simulation basiert auf einem vollfaktoriellen Design mit vier Faktoren (Stichprobengröße, Interkorrelation, Prozentsatz fehlender Werte und fehlende Werte erzeugender Prozeß) mit jeweils 3 Stufen für jeden Faktor, so daß $3 \cdot 3 \cdot 3 \cdot 3 = 81$ verschiedene Bedingungen getestet wurden. Für jede experimentelle Bedingung wurden 10 verschiedene Datensätze mit jeweils 10 multivariatnormal verteilten Variablen mit gegebener Korrelationsstruktur erzeugt¹⁷. Jeder der so entstandenen 810 verschiedenen unvollständigen Datensätze wurde mit 10 verschiedenen MD-Techniken vervollständigt, so daß ein $3 \cdot 3 \cdot 3 \cdot 3$ vollfaktorielles Design mit wiederholter (*10) Messung auf dem letzten Faktor entstand.

Die Stichprobengröße wurde auf 100, 300 und 500 Fälle mit jeweils 10 Variablen festgelegt. Zusätzlich zu den 10 Variablen mit fehlenden Werten

¹⁷ Das Simulationsprogramm wurde für den IBM-VS-FORTRAN Compiler (Level 1.3.0) unter CMS geschrieben. Zur Erzeugung der Pseudozufallszahlen und für die statistischen Subroutinen wurden Subroutinen der Numerical Algorithm Group (NAG) Fortran-Library Mark 10 verwendet. Der grundlegende "multiplicative congruential generator" G05CAF wurde mit der Subroutine G05CBF gestartet. Die multivariaten Normalverteilungen wurden mit den Generatoren G05EAF und G05EZF erzeugt. Die Simulation wurde auf der IBM 4341 der Universität Essen gerechnet und beanspruchte ca. 17 Stunden CPU-Zeit. Weitere Details und ein Programmlisting (ca. 700 Zeilen Fortran) finden sich bei Schnell (1986).

wurden drei Variablen ohne fehlende Werte erzeugt. Die Zusatzvariablen (V1 - V3) wurden als Klassifikationsvariablen und als unabhängige Variablen für die Regressionsmethoden verwendet. Die Zusatzvariablen sollten demographische Variablen wie Geschlecht und Alter simulieren, die in der Regel für jeden Fall eines Datensatzes vorliegen. Alle Berechnungen der Leistungsfähigkeit der Verfahren basieren auf den Schätzungen für die 10 unvollständigen Variablen und ihrer 45 Interkorrelationen. Die Struktur der Korrelationsmatrizen war homogen (in der Population): jedes Nicht-Diagonalelement der Korrelationsmatrix V4...V13 in der Population wurde gleich dem Wert des experimentellen Faktor CM (correlation mean) gesetzt. CM nahm die Werte 0,2, 0,4 und 0,6 an. Für die Simulation eines Thresholdmodelles, bei dem Werte dann fehlen, wenn eine ungemessene Variable einen stochastischen Schwellenwert überschreitet, wurde eine zusätzliche Hilfsvariable (V14) erzeugt, die nur zur Bestimmung der Ausfallwahrscheinlichkeit verwendet wurde. Die Korrelation dieser Variablen mit den Variablen V4 bis V13 wurde auf $CM + 0.1$ festgesetzt. Die resultierende Korrelationsstruktur gibt Tabelle 1 wieder.

Tabelle 1: Gegebene Populationskorrelationsstruktur

	Hilfsvariablen			Hauptvariablen				Selektionsvariable (V14)
	V1	V2	V3	V4	V5	...	V13	
V1		.2	.2	.15	.1515	.15
V2			.2	.15	.1515	.15
V3				.15	.1515	.15

Drei verschiedene MD erzeugende Prozesse wurde simuliert: Missing at Random (MAR), Missing at Random within Classes (MARC) und ein Hazard-Modell: Missing above Threshold (MAT).

Das Ausmaß fehlender Werte (MPER) für jeden Prozeß wurde auf 1%, 5% und 10% festgelegt. Die Löschung fehlender Werte erfolgte für alle Variablen unabhängig voneinander, so daß Fälle mit fehlenden Werten auf allen Hauptvariablen (V4 to V 13) möglich waren. Die Löschungen wurden im MAR-Modell zufällig vorgenommen. Das MARC-Modell basiert auf der Annahme, daß die Stichprobe aus zwei verschiedenen Populationen stammt, die sich in Hinsicht auf die Ausfallwahrscheinlichkeiten und die Interkorrelationsstruktur unterscheiden. Innerhalb der Populationen fehlen die Daten zufällig. In der Simulation wurden die erzeugten Stichproben mit einer Wahrscheinlichkeit von 0.8 aus einer Subpopulation mit einer Interkorrelation von

CM und mit der Wahrscheinlichkeit 0.2 aus einer Subpopulation mit einer Interkorrelation von 0.1 gezogen. Das Verhältnis der Ausfallwahrscheinlichkeiten wurde so auf 1:10 festgelegt, daß der Anteil fehlender Werte im kombinierten Sample dem Prozentsatz der experimentellen Bedingung (MPER) entsprach. Der dritte simulierte Prozeß war ein Hazard-Modell. Falls für einen gegebenen Fall eine ungemessene Variable einen unbekanntem Schwellenwert überschreitet, fällt der Fall aus. Dieses MAT-Modell wurde implementiert, indem eine einfache Zufallstichprobe erzeugt wurde, bei der dann diejenigen Fälle gelöscht wurden, bei denen eine Selektionsvariable (V14) einen Schwellenwert überschritt. Für jeden der drei Level von MPER wurde ein Schwellenwert gewählt, der zu dem gewünschten Ausmaß fehlender Werte führte.

Zehn verschiedene MD-Techniken (PROG) wurden verwendet:

1. - LISTWISE (vollständige Fälle)
2. - PAIRWISE (paarweise vorhandene Werte)
3. - MEAN (Mittelwertersetzung)
4. - CELLMEAN (Zellenmittelwertersetzung)
5. - HOT-DECK (Sequentielle variablenweise Fallverdoppelung)
6. - REGRESSION (Einfache Regression)
7. - HAZARD (Regression mit MD-Indikator)
8. - DEAR (Hauptkomponente)
9. - MULT (Multiple Regression)
10. - ILS (iterierte multiple Regression)

Für jede unvollständige Variable verwendete CELLMEAN den Zellenmittelwert der vorhandenen Daten einer Crossbreak-Tabelle der Variable mit den ersten beiden zusätzlichen Variablen (V1,V2). HOT-DECK verwendete eine Kreuztabelle der Fälle mit den Variablen V1 und V2 als Imputationsklassen. REGRESSION verwendete die am stärksten korrelierende Variable als Prädiktor. HAZARD verwendete neben derselben Variable wie REGRESSION die Anzahl der fehlenden Werte des gegebenen Falles als Prädiktor. DEAR ersetzte die fehlenden Werte durch Schätzungen auf der Basis der ersten Hauptkomponente, die aus der PAIRWISE-Korrelationsmatrix berechnet wurde. MULT verwendete eine multiple Regression mit allen verfügbaren Variablen eines gegebenen Falles (die Prozedur entspricht damit der von Buck, allerdings ohne Korrektur der Kovarianzmatrix). Die resultierenden Schätzungen von MULT bildeten den Ausgangspunkt für ILS.

ILS wiederholte die multiplen Regressionen bis kein Korrelationskoeffizient in zwei aufeinanderfolgenden Schritten sich um mehr als 0.0001 vom vorherigen unterschied.

Zwei Aspekte der Leistungsfähigkeit der MD-Techniken sollten untersucht werden: Die Güte der Schätzung von Korrelationskoeffizienten und die Güte der Schätzung einzelner fehlender Werte.

Der Bias der geschätzten Korrelationsmatrizen wurde berechnet als die mittlere Abweichung der geschätzten Korrelationen von den tatsächlichen Korrelationen:

$$RDEV = \frac{\sum_{i=1}^{p-1} \sum_{j=i+1}^p RT_{i,j} - R_{i,j}}{p(p-1)}$$

Timm (1970:426) verwendete die euklidische Norm der Differenzmatrix: $\text{Tr} [(RT-R)(RT-R)']$; in Anlehnung an Gleason/Staelin (1975:242) wurde eine Modifikation dieses Maßes, das "root mean square residual"¹⁸, in der Simulation verwendet:

$$WRSDEV = \left[\frac{\sum_{i=1}^{p-1} \sum_{j=i+1}^p (RT_{i,j} - R_{i,j})^2}{p(p-1)} \right]^{1/2}$$

Als drittes Ähnlichkeitsmaß für die Matrizen wurde der Korrelationskoeffizient der beiden als Datenvektoren aufgefaßten Korrelationsmatrizen berechnet¹⁹. Dieser Pearson-Korrelationskoeffizient wurde dann mit der Fisher-Transformation in z-Werte umgerechnet:

$$ZMKS = 0.5 \ln \left[\frac{(1+r)}{(1-r)} \right]$$

Die Leistungsfähigkeit der Ersetzungsverfahren (DECK, REGRESSION, HAZARD, MULT, ILS)²⁰ in Hinsicht auf die Schätzung²¹ einzelner fehlender Werte wurde durch vier verschiedene Maße erfaßt.

¹⁸ Dies ist mit dem "overall deviation index" von Kim/Curry (1977: 224) identisch.

¹⁹ In GAUSS ist dies mit wenigen Tastendrücken möglich: R=CORRX (VEC (RT) ~VEC(R)).

²⁰ Obwohl MEAN und CELLMEAN fehlende Werte ersetzen, wurden sie als triviale Ersetzungsmethoden aufgefaßt, wenn man an der Schätzung einzelner fehlender Werte interessiert ist.

Das erste Maß ist die "root mean square deviation" (z.B. von Kalton/Santos 1983 verwendet):

$$\text{ERSD} = \left[\frac{\sum \sum (DT_{i,j} - D_{i,j})^2}{M} \right]^{1/2}$$

wobei DT die wahre Datenmatrix, D die geschätzte Datenmatrix und M die Anzahl fehlender Werte darstellt. Die mittlere Abweichung wurde berechnet als:

$$\text{EMD} = (\sum DT_{i,j} - D_{i,j}) / M,$$

die mittlere absolute Abweichung als

$$\text{EMAD} = \sum |DT_{i,j} - D_{i,j}| / M.$$

Die letzte Maßzahl ist die Anzahl der exakt geschätzten fehlenden Werte (geschätzter Wert=wahrer Wert), dividiert durch die Anzahl der fehlenden Werte. Dieses Maß wird als HIT bezeichnet.

Ergebnisse der Simulation

Aufgrund des beschränkten Raumes werden die wichtigsten Ergebnisse als Plots präsentiert²². Die Y-Achse gibt den Mittelwert der abhängigen Variable über die 10 Wiederholungen unter denselben experimentellen Bedingungen für eine gegebene MD-Technik wieder. Die X-Achse (experimentelle Bedingung) wurde hier berechnet als $\text{COND} = (\text{MTYPE}-1)*27 + (\text{MPER}-1)*9 + (\text{CM}-1)*3 + N$. Daher entspricht $\text{COND}=1$ der Bedingung MAR ($\text{MTYPE}=1$), 1% MD ($\text{MPER}=1$), mittlere Interkorrelation gleich 0.2 ($\text{CM}=1$) und Stichprobengröße gleich 100 ($N=1$). Folglich wechselt in den Plots der fehlende Werte erzeugende Prozeß alle 27 Einheiten, MPER alle 9 Einheiten, CM alle 3 Einheiten und N mit jeder Einheit. Um einen bildlichen Eindruck der Leistung der Verfahren zu geben, wurden die jeweils 81 Meßpunkte einer MD-Technik mit einer Linie zu einem Profil über die experimentellen Bedingungen verbunden.

²¹ Mit der Ausnahme des Hot-Decks erzeugen alle Verfahren keine ganzzahligen Schätzungen. Da die Eingabewerte ganzzahlig waren (nach der Erzeugung kategorisiert in 7 Kategorien), wurden für die Berechnung der Datenähnlichkeitsmaße die Schätzungen kategorisiert (FORTRAN: DNINT).

²² Tabellen und multivariate Analysen (MANOVA's, Regressionen, Clusteranalysen) für eine Reihe spezieller Fragen finden sich bei Schnell (1985, 1986). Der Datensatz (SPSS-PC-Systemfiles und ASCII-Version) wird auf Anfrage zur Verfügung gestellt.

Plot 1 zeigt die mittlere Abweichung des geschätzten Korrelationskoeffizienten für die 81 experimentellen Bedingungen. Der Einfluß der experimentellen Bedingungen ist deutlich sichtbar. Obwohl die Profile ähnlich erscheinen, gibt es deutliche Unterschiede zwischen den Verfahren. Plot 2 zeigt dies am Beispiel der Profile für LISTWISE und PAIRWISE. Unter nahezu jeder getesteten Bedingung zeigt LISTWISE schlechtere Resultate. Die Differenzen zwischen LISTWISE und PAIRWISE erreichen bis zu .13 unter denselben experimentellen Bedingungen.

Auf den ersten Blick erscheinen die mittleren Abweichungen nicht sehr groß: Der Mittelwert unter MAR erreicht 0.03, unter MAT 0.05 und -0.04 innerhalb von MARC selbst für die schlechteste Technik (ein positives RDEV impliziert eine Unterschätzung). Für LISTWISE wird das absolute Maximum von RDEV mit 0.15 innerhalb von MAR, 0.18 innerhalb von MAT und -0.17 innerhalb von MARC erreicht. Für ILS beträgt die maximale Abweichung -0.06 innerhalb von MARC. Betrachtet man aber die "Struktur" der Korrelationsmatrix (ZMKS: Plot 3), so zeigt sich, daß scheinbar kleine Abweichungen einzelner Koeffizienten zu großen Abweichungen bei der Struktur der Korrelationsmatrix führen können: Selbst der Mittelwert von ZMKS erreicht Minimumwerte zwischen 0.58 (LISTWISE) und 1.16 (ILS). Selbstverständlich zeigen die Minimumwerte für ZMKS unter den einzelnen Bedingungen noch extremere Werte. So erreicht z.B. innerhalb von MAR LISTWISE den Wert 0.3 ($r=0.29$) als sein Minimum, dagegen ILS 1.22 ($r=.84$). Innerhalb von MARC erreicht ILS sein Minimum mit .76 ($r=.64$). Das Ergebnis zeigt, daß selbst eine der besten Techniken zu irreführenden Ergebnissen führen kann.

Unter praktischen Gesichtspunkten ist der Anteil korrekt geschätzter fehlender Werte (HIT) von Bedeutung (Plot 4). Trotz der Kategorisierung liegt der Anteil korrekt geschätzter fehlender Werte für die meisten Verfahren zwischen 0.25 and 0.40. Eines der besten Verfahren, ILS, erreicht im Mittel 38% innerhalb von MAR, 36% innerhalb von MAT und 33% innerhalb von MARC. HIT erreicht bei hohem CM und großen Stichproben relativ unabhängig von der Anzahl fehlender Werte Maximumwerte von 45%. Aber Plot 4 zeigt bemerkenswerte Unterschiede zwischen den Verfahren. Insbesondere DEAR ist gegenüber dem erzeugenden Prozeß hochempfindlich. Das Hot-Deck ist bei weitem das schlechteste Verfahren. MULT und ILS sind dagegen kaum unterscheidbar (Plot 4a).

Falls man an der Schätzung einzelner fehlender Werte interessiert ist, so ist die mittlere Abweichung der geschätzten Werte von den wahren Werten

von Interesse. Plot 5 zeigt eine starke Abhängigkeit von EMD vom fehlende Werte erzeugenden Prozeß. Die Unterschätzung innerhalb von MAR zwischen 0.017 und 0.025 steigt auf 0.295 und 1.05 innerhalb von MAT. Die Unterschiede zwischen den Verfahren sind deutlich sichtbar, insbesondere die schlechte Leistung des Hot-Decks²³.

Obwohl der Einfluß der experimentellen Faktoren auf die Leistungsfähigkeit der MD-Techniken sowohl von der Technik als auch vom fehlende Werte erzeugenden Prozeß abhängen²⁴, sind ein paar Tendenzen in den Plots (und den hier nicht berichteten Regressionen) evident. Da die Anzahl der Fälle fast niemals einen Effekt besitzt, scheint es angebracht, in weiteren Simulationen auf diesen Faktor zu verzichten. Die Leistung der Verfahren in Hinsicht auf die Schätzung einzelner fehlender Werte (EMD, EMAD, ERSD, HIT) kann nur teilweise durch die experimentellen Faktoren erklärt werden, das R^2 der multiplen Regressionen überschreitet selten 0.30 für HIT und 0.60 für ERSD. Die relative Wichtigkeit von MPER gegenüber CM hängt von einer Reihe von Faktoren ab, im allgemeinen ist CM aber für die multivariaten Verfahren (DEAR, MULT, ILS) wichtiger als für die anderen Techniken. Je mehr Redundanz in den Daten steckt, umso besser sind die Schätzungen. Wie die Plots (und die hier nicht berichteten MANOVA's) zeigen, kann nahezu jeder mögliche Interaktionseffekt gefunden werden, so daß weitere Generalisierungen kaum möglich sind. Für die multiplen Regressionen für die Maße der Güte der Schätzungen der Korrelationen (RDEV, WRSDEV, ZMKS) kann ein weit besserer Fit festgestellt werden (R^2 zwischen 0.75 and 0.87). In diesen Regressionen ist der Einfluß der Anzahl der MD nahezu immer stärker als der Effekt der Stärke der Interkorrelation. "Typische" Beta-Werte für CM liegen zwischen -0.10 und -0.20, für MPER zwischen -0.75 and -0.90.

Zusammenfassung

Die Ergebnisse zeigen die starke Abhängigkeit aller Verfahren vom - in der Regel unbekanntem - fehlende Werte erzeugenden Prozeß. Das wichtigste Resultat der Simulation ist daher die Feststellung, daß ein "bestes" Schätzverfahren nicht existiert. Selbst die "besten" Verfahren brechen unter

²³ Dies Ergebnis stimmt mit den Resultaten von Kaiser (1983) überein.

²⁴ Alle MANOVA-Interaktionsterme des Typs MTYPE*PROG, MTYPE*N, MTYPE*CM, MTYPE*MPER und die meisten der PROG*N, PROG*CM, PROG*MPER Interaktionen sind signifikant mit $p < 0.001$. Die meisten der entsprechenden 3-Wege-Interaktionen sind ebenfalls signifikant.

bestimmten Bedingungen zusammen. Die weitere Suche nach "besten" Schätzverfahren erscheint daher sinnlos. Die Ergebnisse der Simulation legen nahe, daß ein angemessener Umgang mit fehlenden Werten nicht darin bestehen kann, sie zu vernachlässigen oder sie mit einer "automatischen" Prozedur wie BMDPAM zu beseitigen, sondern nur darin, unterschiedliche plausible theoretische Annahmen über den fehlende Werte erzeugenden Prozeß während der Imputation zu simulieren. Solche "multiplen Imputationen" (Rubin 1987) scheinen der einzige Weg bei der Analyse eines Datensatzes mit einem nicht-trivialen Anteil fehlender Werte zu sein.

Literatur

- ANDERSON, A.B./BASILEVSKY, A./HUM, D.J. (1983): Missing Data: A Review of the Literature; in: ROSSI, P.H./ WRIGHT, J.D./ ANDERSON, A.B. (eds.): Handbook of Survey Research, New York, p.415-493.
- BASILEVSKY, A./SABOURIN, D./HUM, D./ANDERSON, A. (1985): Missing Data Estimators in the General Linear Model: An Evaluation of Simulated Data as an Experimental Design; in: Communications in Statistics, Simulation and Computation, 14, 2, p.371 -394.
- BEALE, E.M.L./LITTLE, R.J.A. (1975): Missing Values in Multivariate Analysis; in: Journal of the Royal Statistical Society, Series B, 37, p. 129-146.
- BERGER, M.P.F. (1979): A FORTRAN IV Program for the Estimation of Missing Data; in: Behavior Research Methods and Instrumentation, 11, 3, p.395-396.
- BERK, R.A./RAY, S.C. (1982): Selection Biases in Sociological Data; in: Social Science Research, 11, p.352-398.
- BERK, R.A. (1983): An Introduction to Sample Selection Bias in Sociological Data; in: American Sociological Review, 48, p.386-398.
- BROWN, C.H. (1983): Asymptotic Comparison of Missing Data Procedures for Estimating Factor Loadings; in: Psychometrika, 48, 2, p.269-291.
- BUCK, S.F. (1960): A Method of Estimation of Missing Values of Multivariate Data Suitable for Use With an Electronic Computer; in: Journal of the Royal Statistical Society, Series B, 22, p.302-307.
- CHAPMAN, D.W. (1983): An Investigation of Nonresponse Imputation Procedures for the Health and Nutrition Examination Survey; in: MADOW, W.G./NISSELSON, H./OLKIN, I. (eds.): Incomplete Data in Sample Surveys, Vol.1, p.435-483, New York.

- COHEN, J./COHEN, P. (1975): Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences, Hillsdale.
- DEAR, R.E. (1959): A Principal Component Missing Data Method for Multiple Regression Models. Technical Report SP-86, System Development Corporation, Santa Monica, Calif.
- DEMPSTER, A.P./LAIRD, N.M./RUBIN, D.B. (1977): Maximum Likelihood from Incomplete Data via the EM-Algorithm; in: Journal of the Royal Statistical Society, Series B, 3g, p.1-22.
- DILLON, W.R./GOLDSTEIN, M. (1984): Multivariate Analysis. Methods and Applications, New York.
- FAY, R.E. (1986): Causal Models for Patterns of Nonresponse; in: JASA, 81, 394, S.354-365.
- FEDERSPIEL, C.F./MONROE, R.J./GREENBERG, B.G. (1959): An Investigation of Some Multiple Regression Methods for Incomplete Samples, University of North Carolina Institute of Statistics, Mimeo Series #236.
- FIELDING, A. (1979): Binary Segmentation: The Automatic Interaction Detector and Related Techniques for Exploring Data Structure; in: O'MUIRCHEARTAIGH, C.A./PAYNE, C. (eds.): The Analysis of Survey Data, Vol.1, p.221 -257, Chichester.
- FINKBEINER, C. (1979): Estimation for the Multiple Factor Model when Data are Missing; in: Psychometrika, 44, p.409-420.
- FORD, B.L. (1983): An Overview of Hot-Deck Procedures; in: MADOW, W.G./ OLKIN, I. / RUBIN, D.B. (eds.): Incomplete Data in Sample Surveys, New York, Vol. 2, p.1 85-207.
- FRANE, J. (1978): Missing Data and BMDP: Some Pragmatic Approaches; in: ASA Proceedings of the Statistical Computing Section, p.27-33.
- FUCHS, C. (1982): Maximum Likelihood Estimation and Model Selection in Contingency Tables With Missing Data; in: JASA, 77, 378, S.270-278.
- GLASSER, M. (1964): Linear Regression Analysis With Missing Observations Among the Independent Variables; in: JASA, 59, p. 834-844.
- GLEASON, T.C./STAEIN, R. (1975): A Proposal for Handling Missing Data; in: Psychometrika, 40, p.229-252.
- GREENE, W.H. (1986): LIMDEP User's Manual.
- HAAGENAARS, J.A. (1988): Log-Linear Analysis with Latent Variables and Missing Data; Paper presented at the International Conference on Social Science Methodology, Dubrovnik, May/June 1988.
- HAITOVSKY, Y. (1968): Missing Data in Regression Analysis; in: Journal of the Royal Statistical Society, Series B, 63, p.67-82.

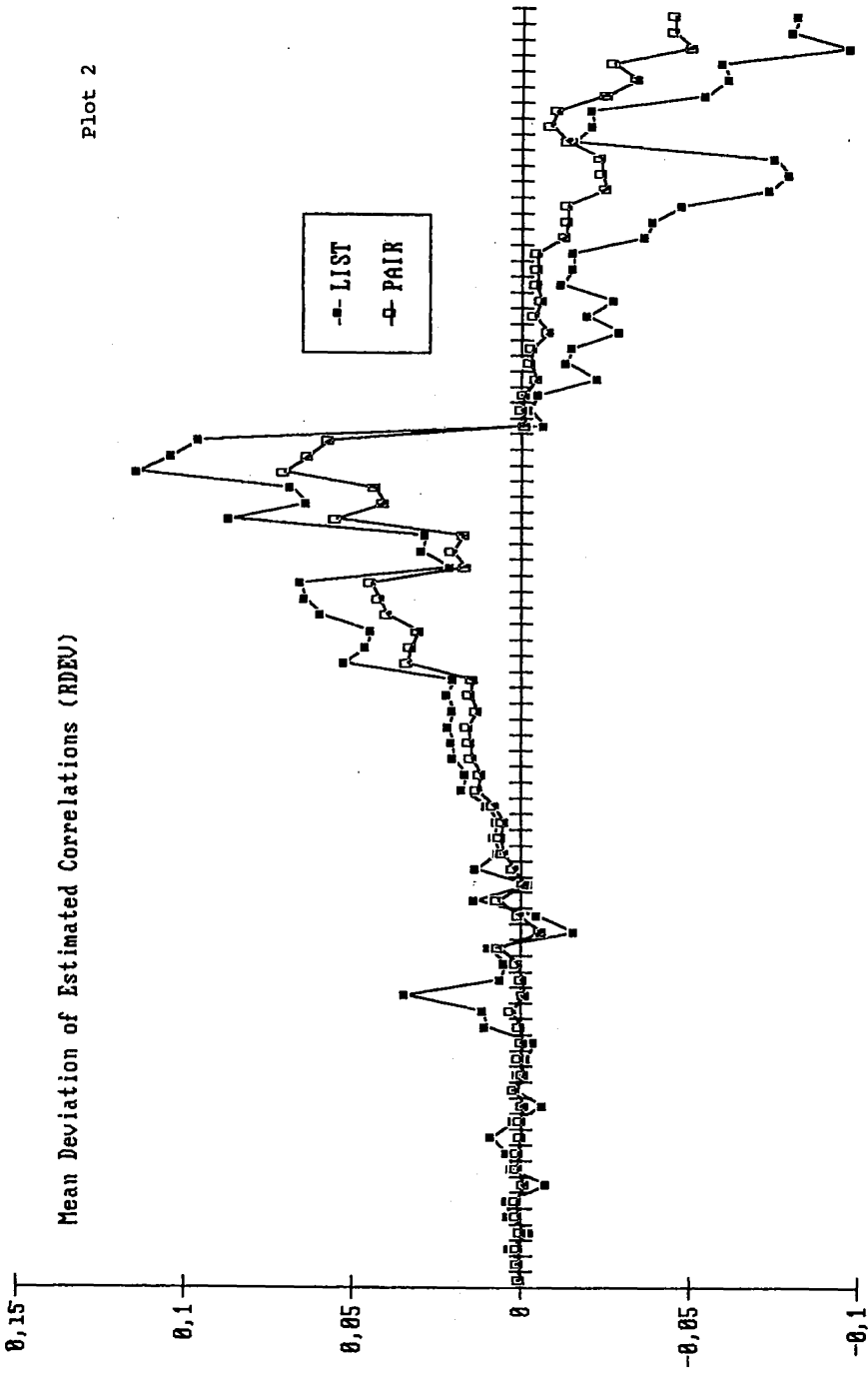
- HAMILTON, M.A. (1975): Regression Analysis When There are Missing Observations - A Survey and Bibliography. Technical Report 1-3-75. Statistical Laboratory, Montana State University.
- HARTIGAN, J.A. (1975): Clustering Algorithms, New York.
- HECKMAN, J.J. (1976): The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for such Models; in: *Annals of Economic and Social Measurement*, 5/4, p.475-492.
- HEIBERGER, R.M. (1977): Regression with the Pairwise-Present Covariance Matrix: A Dangerous Practice; in: *ASA Proceedings of the Statistical Computing Section*, p. 38-47.
- HILL, M./DIXON, W.J. (1981): Missing Data: Search for Patterns; in: *ASA Proceedings of the Statistical Computing Section*, p.57-60.
- KAISER, J. (1983): The Effectiveness of Hot-Deck Procedures in Small Samples; in: *ASA Proceedings of the Section on Survey Research Methods*, p.523-528.
- KAISER, J. (1984): The Estimation of Missing Values, Ph.D. Thesis, University of Kansas.
- KALTON, G./KASPRZYK, D. (1982): Imputing for Missing Survey Responses; in: *ASA Proceedings of the Section on Survey Research Methods*, p.22-31.
- KALTON, G. (1983): Compensating for Missing Survey Data, Survey Research Center, Institute for Social Research, University of Michigan.
- KIM, J.O./CURRY, J. (1977): The Treatment of Missing Data in Multivariate Analysis; in: *Sociological Methods and Research*, 6, 2, p. 215-239.
- LITTLE, R.J.A. (1979): Maximum Likelihood Inference for Multiple Regression with Missing Values: A Simulation Study; in: *Journal of the Royal Statistical Society, Series B*, 41, 1, p.76-87.
- LITTLE, R.J.A. (1983): The Nonignorable Case; in: MADOW, W.G./OLKIN, I./RUBIN, D.B. (eds.): *Incomplete Data in Sample Surveys*, Vol 2, p.383-413, New York.
- LITTLE, R.J.A./RUBIN, D.B. (1987): *Statistical Analysis With Missing Data*, New York (Wiley).
- MÖNTMANN, V./BOL LINGER, G./HERRMANN, A. (1983): Tests auf Zufälligkeit von "Missing Data"; in: WILKE, H.(ed.): *Statistik Software in der Sozialforschung*, p.87-101, Berlin.
- NELSON, F.D. (1977): Censored Regression Models with Unobserved, Stochastic Censoring Thresholds; in: *Journal of Econometrics*, 6, 1977, p.309-327.

- O'GRADY, K.E. (1982): Regression Estimation of Missing Data; in: Behavior Research Methods and Instrumentation, 14, 3, p.359-360.
- OH, H.L./SCHEUREN, F.J. (1983): Weighting Adjustment for Unit Non-response; in: MADOW,W.G./OLKIN,I./RUBIN,D.B. (eds.): Incomplete Data in Sample Surveys, New York, Vol. 2, p.143-184.
- ORCHARD, T./WOODBURY, M.A. (1970): A Missing Information Principle: Theory and Practice; in: Proceedings of the 6th Berkeley Symposium on Mathematical Statistics and Probability, 1, p.697-715.
- RIZVI, H. (1983): An Empirical Investigation of Some Item Nonresponse Adjustment Procedures; in: MADOW,W.G./ NISSELSOHN,H./ OLKIN, I. (eds.): Incomplete Data in Sample Surveys, Vol.1, p.299-366, New York.
- RUBIN, D.B. (1976): Inference and Missing Data; in: Biometrika, 63, p.581 - 592.
- RUBIN, D.B. (1987): Multiple Imputation for Nonresponse in Surveys, New York (Wiley).
- SANDE, G. (1979): Numerical Edit and Imputation; in: Bulletin of the International Statistical Institute, 42nd Session of the International Statistical Institute, p.455-463.
- SANDE, I.G. (1982): Imputation in Surveys: Coping With Reality; in: The American Statistician, 36, 3, 1982, p.145-152.
- SANTOS, R.L. (1981): Effects of Imputation on Complex Statistics, Survey Research Center, Institute for Social Research, University of Michigan.
- SCHMEE, J./HAHN, G.J. (1979): A Simple Method for Regression Analysis with Censored Data; in: Technometrics, 21, 4, p.417-432.
- SCHNELL, R. (1985): Zur Effizienz einiger Missing-Data-Techniken; in: ZUMA Nachrichten 17, p.50-74.
- SCHNELL, R. (1986): Missing-Data-Probleme in der empirischen Sozialforschung, Dissertation, Bochum.
- SONQUIST, J.A./BAKER, E.L./MORGAN, J.N. (1971): Searching for Structure, Institute for Social Research, University of Michigan, Ann Arbor.
- TIMM, N.H. (1969): Estimating Variance-Covariance and Correlation Matrices from Incomplete Data, University of California, Berkeley, Ph.D.Thesis.
- TIMM, N.H. (1970): The Estimation of Variance-Covariance and Correlation Matrices from Incomplete Data; in: Psychometrika, 35, p.417-437.
- VACEK, P.M./ASHIKAGA, T. (1980): An Examination of the Nearest Neighbor Rule for Imputing Missing Values; in: ASA Proceedings of the Statistical Computing Section, p.326-331.

- VAN GUILDER, M./AZEN, S. (1981): Conclusions Regarding Algorithms for Handling Incomplete Data; in: ASA Proceedings of the Statistical Computing Section, p.53-56.
- WISHART, D. (1978): Treatment of Missing Values in Cluster Analysis; in: International Association for Statistical Computing: COMPSTAT 1978, p.281 -287, Wien.
- WISHART, D. (1985): Estimation of Missing Values and Diagnosis Using Hierarchical Classifications; in: Computational Statistics Quarterly, 2, 1, 1985, p. 25-134.

Plot 2

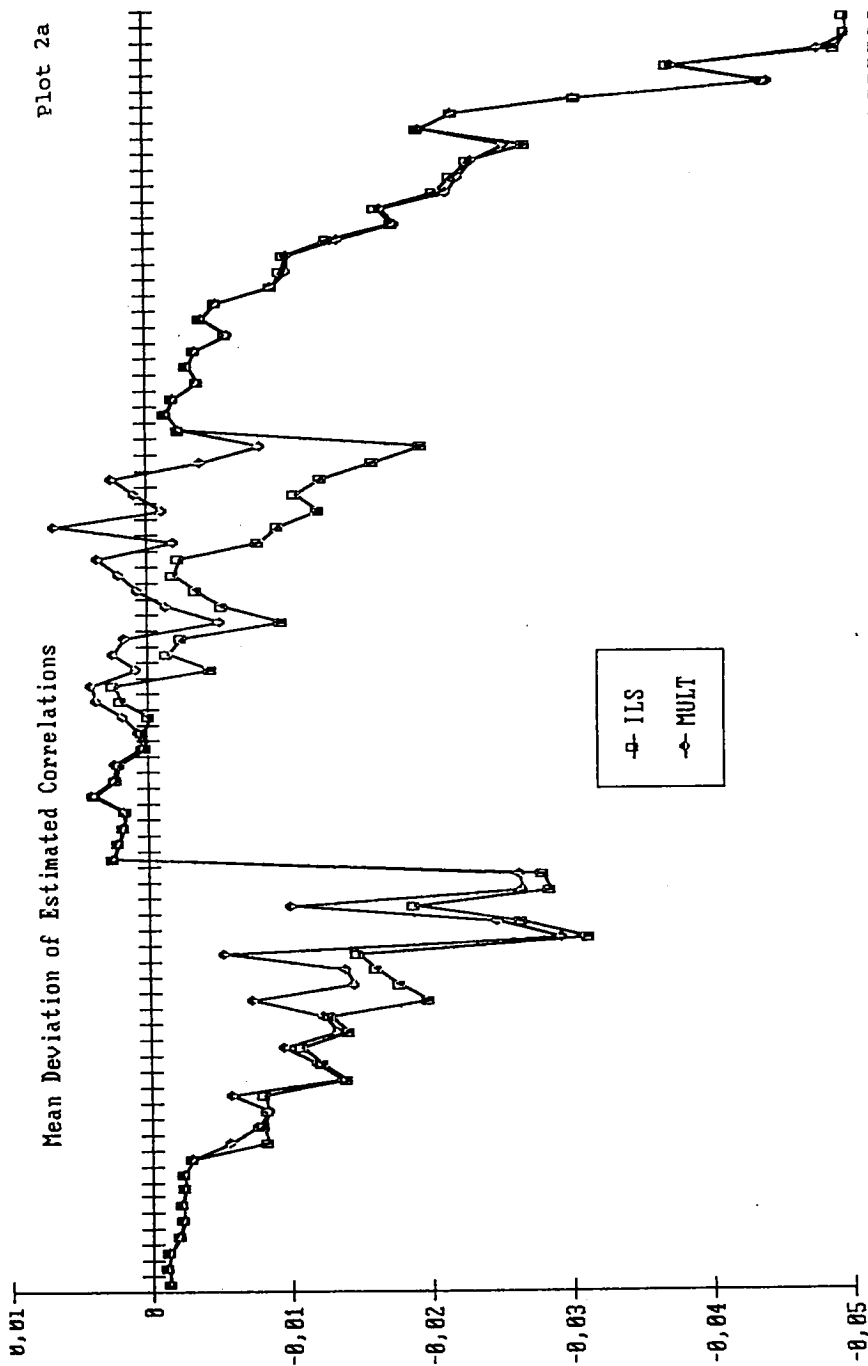
Mean Deviation of Estimated Correlations (RDEV)



12345678911111111122222222333333334444444455555555666666667777777788
 01234567890123456789012345678901234567890123456789012345678901

Experimental condition

Plot 2a



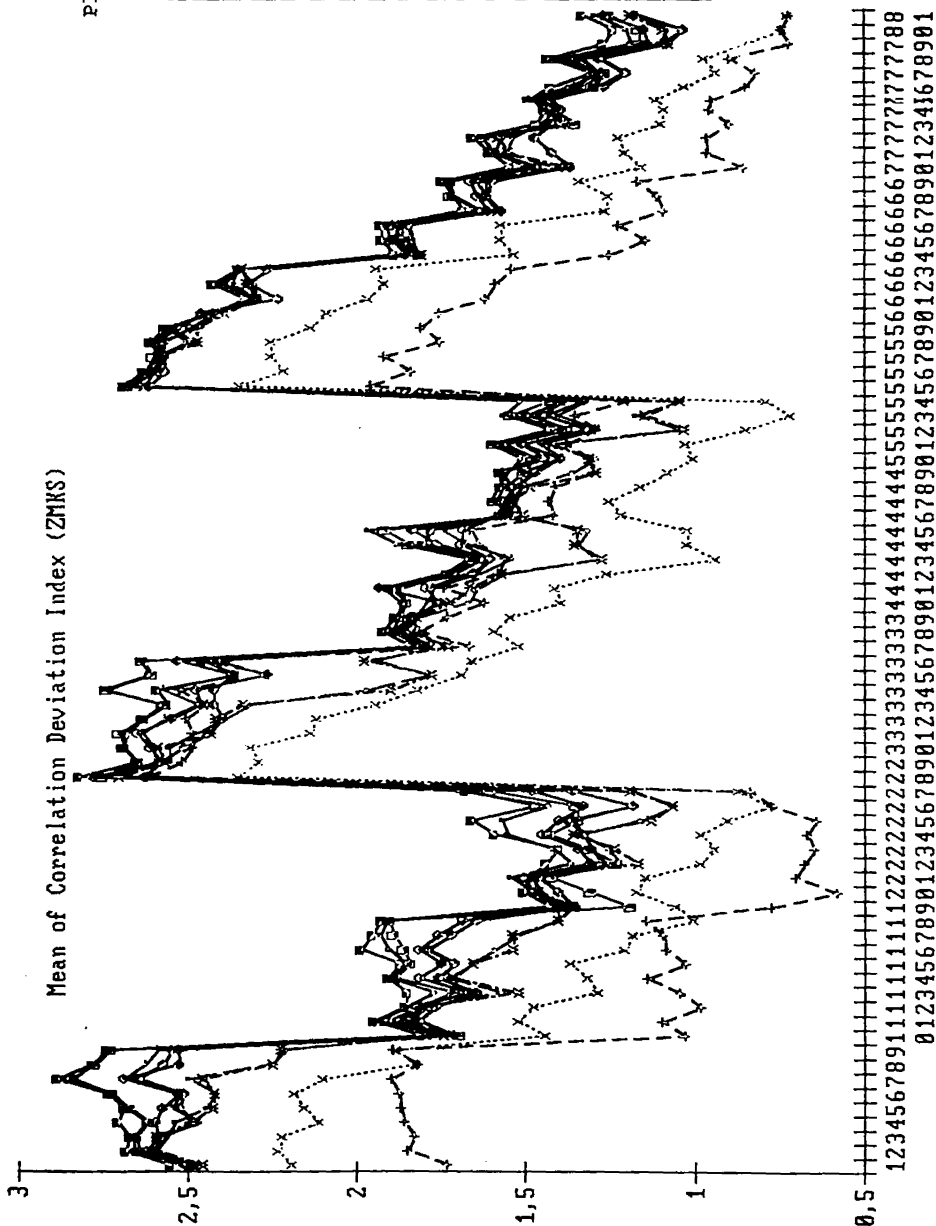
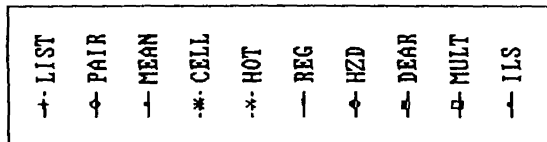
□ ILS
○ MULT

123456789111111122222233333344444455555566666677777788
0123456789012345678901234567890123456789012345678901

Experimental condition

Mean of Correlation Deviation Index (ZMKMS)

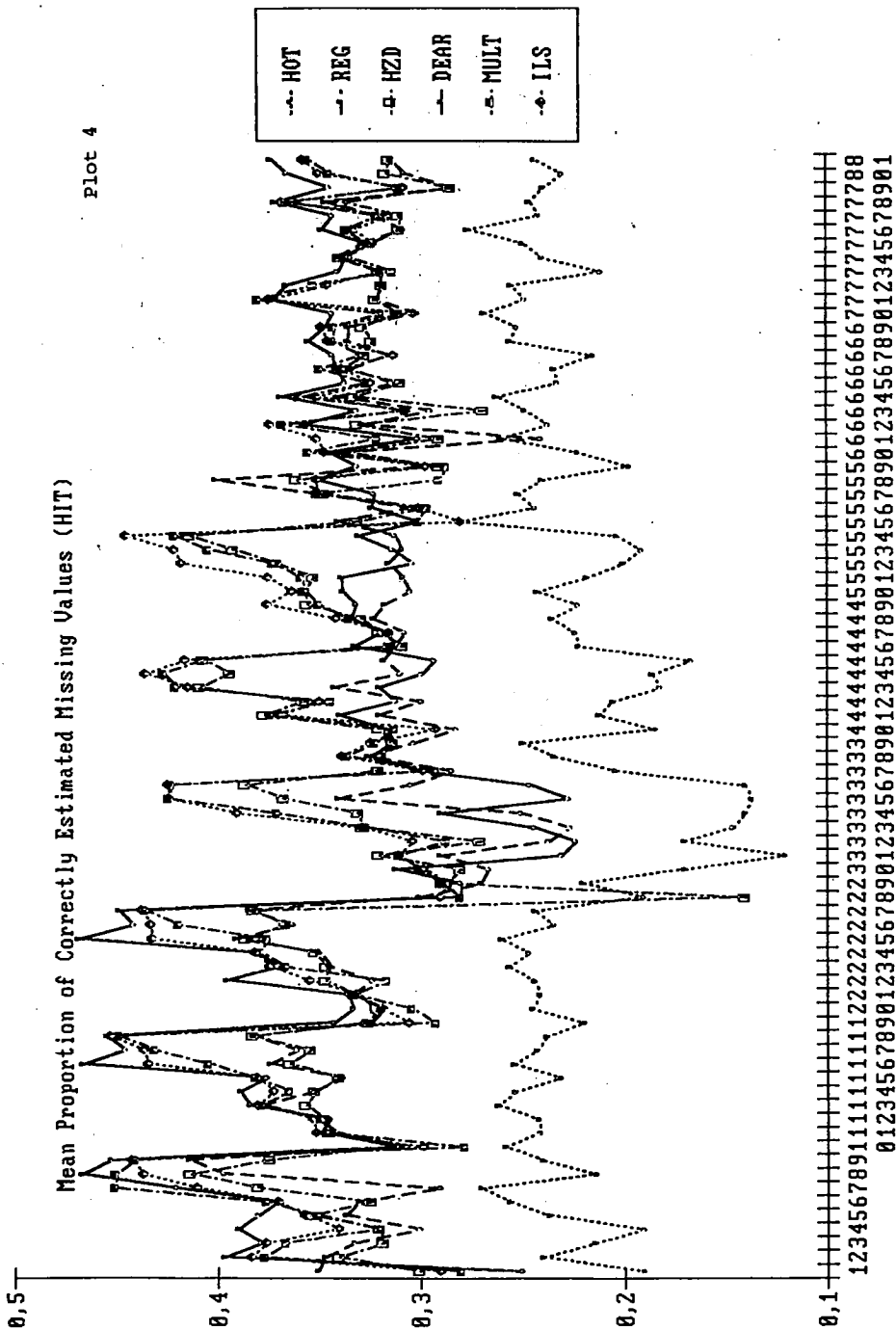
Plot 3



Experimental condition

Plot 4

Mean Proportion of Correctly Estimated Missing Values (HIT)



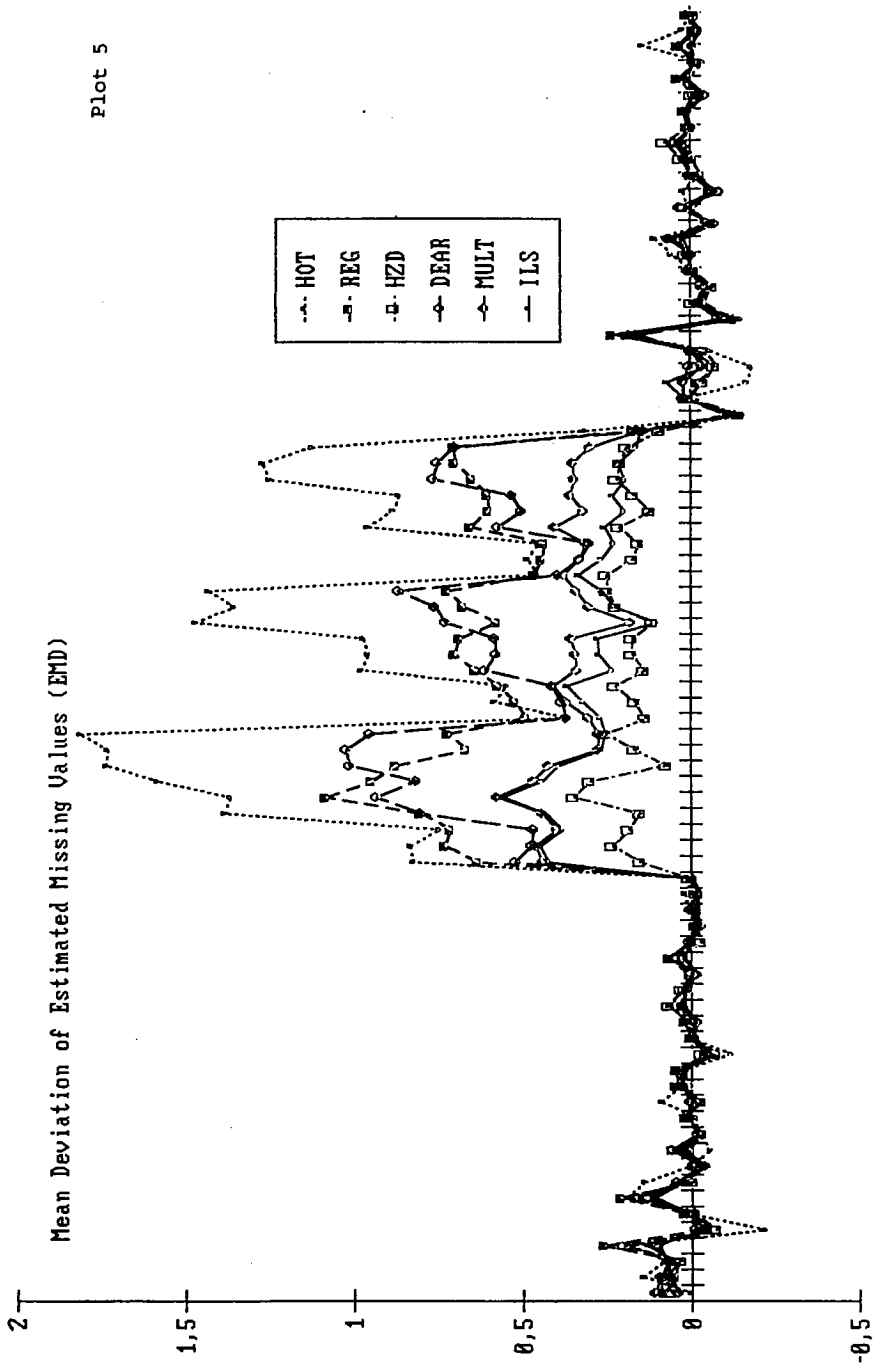
0.5
0.4
0.3
0.2
0.1

1234567891111111122222223333333444444455555556666666777777788
0123456789012345678901234567890123456789012345678901

Experimental condition

Plot 5

Mean Deviation of Estimated Missing Values (EMD)



---△---	HOT
—■—	REG
---□---	HZD
-○-	DEAR
—○—	MULT
—△—	ILS

123456789111111111222222223333333344444444555555566666667777777788
 01234567890123456789012345678901234567890123456789012345678901

Experimental condition

Probleme formaler Modelle in den historischen Wissenschaften

von Johannes Gordesch

1 Modelltechniken in den historischen Wissenschaften

Das Studium vereinfachter, aber streng gefaßter Strukturen spielt in den zeitgenössischen Formen von Sozial- und Geisteswissenschaften eine hervorragende Rolle. In den historischen Disziplinen wurde mit dem Aufkommen der Strukturgeschichte im Gegensatz zur vormals vorherrschenden Ereignisgeschichte der Boden bereitet für ein Denken in Modellen, in "simplified skeletal systems" (BEACH 1957). Theorien lassen sich dann von Modellen bestenfalls durch einen größeren Anwendungsbereich oder eine differenziertere Ausarbeitung unterscheiden.

Damit begibt sich die Wissenschaft in deutlichen Gegensatz zum Vorrang einer allumfassenden, überdauernden Theorie, die sich oftmals zu einer Heilslehre entwickelt. Dieses Vorgehen konnte nicht ohne Widerspruch bleiben. Bisweilen suchte und sucht man polemisch eine Verbindung zu nationalen oder politischen Denkstilen: Den Deutschen oder den Russen (principlalnost') wird eher der Hang zur großen Theorie zugeschrieben, den pragmatischen Angelsachsen oder den Vertretern einer oberflächlichen, allerlei wirtschaftlichen oder politischen Interessen unterworfenen Konsumhaltung die Betonung unverbindlicher, kurzfristig genutzter Modelle. Anderen ist wiederum der große Geist abhanden gekommen (man setze eine an der Kunst der Griechen und Römer orientierte Archäologie gegen eine heutige Archäometrie!). Wie die Argumente auch immer lauten, Zeitgeist und praktische Bewährung haben ihr Urteil gesprochen.

Besonders schön läßt sich der Modellbegriff an der modernen Archäologie demonstrieren. Modell (im Maßstab 1:1) ist zunächst Schauplatz der experimentellen Archäologie, wo die materielle Kultur (Herstellung von Gerätschaften, Bau von Wohnstätten, Gewinnung von Erzen, Bewältigung des Nahrungsmittelproblems) im Mittelpunkt der Betrachtung stehen. Modell ist hingegen weitaus häufiger Denkmodell. Im einfachsten Fall besteht so ein Modell nur aus zwei (Streudiagramm, Scatter Plot) oder drei Variablen

(Trend Surface Analysis, wo zwei Variablen wie üblich im Diagramm, die dritte als Kote oder Niveaulinie wie bei Landkarten auftreten). Diese Beziehungen werden durch eine Regressionsgleichung algebraisch beschrieben (z.B. die klassische lineare Regression mit ein oder zwei unabhängigen Variablen; vgl. HODDER und ORTON 1976). Das andere Extrem stellen die großen ökonometrischen Modelle mit mehreren tausend Variablen dar. Die Modelle der historischen Sozialforschung können für mathematische Modelle in den Geschichtswissenschaften als typisch angesehen werden: Sie gehen so gut wie nie über zwanzig oder dreißig Variablen hinaus.

2 Statistisches Modellieren

Erstaunlicherweise tritt an die Seite der diffizilen realwissenschaftlichen Arbeit häufig eine völlig unreflektierte statistische Methodik: Bald suchen Realwissenschaftler durch die Kenntnis möglichst vieler statistischer Tests ihre Hypothesenprüfung zu unterstützen (der klassische Fall) oder (die neuere Mode) durch den Einsatz immer weiterer statistischer Programmpakete zu imponieren. Damit gehen aber die Vorzüge der statistischen Methodik, nämlich der Gewinn an Sicherheit und Präzision, verloren. Schon bei den einfachsten statistischen Tests steht ja ein mathematisches Modell dahinter, z. B. bei der Varianzanalyse die Beschreibung realwissenschaftlicher Sachverhalte durch Mittelwerte (die eine starke Informationsreduktion bewirken) und deren Veränderlichkeit (gemessen durch die Varianz). Das ist mehr als bloß die stochastische Voraussetzung der Normalverteilung (einer Verteilung, die durch die ersten zwei Momente bestimmt ist); auch unabhängig von einer statistischen Betrachtung wird ein bestimmter funktionaler Zusammenhang angenommen, der dem Sachverhalt entsprechen kann oder nicht.

Nun hat sich die statistische Mode geändert (wenn man will, es hat sich ein Paradigmenwechsel vollzogen): Die Korrelationsrechnung (Besteht ein signifikanter Zusammenhang?) wurde durch die Regressionsrechnung (Wie berechne ich eine Größe aus anderen?) abgelöst, desgleichen ein Wechsel von bivariaten Fragestellungen zu multivariaten vollzogen. Der klassische Einsatz statistischer Tests erfolgte im Sinne der Korrelationsrechnung zur Prüfung von Zusammenhängen und ist somit ebenfalls im Rückgang. Beim Regressionsmodell ist natürlich die Frage nach der Stichprobenabhängigkeit der Koeffizienten ebenso sinnvoll (Wie ändern sich die Koeffizienten in den Gleichungen bei einer Wiederholung der Untersuchung?). Die Modelle sind zu-

meist lineare Modelle, da sie statistisch wie numerisch wesentlich leichter zu handhaben sind als nichtlineare. Die leistungsfähigen und preisgünstigen Computer mit der benutzerfreundlichen Software ermöglichen einem großen Benutzerkreis, hochgezüchtete statistische Modelle einzusetzen. Der Fortschritt gegenüber der blinden Testfreudigkeit ist trotzdem gering, weil für gewöhnlich die Adäquatheit der jeweiligen statistischen Modelle nicht überprüft wird. Die Ursachen liegen zum Teil in schlichter Unwissenheit, zum Teil darin, daß die Tests innerhalb der heutigen komplexen Modelle oftmals nicht bekannt oder von geringer Entscheidungsfähigkeit sind.

Es ist unbedingt vonnöten, bereits bei der Erstellung des realwissenschaftlichen Modells formalwissenschaftliche Vorstellungen einzuarbeiten, wie dies seit langem in den physikalischen Wissenschaften Standard ist. "Formalwissenschaftlich" bedeutet in erster Linie "mathematisch", wie in der Physik auch, in zweiter Linie "statistisch", wenn die Anpassung an die Daten vollzogen wird. Die Physiker nehmen den statistischen Teil recht leicht, können vielleicht dies bei ihren Daten auch tun, legen aber größtes Augenmerk auf die Formulierung mathematischer Modelle und deren (noch theoretische, das heißt vor der Anpassung an die Daten erfolgende) Interpretation. In den Sozialwissenschaften hat eine verstärkte Hinwendung zu (statistischen) Modellierungstechniken mit Lazarsfeld, Blalock, Boudon, Coleman vor mehreren Jahrzehnten eingesetzt, oftmals unter dem Titel "kausale Modelle". Die speziellen statistischen Modelle mögen heute als veraltet gelten, das Prinzipielle ist geblieben. Die historischen Wissenschaften sind in ihrer breiten Masse leider nur selten diesem Weg gefolgt, wenngleich es namentlich im angelsächsischen Kulturbereich herausragende Beispiele gibt.

Wenigstens sollte jedoch eine wohlüberlegte Vorentscheidung nach folgenden Gesichtspunkten geschehen: Kausalität, Finalität, Werturteile; Stichprobe und Grundgesamtheit; Automat (Systeme mit Eingabe, Ausgabe und beschreibenden Zuständen), Boolesche Funktionen (qualitative Modelle, z.B. Entscheidungstabellen) und andere mathematische Modelle; Linearität, Nichtlinearität, Zeitabhängigkeit (Prozesse).

3 Modellbegriff

Ein Modell umfaßt - wie etwa von TINBERGEN und BOS (1962) für die Wirtschaftswissenschaften konzipiert - folgende Teile:

- (1) Eine Liste von Variablen (das heißt von Faktoren, die das Geschehen beschreiben können).
- (2) Eine Aufgliederung der Variablen in zunächst zwei Gruppen, nämlich in bekannte und in erst zu berechnende Größen, oder in exogene und endogene Variablen.
- (3) Eine Liste von Relationen oder spezieller von Gleichungen, welche die Beziehungen zwischen den einzelnen Variablen festlegen. Die Relationen oder Gleichungen werden je nach ihrer Funktion (Definitionen, Beziehungen in besonderen Subsystemen und dergleichen) untergliedert.
- (4) Bestimmende Konstanten ("Parameter"), welche das zunächst noch recht allgemeine Modell dem realen Einzelfall anpassen.

4 Meßproblematik

Von größter Bedeutung ist es, daß in keiner Wissenschaft Messen ohne Bezug auf einen speziellen Kontext (Modell, Operationalisierung) und ohne normative Festlegungen (wie das Anforderungsniveau) erfolgen kann. In den historischen Wissenschaften vereinen sich naturwissenschaftliche Methoden (beispielsweise in der Archäometrie) mit geisteswissenschaftlichen. Daher sei im folgenden eine Synthese zwischen sozialwissenschaftlicher und naturwissenschaftlicher Meßtheorie versucht und darin die wesentlichen Meßprobleme aufgezeigt. Zur Illustration denke man etwa an so gegensätzliche Beispiele wie die Vermessung von Grundrissen oder die Entwicklung von Getreidepreisen.

Messen heißt, den unbekanntem Wert einer exogenen Variablen des Modells auf empirische Weise zu ermitteln. Dies geschieht stets durch einen Vergleich mit der Einheit der betreffenden Größenart, die alleine innerhalb des Kontexts eines Modells und durch eine operationale Definition bestimmt ist.

Die Operationalisierung der Begriffe und Hypothesen entspricht dabei dem Meßprinzip in der naturwissenschaftlichen Metrologie. Die Meßmethode gibt Auskunft darüber, wie der Vergleich mit der Einheit durchgeführt wird. Im allgemeinen wird es sich im sozial- und geisteswissenschaftlichen Bereich um die Differenzenmethode (wieviel mehr als der Ausgangswert?) handeln, so daß der Begriff "Meßmethode" aus der Metrologie uner-

giebig wird. Man kann daher in den historischen Wissenschaften die Begriffe des Meßprinzips und des Meßverfahrens (Meßprinzip plus Meßmethode) zusammenfallen lassen.

Die Schwierigkeiten liegen einmal in einer geeigneten Operationalisierung, zum anderen jedoch in der oft nicht einmal als Problem erkannten "Kalibrierung" oder des "Einmessens", sowie der Eichung. Die Kalibrierung wird in Psychologie und Soziologie für gewöhnlich "Skalierung" genannt. Skalierung setzt ein bestimmtes, auszuwählendes Skalierungsverfahren voraus, das nicht ohne weitreichende Annahmen über die zugrundeliegende Wirklichkeit auskommen kann. Eichung ist eine normative Feststellung, daß das Meßmittel den vorgeschriebenen Anforderungen entspricht (z.B. Validität und Reliabilität). Die Eichung muß von Fall zu Fall und bei Vorliegen besonderer Umstände wiederholt werden.

Ein besonderes Problem stellt die Robustheit der Modelle dar, das heißt, wie weit sich ein Modell bei geringfügigen Änderungen in den Daten selbst ändert.

Der durch die Messung in das Modell gebrachten Unsicherheit wird innerhalb des Modelles durch die statistischen Verfahren zu begegnen gesucht. Die Ansätze reichen dabei von der klassischen Fehler- und Ausgleichsrechnung (Gauß, Laplace, Legendre) bis zu den heute üblichen Schätzverfahren der mathematischen Statistik. Die Grundfragen sind dieselben geblieben: überschüssige Beobachtungen (mehr Beobachtungen als zu schätzende Parameter), regelmäßige (systematische) und zufällige "Fehler" (Abweichungen vom Modellwert), Begriff der Meßgenauigkeit (Streuungsmaße), Fehlerfortpflanzung (Bestimmung des Fehlers einer Funktion, wenn die Fehler der einzelnen Variablen bekannt sind).

In der sozialwissenschaftlichen Meßtheorie und in der Praxis der Auswahl statistischer Verfahren spielt seit langem das Skalenniveau (metrische, ordinale, qualitative Skalen) eine bedeutsame Rolle, während in den Naturwissenschaften die metrischen Skalen vorherrschen und eine Anwendung schwächerer Meßniveaus kaum interessant erscheint. Jedes statistische Modell muß jedoch mehr (z. B. LISREL) oder weniger (normalverteilte Meßdaten, kein besonderes Meßmodell) das Meßproblem berücksichtigen.

5 Kausale Modelle

Kausales Denken gilt oft als die wissenschaftliche Denkform schlechweg. Legt man den Werkzeugcharakter des Denkens zugrunde, so muß sich Kausalität als Denkform im Erkenntnisprozeß bewähren, was gerade im geschichtswissenschaftlichen Bereich nicht immer der Fall zu sein scheint.

Kausalität ist einmal Ziel-Mittel-Denken (technischer Kausalitätsbegriff), dann aber auch ein Interpretationsbegriff wie Zufall und Notwendigkeit (historischer Kausalitätsbegriff).

Kausalität stellt eine lineare Ordnung von Ereignissen (eine "Kette") dar. Dies ist der formal faßbare Aspekt; in der Ordnung liegt noch eine Art von "Notwendigkeit". Bereits allgemeiner ist der Begriff der partiellen Ordnung von Ereignissen (partially ordered sets, "posets"). Einzelne partiell geordnete Ereignismengen bilden einen Verband (z. B. Mengenverband). Andere Ereignisverbindungen stellen kybernetische Systeme dar (Regelkreise, Automaten). Der einfachen Ereignisordnung der Kausalität entspricht ein funktionaler Zusammenhang ("Causal Chains", WOLD 1954) der Art

$$y = Ax,$$

$x = (x_1 \dots x_n)$ Ursachen, $y = (y_1 \dots y_m)$, Wirkungen,

A eine untere Dreiecksmatrix.

Weitergehende Formalisierungen wurden von verschiedenen Seiten vorgeschlagen (MALINVAUD 1964; GORDESCH 1972).

Nun ist jede Denkform nur in einem bestimmten Erfahrungsbereich gültig, eine Übertragung auf einen umfassenderen Bereich nicht mehr zulässig. Klassische Beispiele sind der Begriff der Bewegung in der griechischen Philosophie: Jede Bewegung setzt eine andere Bewegung voraus, bis schließlich der "unbewegte Bewegter" einen ersten Anfang setzt. Das ist aber ein Widerspruch, oder heißt nur, daß das Bewegungsprinzip lediglich in einem eingeschränkten Bereich gültig ist. Ebenso wird allgemein Kausalität nur in einem bestimmten Erkenntnisbereich sinnvoll verwendet. Akausalität bedeutet dann allerdings nicht Ursachenlosigkeit, sondern nur die Nichtanwendbarkeit der Denkkategorie Kausalität.

So ist der Funke des Feuersteins die physikalische Ursache der Entzündung des Pulvers in einer Muskete. Es ist aber nicht mehr oder zumindest nicht ohne eingehende Prüfung aussagbar, ob die Entwicklung der Astrono-

mie, Kartographie und dergleichen im 16. Jahrhundert Schifffahrt und Seehandel gefördert hat, oder Schifffahrt und Seehandel die Wissenschaft. (Die umgekehrte Erklärungsrichtung müßte ausdrücklich falsifiziert werden). Kausalität ist somit keine globale Eigenschaft einer Ereignismenge, sondern eine lokale, das heißt, eine Beziehung, die genügend genau bestimmten Einzelergebnissen zukommen kann. Solche in einer Kausalitätsbeziehung stehenden Ereignisse können dann zu größeren Systemen wie Regelkreisen (Rückkopplung!), Automaten und dergleichen zusammengefügt werden. Diese übergeordneten Systeme sind realwissenschaftlich (also historisch) genau so wirklich oder unwirklich (reine Denkformen) wie die Kausalität, jedoch für einen größeren Bereich von Erscheinungen anwendbar. Fehlt diese Präzisierung, so werden die gemachten Behauptungen weder wahr noch falsch; sie stellen keine Aussagen (im Sinne der Logik) dar, sondern bestenfalls Aussageformen, die noch die Möglichkeit besitzen, zu Aussagen zu werden.

Wie bei der Prognose sucht man dem Problem der nicht immer mit Sicherheit feststellbaren Kausalbeziehung in altbewährter Weise durch eine Stochastisierung Herr zu werden. Man spricht dann nicht mehr davon, daß "A die Ursache der Wirkung B" sei, sondern nur noch davon, daß "mit der Wahrscheinlichkeit p A als die Ursache von B" angesehen werden könne (vgl. dazu SUPPES 1970). Dieser Ansatz läßt sich nochmals verallgemeinern, indem man den Wahrscheinlichkeitsbegriff durch den inhaltlich unbestimmteren, jedoch weiterreichenden Begriff der Fuzzyness ersetzt (Die Fuzzy Set-Theorie ersetzt im großen und ganzen den additiven Mengenverband in der Definition der Wahrscheinlichkeit durch einen subadditiven: Die Wahrscheinlichkeit der Oderverbindung zweier sich ausschließender Ereignisse ist nicht mehr gleich der Summe der Einzelwahrscheinlichkeiten, sondern kleiner oder gleich dieser Summe).

Das Denken in Modellen kommt einerseits dem kausalen Denken entgegen, da es sich auf einen recht kleinen Erfahrungsbereich beschränken läßt, was aber das Denken in größeren Systemen natürlich nicht ausschließt. Die mathematische Formulierung von Kausalität führt jedoch zu größten Schwierigkeiten, da eigentlich immer nur eine funktionale Beziehung zwischen Ursachen und Wirkungen sowie eine Ordnungsbeziehung (die Ursache kommt vor der Wirkung) erfaßt werden. Was aber diese sonderbare "Notwendigkeit" des Vorher-Nachher ausmacht, entgleitet weitgehend. So verfallen eigentlich alle, noch so kunstvoll ausgebildeten kausalen Modelle dem klassischen Fehlschluß des "post hoc, ergo propter hoc".

Diesem Dilemma entkommt man bei der Modellbildung nur, indem man entweder auf das kausale Denken verzichtet und beispielsweise in vernetzten Systemen denkt, oder indem man von außerhalb des Modelles weiß, welche Beziehung eine kausale und welche eine rein funktionale ist. Woher aber dieses Wissen außerhalb des schon genügend Schwierigkeiten bereitenden Modelles kommt, bleibt offen.

Abschließend sei noch auf die Problematik des Kausalitätsbegriffes bei nichtlinearen Modellen hingewiesen, wo das "schwache Kausalitätsprinzip" (gleiche Ursachen, gleiche Wirkungen) durch das "starke Kausalitätsprinzip" (ähnliche Ursachen, ähnliche Wirkungen) ersetzt werden muß, wenn es überhaupt noch anwendbar sein soll (vgl. die Abschnitte 12 und 13).

5.1 Stufen des Modellbaus

Der Aufbau eines kausalen Modells erfolgt in folgenden Schritten:

- (1) Phänomenologische (Verbale) Analyse
- (2) Aufstellen der Zusammenhangsmatrix bzw. des Zusammenhangsgraphen
- (3) Bewertung des Graphen (z. B. durch Korrelationskoeffizienten oder Kovarianzen)
- (4) Aufstellen des linearen Gleichungssystems
- (5) Elimination überflüssiger Variablen und Relationen
- (6) Prüfung des linearen Gleichungssystems auf Identifiziertheit und Überbestimmtheit
- (7) Treffen der für die Identifikation nötigen Annahmen
- (8) Bestimmung der Parameter des Modells, zumeist durch statistische Schätzverfahren oder auch nur einfache Ausgleichung
- (9) Überprüfung der Güte des Modells durch Rückeinsetzen und statistische Tests
- (10) Modifikation des Modells (z. B. Ausscheiden von Variablen mit zu kleinen Koeffizienten)
- (11) Wiederholung ab Schritt (6), bis eine genügend gute Anpassung erreicht ist

(12) Gegebenenfalls Normierung der Koeffizienten des linearen Gleichungssystems

6 Normative Aspekte

Unser tägliches Leben ist voll von Wertungen aller Art, aber auch tief und bisweilen unkontrolliert dringen Wertungen in die Wissenschaft ein. Man könnte F. C. Schiller zitieren: "Expellas hominem, logica, tamenusque recurrent!" Es ist jedoch schwierig und bisher nicht zufriedenstellend gelöst, Werturteile (Stellungnahmen zu einem Sachverhalt im Gegensatz zur Information über einen Sachverhalt) in formale Modelle einzubauen.

Einfacher wird es, wenn Wertungen Gegenstand der Forschung sind. Normen z. B. kann man im Sinne einer deontischen Logik (vgl. VON WRIGHT 1968 und WEINBERGER 1970) als konditionale Imperative auffassen, also als Regeln der Art "Falls die Bedingung A erfüllt ist, dann soll die Handlung B ausgeführt werden", oder in Zeichen,

$$A \rightarrow !B.$$

Der Pfeil " \rightarrow " versinnbildlicht dabei den konditionalen Imperativ, der als Vorderglied eine Aussage (Bedingung A), als Folglied einen Sollenssatz (!B) aufweist.

Zur Beschreibung der Normen einer Kultur oder Subkultur z. B. kann man alle in Frage kommenden kategorialen Imperative auflisten. Soweit A oder B Relations- oder Gleichungssysteme darstellen, werden die Parameter theoretisch vorgegeben (die Normen lassen sich eindeutig bestimmen) oder statistisch geschätzt (die Normen lassen sich nur mit einer gewissen Unsicherheit ermitteln). Die Normen bilden dann einen gesonderten Teil des Modells und werden von den einfachen Aussagen deutlich getrennt.

7 Sinnverstehen und Modellmethode

Der Modellansatz ist ideologisch weitaus neutraler als man gemeinhin annimmt. Selbst wenn man sich auf den extremen Standpunkt einer verstehenden Geisteswissenschaft stellt, läßt sich die Modelltechnik gebrauchen.

Verstehen ist nach der Lexikondefinition "... das Erfassen von Bedeutungen und Sinngehalten, wobei der ihnen selbst eigene Sinn- und Bedeutungsraum (Horizont), dem sie entstammen, mit in den Blick gerückt wird...". Es wird als Gegensatz zum Erklären gesehen, wobei "... das zu Erklärende nur

als "Fall" begriffen und allein seine Einfügbarkeit in einen Gesamtzusammenhang hin ..." betrachtet wird (beide Zitate nach dem Kleinen Philosophischen Wörterbuch).

"Sinn" kann man in diesem Zusammenhang - in etwa BRUGGER 1976 folgend - auffassen als teleologischen Sinn (finale Betrachtung), Gestaltsinn (funktionale Betrachtung, Einordenbarkeit in ein übergeordnetes Ganzes), metaphysischen Sinn (Bedeutung größerer Ganzheiten, insbesondere des Seins, für das Leben der in dem jeweiligen Ganzen stehenden Menschen).

Verstehen in der Bedeutung des Erfassens des metaphysischen Sinns fällt wohl aus der heute üblichen Wissenschaft heraus und muß hier außer Betracht bleiben. Mit dem Erklären nach dem teleologischen Sinn und dem Gestaltsinn scheint es mir aber so zu sein wie mit der Kausalität. Sinn ist hier ebenso Interpretationsbegriff wie Kausalität. Auch hier lassen sich gewisse formale Kriterien finden. Kausale Modelle berücksichtigen nur die zeitliche Ordnung, nicht die Notwendigkeit usf. der Kausalbeziehung. Finalität weist genauso eine zeitliche Ordnung der Ereignisse auf, umfaßt aber mehr als bloß diese Ordnung (was gerade Kausalität und Finalität unterscheidet). Da der gleiche formale Aspekt in den statistischen Modellen erfaßt wird, lassen sich alle gängigen kausalen Modelle genauso gut als finale Modelle verwenden. Die Probleme bleiben haargenau dieselben, vor allem, daß der Wissenschaftler außerhalb des statistischen Modells feststellen muß, ob eine Beziehung eine kausale bzw. finale ist oder nicht. Wiederum lassen sich Schlüsse über einen engen Bereich hinaus nicht ziehen und führen rasch zu logisch widersprüchlichen Aussagen, mit denen sich dann eine dem metaphysischen Sinn gewidmete Philosophie abquält.

8 Statistisches Schließen

Wissenschaft strebt nach möglichst weitreichenden oder allgemein gültigen Aussagen. In der heutigen Statistik wird diese Aufgabe als Schluß ("statistical inference") von einer tatsächlich gezogenen Stichprobe auf eine zumeist hypothetische Gesamtheit ("one-sample problem") oder eine weitere, noch zu ziehende Stichprobe ("two-sample problem") gesehen. Die Fragen sind keineswegs geklärt und recht heikel; hier soll nur versucht werden, durch einen Blick zurück auf die Entstehung der Wahrscheinlichkeitsrechnung und Statistik die Kernfragen zu verdeutlichen und die zwei für Historiker wichtigsten Ansätze zu skizzieren.

Die Wahrscheinlichkeitsrechnung hat ihren Ursprung in der Glücksspieltheorie, der Versicherungsmathematik des Quattrocento und - etwas später - in den Naturwissenschaften, vor allem in der Meßtheorie ("Fehlerrechnung"; vgl. HARTWIG 1968). Dem schließt sich die amtliche Statistik (Wirtschafts- und Sozialstatistik) an. Bis in unser Jahrhundert hinein waren Mathematik und Naturwissenschaft in weiten Bereichen eine Einheit; Mathematik war genauso eine Methode, zu Erkenntnissen zu gelangen, wie Beobachtung und Experiment. Würfeln war ein physikalischer Vorgang, und Wahrscheinlichkeit die Eigenschaft eines physikalischen Systems. Wahrscheinlichkeitsrechnung ("Statistik" bedeutete lange Zeit ausschließlich "amtliche Statistik") war genauso mathematisches Modellieren wie das Aufstellen der Grundgleichungen der Newtonschen Mechanik. Das Glanzstück der Wahrscheinlichkeitsrechnung war während der ersten Hälfte des 19. Jahrhunderts die Fehlerrechnung. Dabei herrschte die Vorstellung, daß es einen wahren Wert gäbe, aber auf Grund der Unvollkommenheiten bei der Messung es zu Abweichungen käme, weswegen die Ergebnisse in einem gewissen Widerspruch zu einander stünden (die Erdbeschleunigung z. B. könne nicht mehrere Werte zugleich annehmen) und die daher "ausgeglichen" werden müßten. Man unterschied zunächst zwischen "groben" und "unvermeidbaren" ("kleinen") Fehlern, dann noch zwischen systematischen ("regelmäßigen", d. h. immer wieder auftretenden Fehlern) und zufälligen Fehlern, die Gegenstand der Ausgleichsrechnung waren. Die Rechtfertigung für die "Ausgleichung der Fehler" lieferten Meßmodelle, die sich im Ansatz bereits 1757 bei Thomas Simpson finden und bald darauf die Spitzen der Mathematik beschäftigten (Lagrange, Daniel Bernoulli, Legendre und vor allem Gauß, aber auch zahlreiche heute mehr oder minder vergessene Mathematiker und Naturwissenschaftler). Dabei wurde anstelle des nicht erreichbaren "wahren" Wertes ein "bester" Wert errechnet. Als "bester Wert" wurden der "plausibelste", das ist der wahrscheinlichste, bzw. der "am wenigsten schädliche", das ist der eine (zumeist quadratische) Fehlerfunktion minimierende Wert angesehen. Damit waren die auch heute noch gültigen Ansätze der Maximum Likelihood-Schätzung und der Entscheidungstheorie gefunden. Etwa zur gleichen Zeit wie Simpson hatte Bayes einen völlig anderen Weg zum Meßproblem und zum statistischen Schließen beschritten, womit der letzte bedeutende Ansatz gefunden war (Bayes-Statistik). Ein gesondertes Inferenzproblem tauchte hierbei gar nicht auf, zielte ja die ganze Fehlerrechnung auf den "wahren Wert" und war obendrein Teil von Physik, Astronomie oder Geodäsie (es handelt sich allenfalls um einen Sonderfall des Induktionsproblems).

Das zwanzigste Jahrhundert brachte ein immer stärkeres Auseinanderklaffen von Mathematik und realitätsbezogener Wahrscheinlichkeitsrechnung und, wie man jetzt auch für die mathematisierte Disziplin sagte, Statistik. 1933 veröffentlichte Kolmogorov sein Axiomensystem der Wahrscheinlichkeitsrechnung, welches das für die Mathematiker so wichtige Lebesgueintegral zu verwenden gestattete und die Wahrscheinlichkeitsrechnung zu einer Teildisziplin der Analysis machte. Zwei der damals bedeutendsten Theoretiker der Wahrscheinlichkeitsrechnung, Hans Reichenbach und von Mises, emigrierten mit dem Hochkommen des Nationalsozialismus und wandten sich in ihrer neuen Heimat völlig anderen Themen zu. Von der Statistik her gesehen, wurde Ende der Zwanziger- und Anfang der Dreißigerjahre ein Abschluß gesetzt: R.A. Fisher und Neyman und Pearson (letztere beide gingen von der statistischen Qualitätskontrolle aus) fixierten den heute klassischen Ansatz statistischen Schätzens und Schließens, so wie man ihn im Elementarunterricht lernt mit den Begriffen Fehler 1. und 2. Art, Konfidenzintervall, Hypothese und Alternativhypothese usf.

Man hatte bald die Frage nach dem "wahren Wert" als Scheinfrage fallen lassen, entweder weil man darin einen Zirkel sah, oder weil man den Schritt in die Metaphysik ablehnte. Es blieb aber die Kernfrage, nämlich die der Population. Bei der Fehlerrechnung konstruiert man gemäß dem Modell eine unendliche Population aus den als beliebig oft und ohne Änderung der Voraussetzungen wiederholbar vorausgesetzten Messungen. Die Messungen bilden eine Zufallsfolge, die in bestimmtem Sinne als konvergent angenommen wird. Nichtsdestotrotz wirken die Postulate haarsträubend: unbeschränkt oft wiederholbare Messungen (d. h. unendlich viele), dazu noch bei gleichbleibenden äußeren Bedingungen, und beliebig genaue Messungen (was schon physikalisch unmöglich ist, da längstens die Wärmebewegung der Moleküle die Zeigerablesungen unmöglich machen, im Mikrokosmos keinesfalls Stetigkeit herrscht, und die Lichtgeschwindigkeit als höchste Signalgeschwindigkeit nicht überschritten werden kann). Nicht umsonst witzelt man im Englischen über die unendlichen Zufallsfolgen: "In the long run we are all dead."

Weniger schwerwiegend sind die Voraussetzungen in der klassischen Stichprobentheorie, wie sie in der Umfrageforschung zugrunde gelegt wird: Aus endlichen Populationen werden endliche Stichproben gezogen, so daß wenigstens die beliebig feine Meßbarkeit nicht angenommen werden muß. Die unbeschränkte Wiederholbarkeit unter sich nicht ändernden Bedingungen muß jedoch ebenfalls postuliert werden. Noch bei GEBELEIN 1943 findet sich eine solche Theorie statistischen Schließens. Später, unter dem

Einfluß einer übermächtigen Mathematik, werden zwar noch endliche Stichproben behandelt, indessen erscheinen unendliche Folgen als etwas Selbstverständliches und werden wegen der Reichhaltigkeit und Eleganz der Theorie vorgezogen. Gleichwohl tritt ein neues Problem hinzu: Bei der Fehlerrechnung war die Konstruktion der Population klar vorgegeben, bei dem üblichen Ansatz der Stichprobenerhebung ist sie es nicht:

Betrachten wir als Beispiel die Frage: "Hat der Zentralismus des römischen Reiches normierend auf das Lateinische eingewirkt?". Offenkundig ist "Latein" kein fest umrissener Begriff, und die Sprache wandelt sich im Laufe der Zeit, je nach der Region (Rom - Latium - Italien - außeritalische Provinzen), den Sprechern (oder Schreibern) und den äußeren Umständen. Leider aber ändert sich die Sprache auch mit den Quellen: Inschriften, römische Literatur, Eigennamen, Lehnwörter, Angaben der römischen Grammatiker und Lexikographen. Zum Teil sind in den Quellen Fehler enthalten (wie manche etymologische Herleitungen), die ausgeschieden werden müssen. Manche Quellen sind für statistische Verfahren weniger geeignet wie Inschriften, die wenig Sprachmaterial bieten und deren Auswahl in unkontrollierter Weise zustande gekommen ist, wenngleich ihre Zahl oft nicht gering ist (für das Faliskische z. B. stehen rund 150 Inschriften zur Verfügung; vgl. STOLZ 1966). Angaben der römischen Grammatiker muß man als Untersuchungsergebnis hinnehmen, ohne das Material dazu zu erhalten. Recht gut eignet sich die klassische römische Literatur, die man als maschinenlesbares Corpus zusammenstellen und am Rechner nach diffizilen Kriterien untersuchen kann. Doch trotz der günstigen Datenlage tauchen ohne Ende Schwierigkeiten auf: Kann man die Sprache von Plautus, Tacitus und Augustinus (vielleicht ein extremes Beispiel) zu einem "Lateinischen" zusammenfassen? Oder die Fülle der Probleme, die sich aus der Abkehr von einer normativen Linguistik ergeben - was "Sprache" ist, beginnt zu zerfließen. Die verschiedenen Definitionen oder Operationalisierungen dessen, was "Latein" ist, führen statistisch gesehen zu verschiedenen Populationen und verschiedenen Stichproben, wodurch gleichbleibende Ergebnisse nicht garantiert werden können. Dazu kommen Detailprobleme wie die der Orthographie, Unvollständigkeit und dergleichen mehr.

Wie aus dem Dilemma herausfinden? Der einfachste Ausweg besteht darin, auf den Schluß auf die Gesamtheit überhaupt zu verzichten und sich auf die beschreibende Statistik zu beschränken. Nicht von ungefähr haben die "Analyse des données" (Benzécri und andere in der frankophonen Welt) und die "Data Analysis" als moderne Formen der beschreibenden Statistik den

"Probabilisten" einiges Terrain abgewonnen. Andererseits ist die Begriffsbildung oft nur aus dem probabilistischen Kontext verständlich, und die Beschränkung auf das untersuchte Material mag im Falle einer aus den "Fehlern" konstruierten Population einer Gesamterhebung der zu untersuchenden Einzelfälle angebracht sein, läuft aber sonst den wissenschaftlichen Intentionen zuwider und ist allenfalls in einer explorativen Phase ("exploratory techniques") zu vertreten.

Ein Teil der Statistiker wird auf verschiedene Varianten der angeführten Ansätze verweisen oder weitere Methoden anführen, die im deutschen Sprachraum so wenig bekannt sind, daß ich die englischen Bezeichnungen belasse (BARNET 1973): fiducial inference (R. A. Fisher), likelihood inference, structural inference (D. A. Fraser), information-theoretical approach. Der Grundgedanke des strukturellen Ansatzes von Fraser besteht in einer Trennung von Meßmodell und beschreibendem Modell, wie es etwa bei LISREL geschieht. Das Meßmodell kann dabei bedeutend einfacher gehalten werden, und bekannte Meßmodelle können als modularer Baustein immer wieder verwendet werden. Da allerdings die beiden Modellteile nicht voneinander unabhängig sind, treten trotz alledem Schwierigkeiten auf ("Fehlerfortpflanzung").

Eine weitere Möglichkeit besteht in einer Änderung des Wahrscheinlichkeitsbegriffes. Dem Fehler- wie dem Umfragemodell entspricht die sogenannte Häufigkeitsinterpretation der Wahrscheinlichkeit (Wahrscheinlichkeit als Grenzwert einer Folge von relativen Häufigkeiten). Mehr Messungen ergeben einen genaueren Wert, ebenso unter gleichbleibenden Bedingungen wiederholte Umfragen (Mittelwerte von Mittelwerten geben bessere Ergebnisse; "biasfreie Schätzungen"). Die Schwierigkeiten mit diesem Begriff sind seit langem bekannt, und verschiedene Lösungsversuche wurden vorgeschlagen (STEGMÜLLER 1973). Zumeist läuft das auf den Begriff der "subjektiven Wahrscheinlichkeit" hinaus, in der Keynes'schen Definition "the degree of rational belief" (KEYNES 1962). Praktisch bleibt aber die Bestimmung dieser subjektiven Wahrscheinlichkeiten vage.

Unvermeidbar ist eine bis zu einem gewissen Grad willkürliche, wenngleich wohlüberlegte und genaue Festlegung (Operationalisierung) der Begriffe mit all ihren bekannten Schwierigkeiten, auch im Falle einer beschreibenden Statistik. Für eine Anwendung statistischer Methoden ist eine präzise Operationalisierung (und damit Konstruktion der Population) unabdingbar, will man nicht jeden Gewinn an Exaktheit und Sicherheit in sein Gegenteil verkehren. Gerade die Modellmethode impliziert eine solche Konstruktion.

Da die Voraussetzungen nicht logisch bewiesen oder empirisch verifiziert werden können, bleibt wie sonstwo in der Wissenschaft nur ein pragmatisches Kriterium übrig, nämlich das der Bewährung, wozu mindestens Verträglichkeit mit der empirisch gefaßten Wirklichkeit gehört. Die Fehler- und Ausgleichsrechnung hat sich in weiten Bereichen der Naturwissenschaft und Technik bewährt und ist der Ausgangspunkt für zahlreiche statistische Verfahren auch in den Sozial- und Wirtschaftswissenschaften geworden. In der Mathematik hat Felix Klein (KLEIN 1928), einer der bedeutendsten Mathematiker, die Analysis (und damit die Wahrscheinlichkeitsrechnung) nicht umsonst immer "Approximationsmathematik" genannt, und in philosophischer Überhöhung hat der zu Unrecht heute kaum bekannte Hans Vaihinger seine Philosophie der Fiktion entwickelt (VAIHINGER 1911), die auch den Kern der Statistik trifft.

9 Systemtheorie

Das Wort "System" leitet sich vom griechischen "τὸ σύστημα" her, was soviel wie "das Zusammengestellte, das Geordnete" bedeutet. Es weist damit schon auf den Kern des Systembegriffs und verwandter Begriffe hin, nämlich die Ganzheit, Strukturiertheit und Abgegrenztheit (Konturiertheit). Von der biologischen Seite - die Biologen zählen zu den nachdrücklichsten Vertretern der Systemtheorie - kommt noch der Aspekt der Wechselwirkung mit der Umgebung hinzu. Für den Historiker kann ein System die Gesamtheit aller dem Oberbegriff Kultur unterzuordnende Phänomene umfassen, es kann sich um einen Teilbereich wie Wirtschaft oder Recht handeln oder wie in der Archäologie um sämtliche an einer Siedlungsstätte anzutreffende Phänomene.

Die klassischen mathematischen Methoden der Systemtheorie sind allerdings an der Physik orientiert und gründen zumeist auf Differentialgleichungssystemen. Historisches Material erfüllt die Voraussetzungen dafür nicht oder kaum. Systemtheorie dient dann ausschließlich, um gewisse qualitative Beziehungen modellhaft nachzuvollziehen ("Systemdenken"). So könnten geringfügige genetische Änderungen von Gräsern prähistorische Menschen veranlaßt haben, mehr Gräser dieser Art einzusammeln und sie schließlich anzubauen. Dies führte im Laufe der Zeit zur Züchtung von Getreidesorten mit noch höheren Erträgen, zu Bevölkerungswachstum, abermals intensivierter Landwirtschaft usw. Bei der Aufstellung und Ausarbeitung solcher Gedanken wirken Konzepte aus der Systemtheorie als Leitlinie (wie im Beispiel oben Rückkopplung, Wechselwirkung von System und Umwelt und dergleichen mehr).

Es existieren aber auch mathematische Modelle, die das Systemdenken nicht nur ganz allgemein und vage, sondern recht präzise anzuwenden gestatten und doch nicht die Datenlage für kontinuierliche Systeme voraussetzen. Dazu zählen zunächst die Theorie der formalen Sprachen, Automatentheorie, Boolesche Funktionen, dann Differenzgleichungssysteme, die bereits sprunghafte Änderungen innerhalb eines gleichmäßig ablaufenden Prozesses zu modellieren gestatten.

Ein anderer Weg, um die fehlenden oder unzureichenden Daten für eine Systembetrachtung zu ersetzen, ist die Computersimulation. Sie ermöglicht die Untersuchung des Verhaltens von Kultursystemen unter dem Einfluß zeitbedingter wie auch zeitunabhängiger Faktoren. Typische Fragestellungen für eine Simulation sind: Verbreitung von Handelsgütern, Sprachen und dergleichen (Diffusionsmodelle); Auswirkungen von Veränderungen in der Umwelt auf Bevölkerungszahl, Nahrungsbeschaffung, Siedlungsmuster; die Entwicklung von Siedlungssystemen bei unterschiedlichem Bevölkerungswachstum und unter sich ändernden wirtschaftlichen, technologischen und ökologischen Bedingungen. Treffend läßt sich die Simulation als "computergestütztes Gedankenexperiment" charakterisieren. Die prinzipiellen Vorzüge und Mängel der Methode sind dann vom Gedankenexperiment abzulesen, nur daß eine Leistungssteigerung durch den Rechner eintritt.

10 Automatentheorie

10. 1 Automatentheorie als historische Metatheorie

Die Automatentheorie ist eine Metatheorie von Input-Output-Systemen mit Rückkopplung. Eine Metatheorie hat als Ziel eine Verallgemeinerung der Begriffe und Aussagen spezieller Theorien. Sie vereinigt mehrere bislang getrennte Theorien und sucht dadurch ein effizienteres Werkzeug für den Forscher zu liefern. Das grundlegende Denkschema entspricht dabei vollkommen dem wissenschaftlichen Alltagsdenken: Etwas wirkt ein ("Eingabe"), es entstehen Ergebnisse ("Ausgabe"), und dieses Wirken ist durch allgemeine Bedingungen ("Zustände") bestimmt. Diese Bedingungen werden durch die Einwirkungen verändert und das Ergebnis beeinflußt die ursprünglichen Einwirkungen (Rückkopplung).

In der Geschichtswissenschaft läßt sich häufig "Staat" als System auffassen. Die Zustände beschreiben dann das Staatswesen (z. B. Staatsgebiet, Bevölkerung, Wirtschaftslage, oder auch nur gute bzw. schlechte Lage). Einwirkungen (Einflüsse) können innere (politische Umschichtungen, Wahlen, ...) wie auch

äußere (Politik anderer Staaten) sein. Auswirkungen sind beispielsweise das innen- wie außenpolitische Geschehen, das naturgemäß die Zustände des Staates (z. B. sein politisches System) wie auch die Einwirkungen (Außenpolitik und dergleichen) beeinflussen.

Daß das Grundscheema der Automatentheorie realitätsgerecht ist, ist unzweifelhaft. Manche historische Darstellungen (vgl. PING 1976) lesen sich wie Übungsbücher zum automatentheoretischen Modellbau. Es fragt sich allein, ob die Einzelheiten der Theorie über das Grundlegende hinaus sich als nützlich erweisen oder nur eine scheinbare Wissenschaftlichkeit bringen. Die Antwort ist nicht ganz einfach. Sicherlich wird die übliche verbale Geschichtsschreibung nicht überflüssig. Wenn jedoch eine große Zahl von Einzelfakten berücksichtigt werden muß, entgleitet leicht die Übersicht. Simulationsmodelle erfordern das genaue Einhalten eines festen Schemas: In beiden Fällen lohnt die Formalisierung auch von qualitativen Sachverhalten. Unumgänglich wird die Formalisierung bei umfassenderen quantitativen Modellen, allerdings beschränkt die historische Datenlage solche Modelle wohl auf einen recht kleinen Zeitraum jüngerer Geschichte. Wieweit spezielle Begriffsbildungen wie Minimalautomat, Äquivalenz von Automaten und ähnliches sich als fruchtbar erweisen, hängt nicht zuletzt daran, ob die Historiker bereit sind, solche Begriffsbildungen in ihr Denken aufzunehmen.

10.2 Formale Definitionen

Morphismen (Abbildungen von Strukturen) beziehen zwei Systeme aufeinander (z.B. zwei Staaten):

$$A(X, Y, Z, \delta, \mu) \rightarrow A^*(X^*, Y^*, Z^*, \delta^*, \mu^*),$$

A System; X, Y, Z Mengen; δ, μ Abbildungen;

* bezeichnet die entsprechenden Größen im Bildbereich.

Halbautomaten sind Tripel $(X, Z; \delta)$ mit nichtleerem X, Z, und der Abbildung

$$\delta: X \times Z \rightarrow Z$$

(Entsprechend der Eingabe und dem zuvor herrschenden Zustand folgt der neue Zustand des Systems).

Automaten sind Quintupel (X, Y, Z, δ, μ) über einem nichtleeren Eingangsalphabet X , einem Ausgangsalphabet Y , einer Menge von Zuständen Z , und zwei Abbildungen

$$\delta: X \times Z \rightarrow Z,$$

$$\mu: X \times Z \rightarrow Y$$

(Je nach der Eingabe und dem Zustand erfolgen die Zustandsänderung und die Ausgabe).

Praktisch werden Automaten häufig durch zeitabhängige Differenzen- oder Differentialgleichungen dargestellt:

$$y(t) = f[x(t), z(t)],$$

$$z(t + 1) = g[x(t), z(t)].$$

Der Arbeitsökonomie und der Durchschaubarkeit der Modelle wegen sucht man nicht wesentlich verschiedene ("äquivalente") Zustände auszuscheiden:

Äquivalente' Zustände $z \equiv z^*$ sind solche, für die gilt:

$$\delta(z, x_1, \dots, x_n) = \delta^*(z^*, x_1, \dots, x_n)$$

für alle Folgen $(x_1, \dots, x_n) \in F_x$

(Äquivalente Zustände führen bei gleicher Eingabe zur gleichen Ausgabe).

Äquivalente Automaten $A(X, Y, Z, \delta, \mu)$, $A^*(X^*, Y^*, Z^*, \delta^*, \mu^*)$ sind charakterisiert durch die Bedingungen:

Für alle z gibt es ein z^* , so daß $z \equiv z^*$,
und für alle z^* gibt es ein z , so daß $z \equiv z^*$

(Zu jedem Zustand eines Automaten findet man einen äquivalenten Zustand im jeweils anderen Automaten).

Ein Minimalautomat ist ein Automat ohne äquivalente Zustände:

$$z \equiv z^* \rightarrow z = z^*.$$

10.3 Stochastische Automaten

Im allgemeinen ist das Material, auf Grund dessen der Automat bestimmt werden soll, unvollständig. Das heißt, die Abbildungen eines Automaten lie-

fern bei einer Eingabe die Resultate nur mehr mit einer gewissen Wahrscheinlichkeit. Dies berücksichtigt der Begriff des stochastischen Automaten.

$A = (X, Y, Z; p)$ heißt stochastischer Automat, wenn gilt:

- (1) X, Y, Z sind nichtleere Mengen,
 X Eingabealphabet,
 Y Ausgabealphabet,
 Z Zustandsmenge.
- (2) Für alle $x \in X$, für alle $z \in Z$:
 $p(\cdot, \cdot \mid x, z)$ ist ein bedingtes Wahrscheinlichkeitsmaß auf $Y \times Z$.

Dazu kommen entsprechende rekursive Definitionen der Fortsetzung von p auf die freien Monoide von X und Y , um eine sequentielle und synchrone Arbeitsweise zu beschreiben.

Diese Art von Automaten führt zwangsläufig auf Probleme der Parameterschätzung, da die den Halbautomaten oder Automaten kennzeichnenden Größen im allgemeinen aus Beobachtungen, die Realisierungen eines stochastischen Prozesses darstellen, bestimmt werden. Zahlreiche statistische Methoden sind zweckdienlich: Anpassung von Wahrscheinlichkeitsverteilungen, Analyse latenter Strukturen im Sinne von Lazarsfeld, das Rasch-Modell, Beschreibung stochastischer Automaten durch Regressionsgleichungen mit stochastischen Koeffizienten. Automatentheoretische Sätze lassen sich dabei unter anderem ähnlich wie die Faktorenanalyse zur Reduktion auf eine geringere Zahl von Variablen verwenden. Aussagen über Gleichheit, Äquivalenz usf. von Automaten führen zur Prüfung von für gewöhnlich multivariaten statistischen Hypothesen.

10.4 Regressionsmodell mit Zufallskoeffizienten

Die Anpassung von Verteilungen ist oftmals mit vielen Problemen behaftet. Es erscheint daher empfehlenswert, einen sequentiellen stochastischen Automaten über Regressionsmodelle mit Zufallskoeffizienten zu definieren:

$$y_t = \sum_k z_{tk} (\beta_k + v_{tk}),$$

$$1 \leq t \leq T, 1 \leq k \leq K, v_{tk} \text{ Zufallsvariable.}$$

Im einfachsten Fall ergibt sich das klassische Regressionsmodell

$$y_t = z_t \beta + u_t,$$

$$Eu_t = 0, Eu_t u_{t+k} = \text{Diagonalmatrix},$$

E Erwartungswert bezüglich t, k Lag.

Der allgemeine Fall ist heikler; Schätzverfahren sind z. B. bekannt für:

$$Ev_{tk} = 0, Ev_{tk}^2 = \alpha_k \text{ (unbekannt)},$$

$$Ev_{tk} v_{sj} = 0 \text{ für } t \neq s \text{ und } k \neq j,$$

E Erwartungswert bezüglich t bzw. t und s.

Es läßt sich eine Vielzahl von Schätzverfahren auf das Modell anwenden. Als Beispiel diene die Methode des Maximum Likelihood:

Maximiere die Likelihoodfunktion

$$L = -1/2 \sum_t \ln(\sum_k z_{tk}^2 \alpha_k)$$

$$-1/2 \sum_t (y_t - \sum_k z_{tk} \beta_k)^2 / \sum_k z_{tk}^2 \alpha_k,$$

v_{tk} verteilt nach $N(0, \alpha_k)$.

Unter anderem wurden verschiedene Varianten der Methode der Kleinsten Fehlerquadrate, das Verfahren nach Theil - Mennes (iteratives Verfahren mit einem Wechsel zwischen den v_{tk} und der Diagonale der Kovarianzmatrix der Residuen), Minimum Norm Quadratic Unbiased Estimation (MINQUE) vorgeschlagen (SWAMY 1970, FROEHLICH 1973). Viele Fragen sind noch offen geblieben, wie der Übersichtsartikel CHOW 1987 zeigt, wo sich weitere Literaturhinweise finden.

Zur Durchführung der Parameterschätzungen stehen ökonomische Softwarepakete zur Verfügung, u. a. TSP oder RATS, sofern man nicht eine eigene Programmierung (etwa in GAUSS) vorzieht.

11 Entscheidungstabellen

Entscheidungstabellen sind endliche Input-Output-Systeme und im Gegensatz zu Automaten ohne Zustände und ohne Rückkopplung. Da sie endlich sind, können sie in Tabellenform dargestellt werden. Sie sind gut geeignet, Wenn-Dann-Beziehungen wie Verhaltensregeln oder Gesetzmäßigkeiten darzustellen.

Nehmen wir als Beispiel die Regeln für die Staatskunst, wie sie der Erste Minister Guan Xing des Herzogtums Xi im 7. Jahrhundert formuliert hat. Dort finden sich Regeln wie:

Nur wenn Kleidung und Ernährung angemessen sind, können die Menschen Stolz und Scham empfinden.

Not liefert eine Entschuldigung für Verrat.

Dazu läßt sich leicht eine Tabelle aufstellen:

Guan Xings Regeln der Staatskunst

Bedingungen	Regeln			
	1	2	3	4
Kleidung angemessen	J	J	N	N
Ernährung angemessen	J	N	J	N
Stolz und Scham	*		*	
Verrat		*		*
Handlungen				

Mathematisch gesehen ist eine Entscheidungstabelle (decision table) im allgemeinsten Sinne eine tabellarische Darstellung einer Abbildung Φ einer endlichen Menge B ($\text{card } B = N$) von endlichen geordneten Mengen fester Kardinalzahl n in eine endliche Menge H ($\text{card } H = M$) von endlichen geordneten Mengen der gleichen oder einer anderen Kardinalzahl m . Die Elemente der geordneten Mengen gehören wiederum festen endlichen Mengen X und Y an, die eine algebraische Struktur tragen können.

$$\Phi: B \rightarrow H$$

$$B = \{(x_1^1 \dots x_n^1), (x_1^2 \dots x_n^2), \dots\}, x_i^k \in X, \text{card } B = N,$$

$$H = \{(y_1^1 \dots y_m^1), (y_1^2 \dots y_m^2), \dots\}, y_j^l \in Y, \text{card } H = M,$$

$$x_i^k \in X, y_j^l \in Y, \text{card } X \text{ und } \text{card } Y \text{ endlich.}$$

Für gewöhnlich schränkt man die Mengensysteme wie auch die Abbildung stark ein. Der einfachste Fall ist die klassische Entscheidungstabelle mit den Bedingungen

$$X = Y \in \{0,1\},$$

$$X, Y \text{ Galoisfelder der Charakteristik } 2.$$

Die Abbildung ist eine Abbildung "auf" und nicht "in" ("Boolesche Funktionen"). Die Urbildmenge B wird als Menge der Bedingungen (conditions), die Bildmenge H als Menge der Handlungen (actions) gedeutet. Dann sind die zwei standardmäßigen Darstellungen die obenstehende Tabelle und ihre Spiegelung (Vertauschung von Zeilen und Spalten). Durch Wiederholung der Elemente von H wird eine Gleichmächtigkeit von B und H erzielt ($\text{card B} = \text{card H}$).

Zwei solche Abbildungen Φ, Ω lassen sich zu einer dritten (Γ) zusammensetzen (Verknüpfung zweier Abbildungen):

$$\Gamma = \Omega \cdot \Phi.$$

Die Darstellung dieser Verknüpfungen kann wiederum in Tabellenform erfolgen; man spricht dann von Gitternetzen (oder nur von Gittern oder Netzen, grid charts). Eine graphische Darstellung sprengt allerdings rasch die normalen Papierformate und wird recht unübersichtlich.

Die Abbildungen und ihre Verknüpfungen lassen sich auch in Form von Graphen darstellen, bei größeren Systemen wird aber auch diese Form bald unbrauchbar.

Es liegt daher nahe, für größere und komplexere Probleme Computer einzusetzen. Es existiert dafür eine ausgefeilte und bequem zu benutzende Software, und für besondere Fälle läßt sich leicht in einer höheren Programmiersprache wie C oder Pascal ein effizientes Programm schreiben. Auch Kalkulationsprogramme (wie Boeing Calc oder die Version 3.x von Lotus 1-2-3 mit ihren dreidimensionalen Tabellen, PlanPerfect und andere Pakete, die eine Verbindung von Tabellen zulassen) sind recht nützlich.

12 Dynamische Modelle

Daß Geschichte mit zeitlichem Wandel zusammenhängt, ist trivial. Dennoch ist die statistische Untersuchung historischer Prozesse bei weitem nicht ihrer Bedeutung nach entwickelt. Am ehesten finden sich Verfahren zur Beschreibung stetiger Entwicklungsprozesse wie z.B. die Zeitreihenanalyse. Das bewußte statistische Modellieren findet sich allerdings allenfalls bis zu dem Punkt, daß die zum Funktionieren des statistischen Verfahrens notwendigen Voraussetzungen berücksichtigt werden. Inhaltliche Überlegungen müßten weitaus stärker bei der Aufstellung mathematischer Modelle einbezogen werden.

Eine deterministische Prozeßbeschreibung erfolgt am einfachsten durch ein Gleichungssystem. Die Erklärung von Prozessen erfolgt zumeist besser, wenn die Änderung explizit in das Modell einbezogen und erst nachträglich eine Lösung des Systems von Funktionalgleichungen gesucht wird. Da die Physik das große Vorbild für eine Mathematisierung darstellt, werden häufig Differentialgleichungen zur Beschreibung von Prozessen verwendet. In den Humanwissenschaften fallen die Daten jedoch zumeist in festen Zeitabständen an, weswegen Differenzgleichungen explizit und nicht nur im Zuge des numerischen Verfahrens zur Approximation verwendet werden.

Man kann in Analogie zu den Differentialgleichungen von den (endlichen) Änderungen ausgehen und z.B. schreiben:

$$\Delta w = f(t) \Delta t,$$

w Feldgröße, t Zeit,

Δ Änderungsoperator (Differenzenoperator),

f(t) zeitabhängiger Faktor, der die mit der zeitlichen Änderung Δt sich ergebenden Änderung Δw der Feldgröße bestimmt.

Eine andere Denk- und Schreibweise ersetzt Δw usf. durch die Differenz $w(t) - w(t-1)$ und nimmt die Zeiteinheiten konstant (meist gleich eins) an:

$$w(t) - w(t-1) - f(t) = 0,$$

oder bei Verwendung von Indizes statt der Funktionsschreibweise,

$$w_t - w_{t-1} - f_t = 0.$$

Man betrachtet die entsprechenden Größen zu einem bestimmten Zeitpunkt, eine Zeiteinheit davor, eine zweite Zeiteinheit davor usw. Dieser Ansatz wird im allgemeinen vorgezogen, sofern man nicht bewußt die Anlehnung an Differentialgleichungen sucht.

Dieser Ansatz ist einfach und einleuchtend, führt aber nichtsdestotrotz zu erheblichen mathematischen Problemen, da lediglich für lineare Differenzgleichungen mit konstanten Koeffizienten allgemeine Lösungsverfahren bekannt sind und in den übrigen Fällen gesonderte Untersuchungen erforderlich sind.

Schon folgendes Modell führt auf eine nichtlineare Differenzgleichung (die Werte der Feldgröße zu verschiedenen Zeitpunkten treten nicht nur in der nullten oder ersten Potenz auf):

Die Feldgröße wächst kontinuierlich mit der Zeit, doch nähert sie sich einem Grenzwert, so verlangsamt sich das Wachstum und verschwindet schließlich bei Erreichen des Grenzwertes.

$$b \cdot (w_t - w_{t-1}) - a \cdot w_t \cdot (b - w_t) = 0,$$

b Grenzwert, a Maßstabskonstante.

Erst recht aufwendig werden Lagmodelle. Entsprechend der Lagdifferentialgleichung (wo die Modellvorstellungen noch am deutlichsten zu sehen sind)

$$dw(t)/dt = (a/b) \cdot w(t-t_1) \cdot [b - w(t-t_2)]$$

läßt sich eine Differenzgleichung aufstellen:

$$[w(t) - w(t-1)] - (a/b) \cdot w(t-t_1) \cdot [b - w(t-t_2)] = 0,$$

t_1 unter günstigen Umständen bis zum Wachstum verstreichende Zeit,

t_2 Zeit, die zur Anpassung der Änderungsraten bei sich ändernden Bedingungen (Gewichtung!) verstreicht,

a, b wie oben.

Bei vielen Anwendungen werden mehrere Variablen und somit Gleichungssysteme betrachtet, was die Modelle natürlich nicht einfacher macht.

In diesen Modellen treten Parameter auf, die erst aus empirischem Material bestimmt werden müssen. Die klassische Ökonometrie hat dieses Problem bearbeitet, zuerst die linearen Modelle, in letzter Zeit besonders auch die nichtlinearen. Will man die Funktionalgleichungen lösen, um zu einem geschlossenen Ausdruck zu gelangen, so trifft man auf stochastische Funktionalgleichungen, und die Lösungen sind ebenso wie die Parameter Zufallsgrößen. Die Standardaufgabe, die Verteilungen der Lösungen aus den Verteilungen der Variablen zu bestimmen, ist recht kompliziert und oftmals gar nicht oder nur über Näherungen gelöst.

Bei nichtlinearen Differenzgleichungen und Differentialgleichungen tritt ein völlig neues Problem auf. Der lineare Zusammenhang heißt im Kern: Es geht immer so weiter wie bisher. Lineare Funktionalgleichungen haben zwar zumeist nichtlineare Lösungen, aber wenigstens ist die Änderung gleichmäßig, ohne plötzliche Sprünge. Und genau diese Garantie der Gleichmäßigkeit geht bei nichtlinearen Funktionalgleichungen verloren. Sie spielen aber in den heutigen Natur- und Geisteswissenschaften eine herausragende Rolle, weswegen ihnen ein eigener Abschnitt gewidmet sei.

13 Katastrophentheorie

Seit Menschengedenken finden sich Berichte über plötzliche, zumeist als zerstörerisch angesehene Veränderungen im Leben der Menschen. Aus vorgeschichtlicher Zeit sind Mythen und Sagen überliefert, das Alte Reich der Mayas wurde in der Mitte des 10. nachchristlichen Jahrhunderts aufgegeben, in der Antike finden sich Abhandlungen wie Platos Dialoge (vor allem Kritias mit der Atlantissage), um das Jahr Tausend sorgte ein angeblich bevorstehender Weltuntergang für Aufregung, und in der neuesten Zeit schreiben Autoren wie Däniken und Tons Brunés über (diesmal eher segensreiche) vergangene plötzliche Ereignisse, oder unsere Jugend ist voll Angst vor dem atomaren Weltuntergang. Demgegenüber weist die Wissenschaft erstaunliche Zurückhaltung auf. "Natura non facit saltus" galt auch bis vor kurzem in Physik oder Biologie, die Nationalökonomie bevorzugte Gleichgewichtsmodelle nach dem Vorbilde Walras' oder allenfalls Wachstumsmodelle nach Tinbergen und Nachfolger. In der Geschichtsforschung, zumal in der Urgeschichtsforschung, konnte man den Untergang ganzer Völker und Kulturen (Mykene, Mohendscho Daro, Mayas) in verhältnismäßig kurzer Zeit nicht übersehen. Waren bei den Urvätern der modernen Vorgeschichtsforschung im 18. und 19. Jahrhundert noch Katastrophen, die mit dem Untergang aller Lebewesen einhergingen, in Mode, so folgte bald ein strenger Aktualismus, der auch vergangene Veränderungen der Erdoberfläche und des Lebens auf ihr aus heute genauso ablaufenden natürlichen Prozessen zu erklären suchte, und möglichst nicht aus einer gleichmäßigen Entwicklung herausfallen wollte (vergleiche z. B. das Vorgehen der "prozessualen Archäologie", BINFORD 1963). Wie wir im weiteren sehen werden (mathematische Katastrophentheorie), schließen sich die beiden Ansätze nicht nur nicht aus, sondern bilden sogar eine Einheit.

Gehen wir zunächst aus von der archäologischen Methode der magnetischen Sondierung, wo bestimmte Objekte wie z. B. Herde und Brennöfen, gewisse Elemente von Siedlungsanlagen wie Gräben oder Gegenstände aus Eisen lokale Störungen des Magnetfeldes hervorrufen. Ein einfacheres Differentialgleichungssystem, das den magnetischen Feldgleichungen ähnlich ist, wurde von RÖSSLER (1976) studiert und sei zur Demonstration der methodischen Probleme hier skizziert.

Das Differentialgleichungssystem sei in folgender Form gegeben:

$$dx/dt = - (y + z),$$

$$dy/dt = x + ay,$$

$$dz/dt = b + z(x - c),$$

x, y, z Komponenten der magnetischen Feldstärke,

a, b, c Konstanten.

In diesem System sind die beiden ersten Gleichungen linear, in der dritten Gleichung ist ein Glied zweiter Ordnung enthalten. Wählen wir a und b fest und variieren wir c (Bifurkationsparameter, die für die sprunghaften Änderungen verantwortliche Systemgröße). In einem gewissen Bereich des Parameters c können kleine Änderungen ein völlig verschiedenes Verhalten der Lösungen hervorrufen. Angenommen, c sei eine Materialkonstante, die z. B. für das umgebende Erdreich und den Kessel aus Eisen verschieden ist. Bei einer stetigen Lösungskurve müssen die Materialkonstanten stark voneinander abweichen, um eine größere Änderung in der magnetischen Feldstärke zu erzielen, und die Methode versagt bei Substanzen mit etwa gleichen Materialkonstanten. Demgegenüber würde aber gerade in diesem Fall eine instabile Lösung des Gleichungssystems ein brauchbares Verfahren garantieren und trotz der kleinen Unterschiede in den Materialkonstanten ein deutlich verschiedenes Bild der magnetischen Feldlinien ergeben.

An dem Beispiel wird Verschiedenes deutlich. Zunächst fällt die Komplexität der Eigenschaften der Lösungen bei schon ganz einfachen Systemen auf (Einfach heißt dabei einfacher funktionaler Zusammenhang und wenige Variable, kaum je mehr als drei oder vier). Dann wie natürlich beide, kontinuierliche wie sprunghafte, Veränderungen auftreten. Es mutet eher wie eine unerklärliche Blindheit an, daß seit Newton in der Physik fast ausschließlich stetige Lösungen und Gleichgewichtszustände gesucht wurden und es eine Sache der letzten Jahrzehnte ist, gerade unstetige Lösungen und Zustände fern vom Gleichgewicht zu untersuchen. Letztlich, woher kommt die Kenntnis der Parameter? In der Tat gibt es so gut wie keine passende statistische Theorie. Wohl existieren zahlreiche Verfahren zur Parameterschätzung in nicht-linearen Modellen (ein moderner Forschungsschwerpunkt der Ökonometrie), allein die Stichprobenverteilungen der geschätzten Parameter sind unbekannt oder wenigstens aus den vorhandenen Daten schlecht oder gar nicht zu bestimmen. Aber hat eine Statistik, die auf einer beliebigen Wiederholbarkeit ("in the long run") beruht, überhaupt einen Sinn bei einmaligen und plötzlich auftretenden Ereignissen? Zumindest muß wohl eine Uminterpretation des Wahrscheinlichkeitsbegriffs stattfinden (vgl. Abschnitt 8).

Die Katastrophentheorie im engeren Sinne wurde von René Thom (vgl. THOM 1972; eine gute Einführung bei ARNOLD 1986) entwickelt. Sie ist eine Art qualitativer Theorie spezieller Differentialgleichungssysteme (sogeannter Gradientensysteme):

$$dx(t)/dt = - \text{grad } V(x(t);u(t);t),$$

t Zeit, x Zustandsvariablen, u Eingabevariablen,

V Potentialfunktion,

$$f_i = \text{grad } V = \partial V / \partial x_i,$$

$$\text{rot grad } V = \partial f_i / \partial x_j - \partial f_j / \partial x_i.$$

Dieser Ansatz entspringt dem klassischen Grundkonzept dynamischer Systeme (Zustände, Input, Veränderung der Zustände als Output) und ist mathematisch halbwegs einfach zu handhaben (Vektordifferentialgleichung erster Ordnung und Existenz einer Potentialfunktion). Damit eine Potentialfunktion existiert, muß

$$\text{rot grad } V = 0$$

sein (das Feld muß wirbelfrei sein, wie die Physiker sagen), eine recht spezielle Annahme. Es überrascht, daß überhaupt noch ein praktisch genügend flexibles Modell herauskommt (z.B. Fokker-Planck-Gleichung für diffusionsartige Prozesse). Die Katastrophentheorie stützt sich im weiteren auf die Untersuchung der Potentialfunktion, die nicht nur mathematisch vorhanden, sondern auch realwissenschaftlich deutbar sein muß: Die Potentialfunktion verbindet zeitliche Änderungen der Zustandsgrößen mit Änderungen der Inputgrößen bezüglich der Zustandsgrößen. Gesucht werden nun die stationären Lösungen

$$\text{grad } V = 0$$

(Maxima, Minima, Sattelpunkte). Minima führen auf stabile Zustände. Wenn zusätzlich die Hessesche Determinante (Determinante aus den zweiten Ableitungen der Potentialfunktion) verschwindet, ist das System nicht stabil: Bei kleinen Änderungen der Potentialfunktion ändern sich die Lösungen (Trajektorien) beträchtlich. Thom hat die "Katastrophen" für Systeme mit weniger als fünf Zustandsvariablen und bis auf Diffeomorphismen klassifiziert (die sieben elementaren Katastrophen).

Diese Modelle wurden vorwiegend in den Naturwissenschaften entwickelt (Physik, physikalische Chemie, Biologie), und es überrascht, daß sie eine solche Verbreitung gefunden haben. Das mag sich zum Teil aus dem Prestige

der Naturwissenschaften erklären, vorwiegend dürften es die bisweilen nur dem Namen (Katastrophentheorie, Chaostheorie) nach vermuteten Ansätze zur Bewältigung gewaltsamer, oftmals zerstörerischer Prozesse sein. Manchmal fällt es in den Geisteswissenschaften schon schwer, unter den engen Voraussetzungen überhaupt sinnvolle Modelle zu entwickeln: Potentialfunktion (Existenz einer Art Generalfunktion, die alles erklärt); Beschränkung auf so wenige Variable; Meßniveau und hohe Meßgenauigkeit. Am leichtesten geht das, wo überhaupt naturwissenschaftliche Methoden (z.B. magnetische Sondierung bei Ausgrabungen) eingesetzt werden oder wenigstens Modelle nach naturwissenschaftlichen Vorbildern (Diffusion, Distribution) geschaffen wurden.

Der Hauptnutzen scheint jedoch gar nicht aus dem konkreten Einsatz dieser Modelle zu erwachsen, sondern aus allgemeinen Folgerungen. Ilya Prigogine (PRIGOGINE - STENGERS 1986) betont das Näherrücken von Geisteswissenschaften (er meint zumeist Geschichte) und Naturwissenschaften, diesmal sogar umgekehrt, die Physik entdeckt die Irreversibilität der Zeit, wie sie die Historiker seit eh und je kennen. Am bedeutsamsten scheint mir jedoch bisher zu sein, daß in den Geisteswissenschaften instabile, vom Gleichgewicht entfernte Prozesse modellhaft, aber mit aller Exaktheit der Mathematik studiert und diskutiert werden können. Die Einsichten in solche Prozesse mögen intuitiv gefunden worden sein, sie werden jetzt rational faßbar und damit auch eher lehr- und lernbar. Für den Statistiker stellen diese Modelle eine ungeheure Herausforderung dar, sich aufzumachen von der überflüssigen Vervielfachung statistischer Tests und Schätzverfahren zur Eroberung wahren Neulands.

14 Datenverarbeitungsaspekte

Kleine Modelle kann man mit Papier und Bleistift und allenfalls mit einem Taschenrechner (die besseren verfügen bereits über Statistik- und Numerikfunktionen, bisweilen finden sich sogar Matrizenrechnung und algebraische Umformungen) angehen. Für gewöhnlich ist jedoch der Einsatz eines größeren Rechners notwendig. Die Vielzahl der Benutzer verwendet für anspruchsvollere Modelle spezialisierte Programmpakete, die einzelne Verfahren oder Verfahrensklassen beinhalten wie z. B. PHASER (KOÇAK 1989) für nichtlineare Modelle. Im Idealfall erhält man zugleich eine KI-Unterstützung mitgeliefert wie etwa bei GLIMPSE, einer Fortentwicklung des Statistikpakets GLIM (Generalized Linear Models, verallgemeinerte lineare Modelle). Dieses Vorgehen ist zunächst sehr bequem und bringt rasch Erfolge, weist aller-

dings mehrere Nachteile auf. Bleibt man nicht bei einer Modellklasse, so muß man andauernd die Pakete wechseln, was kostspielig ist und das Erlernen immer neuer Systeme bedeutet. Es stehen auch nur bestimmte Verfahren zur Verfügung, und häufig besteht keine Kontrolle darüber, welche algorithmischen Details programmiert wurden oder wie die Rechnung abläuft.

Wer sich mehr mit der statistischen Modellierung beschäftigt, muß daher mindestens zur Ergänzung flexiblere Werkzeuge wählen. In Frage kommt zunächst die Programmierung in einer höheren Programmiersprache (heute am ehesten in C oder Pascal, allenfalls in Fortran) und die Verwendung fertiger Routinen (vor allem aus der Numerik). Als besonders bequem hat sich die Benutzung von statistischen oder numerischen Programmierumgebungen erwiesen. Unter UNIX ist dies S, gegebenenfalls unterstützt durch eine Sammlung von Numerikroutinen, unter MS-DOS oder IBM DOS hat GAUSS große Verbreitung gefunden. Im allgemeinen empfiehlt sich, eine Datenreduktion mit Hilfe eines der gängigen Pakete wie SPSS oder SAS vorzuschalten.

So kann man erfolgreich arbeiten, allein die ganze Zukunft ist das nicht, da letztlich recht altmodische Programmierkonzepte wie Datenfluß und logische Verzweigungen dahinterstehen. Das wohl zur Zeit aussichtsreichste Konzept ist das der "objektorientierten Programmierung" (OOPS, Object Oriented Programming System), wo die Programme "Botschaften" ("Messages") zwischen aktiven "Objekten" senden. Damit ist ein Zurückhalten von Information (Encapsulation, "Abfüllen in Kapseln") verbunden, das heißt, nicht jeder Benutzer verfügt über die gleiche Information bezüglich des Gesamtsystems. Das mag zunächst unverständlich oder trivial erscheinen; in Wirklichkeit ist die objektorientierte Programmierung sehr effizient und dem menschlichen Denken gut nachgebildet.

In der Archäologie könnten Objekte im Sinne von OOPS beispielsweise Objekte im Sinne der Archäologie sein, also Bauten, Kunstwerke, Strukturen (wieder im Sinne der Archäologie verstanden). Botschaften wären Beziehungen zwischen diesen Objekten, z.B. eine Zusammenfügung von Strukturen wie Abzugsrinnen, Brunnenschächten, Gräben, Mauern und Erdwällen, Herden, Brennöfen und dergleichen zu Gebäuden oder Siedlungen. Dabei wird dieselbe Botschaft "Gebäude" natürlich je nach dem einzelnen Objekt verschieden interpretiert. Objekte im Sinne von OOPS wären aber auch Einzelpersonen, die miteinander in Beziehung treten und historisch relevante Handlungen setzen. (Die Theorie des Symbolischen Interaktionismus stellt ein sozio-

logisches Gegenstück zu OOPS dar.) Damit diese Objekte und Botschaften für mathematisches oder statistisches Modellieren verwendbar sind, müssen sie mathematisch oder statistisch relevante Information enthalten, von denen der Historiker vielleicht gar nicht zu erfahren braucht (Encapsulation). Beziehungen können wie Werkzeuge nicht nur als Botschaften, sondern auch als Objekte auftreten. Mathematische und statistische Verfahren wären dann solche Objekte.

OOPS ist weder alltägliche Wirklichkeit noch Utopie. Seit längerer Zeit gibt es das Konzept, das auch praktisch als Leitbild bei der Programmierung fungiert, aber auch eine Reihe von Programmpaketen und Programmiersprachen dazu. Der für den Historiker interessanteste Vertreter ist meines Erachtens SMALLTALK (bzw. SMALLTALK V), das sogar eine graphische Darstellung der Objekte beinhaltet. Hier werden aber auch die Probleme von OOPS deutlich: SMALLTALK und seine Varianten sind schwierig zu erlernen und langsam bei der Programmausführung. Es wartet hier noch sehr viel Arbeit, wie ich meine, aber auch eine ungeheure Chance.

LITERATUR

- ARNOLD, V. I.: Catastrophe Theory. 2. Aufl., Berlin usw. 1986.
- BARNET, V.: Comparative Statistical Inference. London - New York - Sydney - Toronto, 2. Aufl. 1982.
- BEACH, E. F.: Economic Models. New York 1957.
- BINFORD, L. R.: Archaeological systematics and the study of cultural process. American Antiquity 31 (1965) 203-210.
- BOUDON, R.: L'analyse mathématique des faits sociaux. Paris 1967.
- BRUGGER, W.: Philosophisches Wörterbuch. Freiburg, 19. Aufl. 1976.
- CHAMPION, S.: A Dictionary of Terms and Techniques in Archaeology. Oxford 1980. Deutsche Ausgabe: DuMont's Lexikon archäologischer Fachbegriffe und Techniken. Köln 1982.
- CHOW, G. C.: Random and Changing Coefficient Models. Z. Griliches und M. D. Intriligator (Hg.), Handbook of Econometrics. Bd. 2, Amsterdam usw. 1984, 1987.
- CLARKE, D. I.: Analytical Archaeology. London 1968.
- CLAUS, V.: Stochastische Automaten. Stuttgart 1971.

- COLES, J.: *Archaeology by Experiments*. London 1973. Deutsche Ausgabe: *Erlebte Steinzeit - experimentelle Archäologie*. München - Gütersloh - Wien 1973.
- FERGUS, R. M.: *Decision Tables - What, Why and How*. J. D. Couger und R. W. Knapp (Hg.), *System Analysis Techniques*. New York usw. 1974.
- FROEHLICH, B. R.: *Some Estimators for a Random Coefficient Regression Model*. *JASA* 68(1973) 329-335.
- GEBELEIN, H.: *Zahl und Wirklichkeit*. Quelle & Meyer, Leipzig 1943.
- GORDESCH, J.: *Causal Models*. *Colloquia Mathematica János Bolyai*, 267 - 276. Budapest 1972.
- GORDESCH, J.: *Mathematik als Hilfswissenschaft in der Geschichte der Statistik und Staatsbeschreibung*. M. Rassem und J. Stagl (Hg.), *Statistik und Staatsbeschreibung in der Neuzeit*. Paderborn usf. 1980.
- GORDESCH, J.: *A Sampling Procedure for Historical Data*. H. Caussinus und P. Etinger (Hg.), *COMPSTAT 1982. Proceedings in Computational Statistics*. Würzburg - Wien 1982, 242 - 248.
- HARDTWIG, E.: *Fehler- und Ausgleichsrechnung*. BI Hochschultaschenbücher 262/262a, Mannheim - Wien - Zürich 1968.
- HODDER, I., und C. ORTON: *Spatial Analysis in Archaeology*, 153-174. Cambridge 1976.
- HOLDEN, A. V. (Hg.): *Chaos*. Manchester 1986.
- KEYNES, J. M.: *A Treatise on Probability*. 7. Aufl., New York 1962.
- KLEIN, F.: *Elementarmathematik vom höheren Standpunkt aus*. 3. Band: *Präzisions- und Approximationsmathematik*. 3. Aufl., Berlin 1928, Nachdruck 1968.
- KOÇAK, H.: *Differential and Difference Equations through Computer Experiments*. 2. Aufl., New York - Heidelberg - Tokyo 1989.
- MALINVAUD, E.: *Pour une Axiomatique de la Causalité*. H. O. A. Wold (Hg.): *La Techniques des Modèles dans les Sciences Humaines*. Monaco 1964.
- MARDIA, K. V.: *Families of Bivariate Distributions*. London 1970.
- MÜLLER, M., A. HALDER u. a. (Hg.): *Kleines Philosophisches Wörterbuch*. Freiburg im Breisgau 1971.
- NICOLIS, G. und I. PRIGOGINE: *Die Erforschung des Komplexen. Auf dem Weg zu einem neuen Verständnis der Naturwissenschaften*. München 1987.
- ORD, J. K.: *Families of Frequency Distributions*. London 1972.
- PING, C., und D. BLOODWORTH: *The Chinese Macchiavelli. 3000 Years of Chinese Statecraft*. London 1976. Deutsche Ausgabe: *Das chinesische Machtspiel. Dreitausend Jahre Staatskunst*. Tübingen 1977.
- PRIGOGINE, I. und I. STENGERS: *Dialog mit der Natur*. 5. Aufl., München - Zürich 1986.

- RENNERT, P., H. SCHMIEDEL und C. WEISSMANTEL (Hg.): Kleine Enzyklopädie Physik. Leipzig, 2. Aufl. 1988.
- RÖSSLER, O. E.: An Equation for Continuous Chaos. Phys. Lett. 57A(1976), 397 ff.
- RULOFF, D.: Historische Sozialforschung. Stuttgart 1985.
- STEGMÜLLER, W.: Personelle und Statistische Wahrscheinlichkeit. 1. und 2. Halbband, Berlin - Heidelberg - New York 1973.
- STOLZ, F. - DEBRUNNER, A. - SCHMID, W. P.: Geschichte der lateinischen Sprache. Sammlung Göschen Bd. 492/492a, Berlin 1966.
- SUPPES, P.: A Probabilistic Theory of Causality. Acta Philosophica Fennica 24, Amsterdam 1970.
- SWAMY, P. A. V. B.: Efficient Inference in a Random Coefficient Regression Model. Econometrica 38(1970), 311 -323.
- TINBERGEN, J. und H. C. BOS: Mathematical Models of Economic Growth. New York 1962.
- THOM, R.: Stabilité Structurelle et Morphogénèse. Reading, Mass., 1972.
- THOME, H.: Grundkurs Statistik für Historiker. Teil I. Deskriptive Statistik. Historical Social Research, Supplement No. 2 (1989). Teil II. Induktive Statistik und Regressionsanalyse. Historical Social Research, Supplement No. 3 (1990).
- VAIHINGER, H.: Die Philosophie des Als-Ob. Seit 1876, 1. Aufl. 1911, 10. Aufl. 1927.
- WEINBERGER, O.: Rechtslogik. Wien 1970.
- WOLD, H. O.: Causality and Econometrics. Econometrica 16(1954).
- WRIGHT, G. H. von: An Essay on Deontic Logic and the General Theory of Action. Acta Philosophica Fennica 21, Amsterdam 1968.

TEIL II:

Analyse von Ereignisdaten und Zeitreihen

Die Mandatsdauer in den deutschen Reichstagen 1867-1918: Eine Anwendung neuer Methoden der Analyse historischer Verlaufsdaten

von Hans-Jürgen Andreß, Heinrich Best und Kurt Sombert

1. Oligarchisierung in einem parlamentarischen System:
Hypothesen zur Entwicklung der Mandatsdauer im Deutschen Reichstag 1867 bis 1918

Die Zeitlichkeit sozialer Strukturen und Prozesse ist das zentrale Thema der historischen Sozialforschung. Methodische Verfahrensweisen, die den Wandel von Kollektiven in der Zeit und Zeit als Kategorie des sozialen Handelns von Individuen systematisch erfassen, haben deshalb für die Entwicklung der historischen Sozialforschung eine besondere Bedeutung. Methoden der Zeitreihenanalyse, die nach deutbaren Regelmäßigkeiten in den Veränderungsdaten quantitativer Charakteristika sozialer Aggregate suchen, sind inzwischen wohl etabliert;¹ Analysen von Verlaufsdaten, deren Gegenstand Sequenzen und Intervalle in verketteten Ereignisfolgen sind, haben dagegen in die historische Sozialforschung bislang noch keinen Eingang gefunden. Das ist ein Mangel, wird doch Zeit in diesen statistischen Modellen als ein wichtiges Element der Gelegenheitsstruktur des Handelns individueller und kollektiver Akteure betrachtet und nicht lediglich als eine "Dimension", in der sich Wandel vollzieht. Damit werden die vielfältigen Zeitordnungen der Ereignisverkettungen in den Lebensgeschichten von Personen und sozialen Gebilden zum Gegenstand der Analyse, aus deren Verflechtung und Überschneidung sich erst "historische Zeit" konstituiert.

Den Zusammenhang von "historischer" und "sozialer" Zeit veranschaulicht auch der Vorgang, den wir als Beispiel für die Anwendung von Methoden der Analyse von Verlaufsdaten gewählt haben: die Professionalisierung der Politik im Kaiserreich unter dem besonderen Aspekt der Verstetigung der Mandatsausübung in den Reichstagen. Wir haben damit ein Thema gewählt, das mit Max Weber und Robert Michels zwei Klassiker der politischen Soziologie beschäftigt hat. Dies ist kein Zufall, denn es ist in besonderer Weise geeignet, die sich gegenüber formal verfaßten Ordnungen durchsetzende "Eigendynamik" von Prozessen politischer Modernisierung aufzuzeigen. Auch kann es als

¹ Vgl. die Beiträge von Müller-Benedict, Stier, Metz und Thome in diesem Sammelband.

ein Schulbeispiel für ungeplante Folgen sozialen Handelns gelten, denn die Vorgänge, die Max Weber unter dem Gesichtspunkt der "Verberuflichung" der Politik und Robert Michels unter dem Stichwort der "Oligarchisierung" abhandelte, widersprachen entschieden den Absichten und dem Selbstverständnis der antagonistischen politischen Akteure im Kaiserreich: ² Während Bismarck gegenüber dem perhorreszierten "gewerblichen Parlamentarismus" eine parlamentarische Vertretung durch wirtschaftlich unabhängige und nur episodisch in der Politik engagierte Honoratioren anstrebte, war der "wesentlichste Lebenszweck" der Sozialdemokratie als einer sozialrevolutionären und demokratischen Massenpartei neuen Typs "die Bekämpfung der Oligarchie in all ihren Formen". ³ Dennoch hatte sich gegen Ende des Kaiserreichs in bürgerlichen wie in Arbeiterparteien eine Tendenz zur Verberuflichung politischer Betätigung und zur Verstetigung der Mandatsausübung der Reichstagsabgeordneten durchgesetzt. Diese Veränderungen vollzogen sich, obwohl die institutionellen Bedingungen für die Wahl und Mandatsausübung der Reichstagsabgeordneten - mit Ausnahme der seit 1906 gewährten Diäten - im wesentlichen unverändert blieben.

Den Wandel der sozialen Morphologie parlamentarischer Vertretung analysierte Max Weber 1918/1919 in seiner berühmten Abhandlung über "Politik als Beruf". Er entwarf darin eine Typologie politischer Figuren, in der er den Gelegenheitspolitiker, den haupt- und den nebenberuflichen Politiker unterschied. Diese Typologie läßt sich auf ein Kontinuum übertragen, dessen Pole sich nach dem Grad ausrichten, in dem die Politik die "Lebensorientierung" bestimmt. Zugleich wird eine historische Entwicklung konstatiert, die auf die Durchsetzung des Typs des hauptamtlichen Politikers zielt, und zwar in jener Variante, die auch im Sinne "ökonomischen Erwerbs" von der Politik lebt. Die Stichworte, unter denen Weber diesen Vorgang abhandelt, sind "Spezialisierung", "Rationalisierung" und "Scheidung der öffentlichen Funktionäre". ⁴ Hier treten Konturen des Konzeptes einer "Ausdifferenzierung" des

² Vgl. insbesondere Max Weber, "Wahlrecht und Demokratie in Deutschland", in: ders., Gesammelte politische Schriften (Tübingen 1980, zuerst 1917), S. 272-274 und Robert Michels, "Die oligarchischen Tendenzen der Gesellschaft. Ein Beitrag zum Problem der Demokratie", in: ders., Masse, Führer, Intellektuelle. Politisch-soziologische Aufsätze 1906- 1933 (Frankfurt u. New York 1987, zuerst 1908), S 133- 181.

³ Vgl. Ernst Rudolf Huber, Deutsche Verfassungsgeschichte seit 1789, Bd. 3, Bismarck und das Reich (Stuttgart 1963), S. 893. Robert Michels, Zur Soziologie des Parteiwesens in der modernen Demokratie (Stuttgart 1925), S. 370.

⁴ Vgl. Max Weber, "Politik als Beruf", in: ders., Gesammelte politische Schriften (zuerst 1919), S. 516f., 519, 522f.

politischen Handlungsfeldes hervor, die sich auf der personalen Ebene als eine Mediatisierung der privaten Berufsrolle durch die Rolle des Politikers vollzieht.⁵

Der wichtigste soziale Mechanismus, der die Verberuflichung der Politik begünstigte, ist nach Weber die "Abkömmlichkeit". Er verstand darunter im weitesten Sinne die Opportunitätskosten der Ausübung von politischen Ämtern und Mandaten auf Dauer und fern des Wohnorts. Der Politiker, der nicht *v o n*, sondern *f ü r* die Politik lebt, muß nach Max Weber wirtschaftlich abkömmlich sein, "das heißt, seine Einkünfte dürfen nicht davon abhängen, daß er ständig persönlich seine Arbeitskraft, sein Denken voll oder doch weit überwiegend in den Dienst ihres Erwerbs stellt. Abkömmlich in diesem Sinne ist nun am unbedingtesten: der Rentner, derjenige also, der vollkommen arbeitsloses Einkommen, sei es wie die Grundherren der Vergangenheit, die Großgrundbesitzer und die Standesherrn der Gegenwart aus Grundrenten..., sei es aus Wertpapier- oder ähnlichen modernen Rentenquellen bezieht. Weder der Arbeiter, noch - was sehr zu beachten ist - der Unternehmer, auch und gerade der moderne Großunternehmer, ist in diesem Sinn abkömmlich. Denn auch und gerade der Unternehmer - der gewerblich sehr viel mehr als, bei dem Saisoncharakter der Landwirtschaft, der landwirtschaftliche Unternehmer - ist an seinen Betrieb gebunden und nicht abkömmlich. Es ist für ihn meist sehr schwer, sich auch nur zeitweilig vertreten zu lassen. Ebenso wenig ist dies zum Beispiel der Arzt, je hervorragender und beschäftigter er ist, desto weniger. Leichter schon, aus rein betriebstechnischen Gründen, der Advokat - der deshalb auch als Berufspolitiker eine ungleich größere, oft eine geradezu beherrschende Rolle gespielt hat".⁶ Im Mittelpunkt von Webers Argument steht also die wirtschaftliche Abkömmlichkeit, operationalisierbar durch den Hauptberuf der Mandatsträger. Daneben lassen sich aber un schwer weitere, die Opportunitätskosten der Mandatsausübung bestimmende, Komponenten der sozialen Lage von Abgeordneten einführen. Dies gilt etwa für die geographische Herkunft, wobei wir hier von der einfachen Überlegung ausgehen können, daß ein Abgeordneter c.p. Berufsnotwendigkeit und Mandat dann besser vereinbaren kann, wenn er nahe am Sitzungsort wohnt.

Während bei M. Weber der Zeitaufwand für die Mandatsausübung als Kostenfaktor wirkt, Abgeordnete mit geringer Abkömmlichkeit deshalb auch

⁵ Zur politischen Professionalisierung vgl. Dietrich Herzog, *Selektion und Professionalisierung politischer Führungsgruppen* (Opladen 1975) und Hans-Jürgen Hohm, *Politik als Beruf. Zur soziologischen Professionalisierungstheorie der Politik* (Opladen 1987).

⁶ M. Weber, *Politik als Beruf*, S. 514.

eine geringe Mandatsverhaftung aufweisen sollten, sind für Robert Michels Mandate Ressourcen, die es politischen Akteuren unter anderem ermöglichen, der Elitenzirkulation zu entgehen und ihre Amtsdauer zu erhöhen. Der Vorsprung an Erfahrung und innerer Vernetzung, den das Leitungspersonal und die Repräsentanten politischer Verbände mit wachsender Routine gegenüber Wählern und Mitgliedern erlangen, ist eine wichtige strukturelle Voraussetzung für den Prozeß der "Oligarchisierung", dessen Kern die Formierung eines hauptamtlichen und amtsverhafteten Führungspersonals ist. Seine Argumente faßte Robert Michels in dem berühmt gewordenen "ehernen Gesetz der Oligarchie" zusammen: "Die Organisation ist die Mutter der Herrschaft der Gewählten über die Wähler, der Beauftragten über die Auftraggeber, der Delegierten über die Delegierenden. Die Bildung von Oligarchien im Schoße der mannigfaltigen Formen der Demokratie ist eine organische, also eine Tendenz der j e d e Organisation, auch die sozialistische, selbst die libertäre, notwendigerweise unterliegt".⁷

Im Zusammenhang unseres Beitrags ist nun die theoretische Bedeutung der Mandatsdauer von besonderem Belang. Obwohl sie bei Michels als Ressource für die Etablierung oligarchischer Strukturen, bei Weber als Element der Opportunitätskosten politischer Betätigung eingeführt wird, handelt es sich nicht um konkurrierende Deutungen, denn die Mandatszeiten verwandeln sich spätestens in dem Augenblick von Kosten in Ressourcen, wenn der Übergang vom neben- zum hauptberuflichen Politiker vollzogen ist und sich keine Konkurrenzen mehr zwischen der "privaten" Berufsrolle und der Politikerrolle ergeben. Die Argumente von Weber und Michels konvergieren dann auch in modernen Konzepten der "Professionalisierung" politischer Rollen und der "Institutionalisierung" politischer Rekrutierung. Danach erzeugen formale Verfahren und informelle Praktiken ein zunehmend rigides Regime der Auswahl politischen Personals und der Karrierisierung des Zugangs zu politischen Positionen. Als eine wesentliche Folge dieser Entwicklung wird die zunehmende Kontinuität und Dauer legislativer Karrieren diagnostiziert. Die Mandatsdauer in Parlamenten ist demnach ein wichtiger Indikator für den Grad der Institutionalisierung repräsentativer Systeme und der Professionalisierung politischer Rollen.⁸

⁷ R. Michels, Soziologie des Parteiwesens, S. 370f.

⁸ Die Mandatsdauer der Abgeordneten und ihre Fluktuationsraten nach Wahlen haben die historische Parlamentarismusforschung in Deutschland bislang kaum interessiert, was angesichts der Beiträge der Klassiker recht überraschend ist. Nur gelegentlich finden sich einzelne statistische Nachweise. Vgl. etwa Peter Molt, Der Reichstag vor der improvisierten Revolution (Köln u. Opladen 1963), Tab. 30, S. 324; Bernhard

Aus den oben skizzierten Basistheoremen lassen sich folgende Hypothesen über die historische Entwicklung der Mandatsdauer und ihre Abhängigkeit vom beruflichen Hintergrund der Abgeordneten, deren Stellung im Geflecht politischer Positionen und Parteizugehörigkeit ableiten:⁹

1. Im Beobachtungszeitraum wird die durchschnittliche ("aggregierte") Mandatsdauer im Reichstag zunehmen. Ein solcher Trend läßt sich mit Michels' Oligarchisierungstheorie ebenso wie mit neueren Professionalisierungs- und Institutionalisierungstheorien begründen.
2. Konzipiert man politische Professionalisierung individual-biographisch als longitudinalen, sich selbst stabilisierenden Prozeß, werden auch die individuellen Karrieren mit zunehmender Dauer an Stetigkeit gewinnen. Diese Vermutung läßt sich mit Michels' Oligarchisierungstheorie begründen: Amtsinhaber gewinnen mit der Zeit einen Vorsprung an Erfahrung und

Mann, Biographisches Handbuch für das Preussische Abgeordnetenhaus 1867-1918 (Düsseldorf 1988), Tab. 1, S. 515.

Zur Entwicklung des "Mandatsalters" der Reichstagsabgeordneten 1867-1918 vgl. Heinrich Best, "Politische Modernisierung und parlamentarische Führungsgruppen in Deutschland 1867- 1918", in: Historical Social Research/Historische Sozialforschung, 13 (1988), S. 29 -31 u. Abb. 13. In der amerikanischen Parlamentssoziologie bilden 'membership turnover', 'career stability' und 'personal circulation' dagegen schon seit langem zentrale Aspekte von Strukturanalysen parlamentarischer Führungsgruppen. Vgl. zuletzt Peverill Squire, "Career Opportunities and Membership Stability in Legislatures", in: Legislative Studies Quarterly 13 (1988), S. 65-82; Richard G. Niemi und Laura R. Winsky, "Membership Turnover in U.S. State Legislatures: Trends and Effects of Districting", in: Legislative Studies Quarterly, 12 (1987), S. 115- 123; R. Darcy u. James R. Choike, "A Formal analysis of Legislative Turnover: Woman Candidates and Legislative Representation", in: American Journal of Political Science, XX (1986), S. 237-255; Paul Brace, "A Probabilistic Approach to Retirement from the U.S. Congress", in: Legislative Studies Quarterly 10 (1985), S. 107-123; Donald D. Mathews, "Legislative Recruitment and Legislative Careers", in: Gerhard Loewenberg u.a. (Hrsg.), Handbook of Legislative Research (Cambridge u. a. 1985), S. 17-55 (darin auch weitere Literatur). Beispiele für historische Untersuchungen sind Mogens N. Pedersen, "The Personal Circulation of a Legislature: The Danish Folketing 1848-1968", in: William O. Aydelotte (Hrsg.), The History of Parliamentary Behavior, (Princeton 1977), S. 66-101; James Q. Graham, "Legislative Careers in the French Chambers and the U.S. House, 1871-1940", Legislative Studies Quarterly 7 (1982), S. 37-56.

⁹ Eine nutzentheoretische Formalisierung von Michels' Oligarchisierungstheorem findet sich bei Reinhard Wippler, "Die Entstehung oligarchischer Strukturen in demokratisch verfaßten Organisationen", in: Günter Büschges und Werner Raub (Hrsg.), Soziale Bedingungen - individuelles Handeln - soziale Konsequenzen (Frankfurt a.M. u.a. 1985), S. 23-48. Darin wird Oligarchisierung als ein umkehrbarer kumulativer Prozeß konzipiert, bei dem die Variable Zeit eine zentrale Rolle spielt (vgl. unsere Hypothesen 2 und 5).

innerer Vernetzung, der sie gegenüber möglichen Konkurrenten überlegen macht.

3. Gegenläufig wirkt ein biologisch-demographischer Mechanismus: Zunehmendes Alter verringert das physische Potential im Konkurrenzkampf um politische Mandate.
4. Ein "privater" Beruf mit geringer Abkömmlichkeit - typischerweise eine nicht delegierbare unternehmerische Tätigkeit - wird die Mandatsdauer verkürzen; verlängern wird sie dagegen ein "politiknaher" Beruf, dessen Tätigkeitsmerkmale dem Bild politischer Rollen entsprechen und bei dem private Karriere- und Einkommenschancen aus politischer Tätigkeit zu wachsen.
5. Vertikale Ämterkumulation wird die Mandatsdauer erhöhen. Durch die Rückbindung an lokale und regionale Unterstützungssysteme verbessert der Abgeordnete seine Wiederwahlchancen. Ähnliches gilt für horizontale Ämterkumulation: Spitzenpositionen sind nicht nur Ziele, sondern auch Ressourcen im Konkurrenzkampf um die Macht. Sie verleihen einen Erfahrungsvorsprung, öffnen den Zugang zu rekrutierungsrelevanten Zirkeln und die Möglichkeit, durch die Gewährung von Patronage die persönliche Unterstützung von Gefolgschaften zu gewinnen. Hinzu kommt im gegebenen Untersuchungszeitraum der unmittelbar stabilisierende Effekt einer parallelen Mitgliedschaft in einzelstaatlichen Parlamenten, die Abgeordneten Tagegelder zahlten, während für den Reichstag bis 1906 das Bismarcksche Diätenverdict galt.
6. Die Mandatsdauer der Abgeordneten wird schließlich entscheidend durch die Wiederwahlchancen mitbestimmt, die sich aus der Stellung ihrer Parteien auf dem "support market" des Elektorats ergeben. Haben Parteien einen stabilen oder wachsenden Stimmenanteil, begünstigt dies c.p. lange Mandatsdauern, während bei schrumpfendem oder stark fluktuierendem Stimmenanteil die durchschnittliche Mandatsdauer abnehmen wird. Ein weiterer Effekt des "support market" der Parteien auf die Mandatsdauer resultiert aus dem territorialen Mehrheitswahlrecht des Kaiserreichs: konzentrierte sich das Elektorat einer Partei räumlich in bestimmten Wahlkreisen, erhöhte das die Wiederwahlchancen der Abgeordneten, und c.p., die Mandatsdauer.

2. Statistische Probleme der Analyse von Zeitdauern

Grundlage der folgenden Auswertung sind die Zeitdauern vom erstmaligen Eintritt eines Abgeordneten in einen deutschen Reichstag bis zum Ende

seines letzten Mandats innerhalb des Kaiserreichs. Ein Mandat endet, wenn der Abgeordnete nicht mehr wiedergewählt wird oder aus anderen Gründen ausscheidet (z.B. wegen seines Todes).¹⁰ Nur der erste Aspekt (keine Wiederwahl) ist in diesem Zusammenhang von Interesse und wird daher als reguläres Ende eines Mandats bezeichnet. Der gesamte Untersuchungszeitraum erstreckt sich auf die Jahre 1867 (Konstituierender Norddeutscher Reichstag) bis 1918 (Ende des Kaiserreichs). Abgeordnetenmandate in der Weimarer Republik werden also nicht berücksichtigt, so daß für Abgeordnete, die über das Kaiserreich hinaus Mitglied des Reichstages waren, nur eine unvollendete (technisch: zensierte) Mandatsdauer angegeben werden kann.

Eine mögliche Analysestrategie bei der Betrachtung von Mandatsdauern wäre die Berechnung von Durchschnittswerten für verschiedene Subgruppen der Stichprobe. Auf diese Weise ließe sich beispielsweise die durchschnittliche Mandatsdauer der Abgeordneten aus den früheren Reichstagen mit der Mandatsdauer aus den späteren Reichstagen vergleichen. Leider sind diese Statistiken jedoch durch mindestens zwei Gründe verzerrt:

- a) Eine abgeschlossene Mandatsdauer, d.h. die Zeitdauer bis zum regulären Ende (s. oben) des Mandats, kann für die Abgeordneten der letzten Reichstage nicht mehr berechnet werden, da der Untersuchungszeitraum 1918 endet. Ein Teil der Abgeordneten ist zu diesem Zeitpunkt noch Mandatsträger und wird es auch in der Weimarer Republik bleiben.
- b) Das Ende eines Reichstagsmandates tritt definitionsgemäß nicht nur dann ein, wenn eine Person nicht wiedergewählt wird, sondern auch dann, wenn sie stirbt, befördert wird oder aus anderen Gründen den Reichstag verläßt. Wären diese konkurrierenden Risiken nicht aufgetreten, hätte das Mandat noch weiter bestehen können.

Beide Aspekte (Zensierung durch Begrenzung des Untersuchungszeitraums, konkurrierende Risiken) sind typisch für die Analyse von Zeitdauern. Werden sie nicht durch angemessene Methoden kontrolliert, ist mit einer Unterschätzung der tatsächlichen Zeitdauern und mit einer Verzerrung der Strukturvergleiche zu rechnen. Die Auswirkungen auf das Untersuchungsbeispiel lassen sich in Tabelle 1 studieren.

¹⁰ Zwischen dem Beginn des ersten und dem Ende des letzten Mandats innerhalb des Kaiserreichs sind Unterbrechungen möglich, weil Abgeordnete nicht Mitglied aller zwischenzeitlichen Reichstage waren. Solche Unterbrechungen treten jedoch selten auf. Die hier untersuchten Zeitdauern berücksichtigen lediglich die Netto-Mandatsdauer (Differenz Ende-Beginn abzüglich Unterbrechungszeiten). Gegenüber früher publizierten Ergebnissen unterscheidet sich diese Untersuchung auch durch die Kontrolle konkurrierender Risiken.

Tabelle 1: Mittelwert und Median der Mandatsdauer

Variable	Häufigkeit		Mittelwert		Mittelwert		Median	
	Alle	zens. Fälle	Abs.	Diff.	Abs.	Diff.	Abs.	Diff.
Hauptberuf:								
Rest	137	48 35,04%	5,96	-1,20	4,70	-1,57	5,89	-1,25
Staatsdienst	783	260 33,21%	6,30	-0,86	5,78	-0,48	6,19	-0,95
Industrielle/Großkaufleute	278	66 23,74%	6,60	-0,56	6,05	-0,21	6,06	-1,08
Freie Berufe	249	86 34,54%	7,29	0,13	6,45	0,19	7,49	0,35
Handwerk/Kleinhandel	174	64 36,78%	7,48	0,32	6,13	-0,13	8,03	0,89
Geistliche	126	54 42,86%	7,56	0,40	6,44	0,18	8,13	0,99
Gutsbesitzer	731	233 31,87%	7,57	0,40	7,05	0,78	7,50	0,36
Funktionäre/Journalisten	291	180 61,86%	9,19	2,03	6,47	0,21	17,67	10,53
Fraktionszugehörigkeit:								
Rechtsliberale	638	170 26,65%	5,70	-1,46	5,53	-0,73	5,79	-1,35
Linksliberale	380	108 28,42%	6,43	-0,73	5,78	-0,49	6,16	-0,98
Rest	425	139 32,71%	6,52	-0,64	5,98	-0,28	6,19	-0,95
Konservative	629	217 34,50%	7,06	-0,10	6,39	0,12	6,89	-0,25
Zentrum	488	211 43,24%	9,22	2,06	8,18	1,92	11,41	4,27
Sozialdemokraten	214	149 69,63%	9,65	2,49	5,83	-0,43	42,00+	34,86
Wohnort:								
im Wahlkreis	1516	498 32,85%	6,82	-0,34	6,24	-0,02	6,80	-0,34
nicht im Wahlkreis	1258	496 39,43%	7,57	0,41	6,28	0,02	7,60	0,46
Politische Ämter:								
keine	1584	565 35,67%	6,40	-0,76	5,60	-0,66	6,35	-0,79
nur lokale	508	199 39,17%	7,82	0,66	6,72	0,46	8,40	1,26
nur regionale	425	146 34,35%	7,98	0,82	7,28	1,02	7,88	0,74
lokale und regionale Ämter	257	84 32,68%	9,20	2,04	7,68	1,42	9,49	2,35
Einzelstaatl. Parlamente:								
keine Mitgliedschaft	1498	493 32,91%	5,65	-1,51	5,01	-1,26	5,70	-1,44
andere einzel. Parlamente	664	247 37,20%	8,62	1,46	7,42	1,16	9,38	2,24
Preuss. Abgeordnetenhaus	612	254 41,50%	9,26	2,10	8,44	2,18	10,79	3,65
Verbandsaffiliation:								
kein Verbandsmitglied	1468	459 31,27%	5,72	-1,44	5,26	-1,00	5,63	-1,51
nicht-pol. Verband	1021	379 37,12%	8,50	1,34	7,43	1,17	9,29	2,15
politischer Verband	285	156 54,74%	9,78	2,62	8,29	2,03	14,68	7,54
Beginn des ersten Mandats:								
1867-1873	775	185 23,87%	6,15	-1,01	5,68	-0,59	4,90	-2,24
1874-1883	629	163 25,91%	7,31	0,15	6,54	0,28	6,35	-0,79
1884-1892	423	113 26,71%	7,87	0,71	6,63	0,36	6,73	-0,41
1893-1902	399	166 41,60%	9,21	2,05	7,48	1,22	10,30	3,16
1903-1918	548	367 66,97%	6,37	-0,79	5,26	-1,00	16,00+	8,86
Insgesamt	2774	994 35,83%	7,16		6,26		7,14	

Insgesamt beträgt die durchschnittliche Mandatsdauer 7,16 Jahre. Von den insgesamt 2774 Abgeordneten sind jedoch 994 (35,83%) vorzeitig ausgeschieden oder am Ende des Kaiserreichs noch im Reichstag (zensierte Fälle und konkurrierende Risiken). Nur bei 1780 Abgeordneten konnte innerhalb des Untersuchungszeitraums ein reguläres Mandatsende beobachtet werden (Spalte "Ereignisse"). Diese Abgeordneten waren durchschnittlich 6,26 Jahre Mitglied des Reichstages. Dieser Durchschnittswert unterschätzt natürlich die tatsächliche Mandatsdauer ganz erheblich. Aber auch der vorher genannte Wert von 7,16 dürfte zu gering sein, denn einige der Abgeordneten setzten

auch noch in der Weimarer Republik ihre Laufbahn fort oder wären weiter im Reichstag geblieben, wären sie nicht wegen Beförderung, Krankheit, Tod oder sonstigen Gründen frühzeitig ausgeschieden.

Mit den weiter unten beschriebenen Methoden kann man nun die Wahrscheinlichkeit schätzen, daß Abgeordnete über einen bestimmten Zeitraum ihr Mandat behalten. Auf dieser verbesserten Schätzung, die Zensierungen und konkurrierende Risiken kontrolliert,¹¹ beruhen die Zahlen in den beiden letzten Spalten der Tabelle 1. Danach sind 50% aller Abgeordneten mindestens 7,14 Jahre (Median) Mitglied des Reichstages.¹² Zum Vergleich: Die entsprechenden unkorrigierten Medianwerte (nicht ausgewiesen) für alle 2774 Abgeordneten bzw. alle 1780 Abgeordneten mit regulärem Mandatsende liegen weit darunter, nämlich jeweils bei 5 Jahren.

Solange Durchschnittswerte die tatsächliche Mandatsdauer immer um den gleichen Betrag unterschätzen, ist darin noch kein großes datenanalytisches Problem zu sehen, denn der exakte numerische Wert einer Zeitdauer dürfte im Rahmen dieser historischen Untersuchung eher von nebensächlicher Bedeutung sein.¹³ Das Ausmaß der Unterschätzung variiert jedoch mit dem Anteil zensierter Beobachtungen und da dieser Anteil (notwendigerweise) mit den hier interessierenden Strukturmerkmalen zusammenhängt, ist absehbar, daß auch der Vergleich von Durchschnittswerten zwischen einzelnen Subgruppen nur bedingt richtige Ergebnisse liefert. So ist z.B. der Anteil zensierter Mandatsdauern bei den Sozialdemokraten besonders hoch (69,63%), weil die Mitgliederzahl der sozialdemokratischen Fraktion erst seit 1890 deutlich wuchs und dementsprechend viele sozialdemokratische Abgeordnetenmandate in die Weimarer Republik hinüberreichten. Noch deutlicher ist dieser Effekt, wenn man die Mandatsdauern nach dem zeitlichen Beginn differenziert: Der Anteil zensierter Beobachtungen nimmt (notwendigerweise) im Zeitablauf zu.

¹¹ Solange man nicht an dem konkreten Beendigungsgrund interessiert ist (außer dem oben definierten regulären Ende), kann man diese konkurrierenden Risiken behandeln wie zensierte Beobachtungen. Vgl. dazu H.J. Andreß, *Multivariate Analyse von Verlaufsdaten*. ZUMA Methodentexte Bd. 1 (Mannheim 1985), S. 206-209. Im folgenden werden wir daher die Differenzierung von zensierten Beobachtungen und konkurrierenden Risiken nur noch im Bedarfsfall verwenden.

¹² Wenn der Median in Tabelle 1 mit einem "+" Zeichen versehen ist, bedeutet das, daß weniger als 50% der Gruppe innerhalb des Untersuchungszeitraums ein reguläres Mandatsende hatte. In diesem Fall ist eine Schätzung der 50%-Überlebenswahrscheinlichkeit und damit auch des Medians nicht möglich. Der tatsächliche Median liegt daher über (+) dem ausgewiesenen Wert.

¹³ Diese Frage ist natürlich im Rahmen einer prognostischen Untersuchung (z.B. in der Ökonomie oder Soziologie) ganz anders zu beurteilen.

Im Hinblick auf die Ausgangshypothese einer zunehmenden Professionalisierung ist diese zeitliche Differenzierung von entscheidender Bedeutung. Betrachtet man nur die unkorrigierten Durchschnittswerte, so läßt sich zwar bei beiden Varianten (Spalten "Alle" und "Ereignisse") eine Zunahme der Mandatsdauern bis zur Jahrhundertwende erkennen, für die letzten Reichstage liegen die Durchschnittswerte jedoch wieder unter dem Gesamtdurchschnitt. Korrigiert man dagegen zensierte Beobachtungen und konkurrierende Risiken (s. Spalte Median), dann ist eine deutliche Zunahme der Mandatsdauern über den gesamten Zeitraum erkennbar. Anders ausgedrückt: Ohne geeignete Methoden, die den spezifischen Charakter von Verlaufsdaten berücksichtigen, wäre die Hypothese einer zunehmenden Professionalisierung mit entsprechend längeren Abgeordnetenmandaten nicht entscheidbar. Im folgenden Abschnitt sollen die wesentlichen Konzepte und Auswertungsstrategien der Verlaufsdatenanalyse vorgestellt werden, ehe wir zu den Ergebnissen unserer Datenanalyse kommen.

3. Verlaufsdatenanalyse: Statistische Konzepte und Auswertungsstrategien

3.1 Definition

Verlaufs- oder auch Ereignisdatenanalyse wird überall dort eingesetzt, wo es um die Längsschnittanalyse von diskreten abhängigen Merkmalen geht (z.B. Berufswechsel, Änderungen des Familienstands, des Wohnorts oder - wie im gegebenen Fall - der Mitgliedschaft im Reichstag). Die Veränderungen des diskreten Merkmals im Zeitablauf bezeichnet man auch als Ereignisse (event, failure), die zeitliche Abfolge mehrerer Ereignisse als Verlaufsdaten (event-history).

Gegenstand des eigentlichen statistischen Modells ist nicht nur das Ereignis selbst, sondern auch und vor allem die Zeitdauer bis zum Eintreten dieses Ereignisses (Wartezeit, waiting/ failure time, duration data): In unserem Falle also die Mandatsdauer, die je nach Art der Beendigung (keine Wiederwahl, Krankheit, Tod usw.) mit einem bestimmten Ereignis abgeschlossen wird, vorausgesetzt innerhalb des Untersuchungszeitraums konnte überhaupt ein Mandatsende beobachtet werden.

Wie aus dem vorhergehenden Abschnitt deutlich wurde, müssen dabei zwei Spezifika von Verlaufsdaten berücksichtigt werden:

- a) Untersuchungseinheiten ohne Ereignis im Untersuchungszeitraum (zensierte Beobachtungen);
- b) Veränderungen (z.B. Tod), ohne deren Auftreten das eigentlich interessie-

rende Ereignis (reguläres Mandatsende) beobachtbar gewesen wäre (konkurrierende Risiken).

Alle diese statistischen Modelle benutzen Ergebnisse der Theorie stochastischer Prozesse. In der entsprechenden Fachliteratur werden sie auch häufig als Survival Analyse oder als Modelle zur Analyse von Zeitdauern bezeichnet. Der interessierte Leser findet hierzu mittlerweile eine ganze Reihe deutsch- und englischsprachiger Lehrbücher.¹⁴

3.2 Überlebensfunktion und Raten

Zwei statistische Konzepte sind für die Verlaufsdatenanalyse zentral:

- die Überlebenswahrscheinlichkeit und
- die Übergangsrate (Hazardrate, Risikofunktion oder kurz: Rate).

Grob gesagt, handelt sich hierbei um die Wahrscheinlichkeit, daß ein bestimmter Zeitpunkt ohne Ereignis "überlebt" wird, bzw. um die Rate, mit der Ereignisse pro Zeiteinheit auftreten. Eine formale Definition beider Konzepte findet sich in der einschlägigen Literatur.¹⁵

An dieser Stelle soll der Hinweis genügen, daß beide Konzepte zensierte und unzensierte Beobachtungen (Ereignisse) gleichermaßen berücksichtigen. Bei der Schätzung wird im Prinzip die Anzahl der Ereignisse auf die Zeit bezogen, die die Untersuchungseinheiten im Ausgangszustand verbracht haben, bis entweder das Ereignis eingetreten ist oder die Untersuchung beendet

¹⁴ H.J. Andreß, Multivariate Analyse von Verlaufsdaten (Mannheim 1985); H.P. Blossfeld, A. Hamerle und K.U. Mayer, Ereignisanalyse, Statistische Theorie und Anwendung in den Wirtschafts- und Sozialwissenschaften (Frankfurt 1986); A. Diekmann und P. Mitter, Methoden zur Analyse von Zeitverläufen: Anwendungen stochastischer Prozesse bei der Untersuchung von Ereignisdaten (Stuttgart 1984); J.D. Kalbfleisch und R.L. Prentice, The statistical analysis of failure time data (New York 1980); J.F. Lawless, Statistical models and methods for lifetime data (New York 1982); N.B. Tuma und M.T. Hannan, Social dynamics: models and methods (New York 1984).

¹⁵ Sei T die Zufallsvariable "Wartezeit", dann entspricht die Überlebenswahrscheinlichkeit $S(t)$ der Wahrscheinlichkeit, daß die Wartezeit mindestens so groß ist wie t :

$$S(t) = \Pr(T > t).$$

Die Überlebensfunktion entspricht 1 minus der (kumulierten) Verteilungsfunktion $F(t)$.

Sei $q_{jk}(t, t+dt)$ die Wahrscheinlichkeit eines Wechsels zwischen Zustand j und Zustand k im Intervall $(t, t+dt)$, dann entspricht die Rate $r(t)$ dem Grenzwert dieser Wahrscheinlichkeit für sehr (infinitesimal) kleine Intervalle:

$$r(t) = \lim_{dt \rightarrow 0} q_{jk}(t, t+dt) / dt.$$

Sie mißt quasi die momentane Neigung zum Zustandswechsel. Vgl. auch die Literatur in Anm. 14.

wurde. Weiterhin kann man zeigen, daß beide Konzepte nicht voneinander unabhängig sind. Aus einer Annahme über den zeitlichen Verlauf der Rate (konstant, zu- oder abnehmend) läßt sich eindeutig die Veränderung der Überlebenswahrscheinlichkeit im Zeitablauf (Überlebensfunktion) ableiten.¹⁶ Aus diesen und aus inhaltlichen Gründen hat daher die Übergangsrate in der Verlaufsdatenanalyse einen zentralen Stellenwert.

So lassen sich die alternativen Hypothesen dieser Untersuchung direkt in entsprechende Regressionsmodelle für Übergangsraten übersetzen:

- a) Die durchschnittliche Mandatsdauer variiert nicht im Zeitablauf und ist bis auf zufällige Abweichungen für alle Abgeordneten gleich:

$$r(t) \text{ konstant.}$$

- b) Die durchschnittliche Mandatsdauer ist kürzer für Abgeordnete, die keine Erfahrungen in (politischen) Verbänden haben oder beruflich unabhkömmlich sind:

$$r(t) \text{ individuell verschieden.}$$

- c) Auf Grund eines höheren Professionalisierungsgrades ist die durchschnittliche Mandatsdauer in den späteren Reichstagen länger als in den früheren:

$$r(t) \text{ nimmt im Zeitablauf ab}$$

Aufgrund jeder dieser Annahmen über die Rate ergibt sich jeweils ein bestimmtes Verteilungsmodell der Wartezeiten mit entsprechendem Erwartungswert und umgekehrt. Wie wir später sehen werden, kann man das Erklärungsmodell entweder für die Rate oder für die Zeitdauern spezifizieren, die Ergebnisse sind äquivalent.

Geht man z.B. von einer zeitkonstanten Rate $r(t) = \alpha$ aus, ergeben sich exponentiell verteilte Wartezeiten, deren Erwartungswert dem Kehrwert $1/\alpha$ der Rate entspricht. Unterstellt man einen bestimmten zeitlichen Verlauf der Rate, z.B. $r(t) = \beta\alpha(\alpha t)^{\beta-1}$, dann ergeben sich Weibull-verteilte Wartezeiten usw. Alle diese Verteilungen haben zwei Eigenschaften, die sie (im Gegensatz

¹⁶ Sei $H(t)$ das Integral der Rate $r(t)$ in den Grenzen von 0 bis t , kurz die integrierte Risikofunktion, dann entspricht die Überlebenswahrscheinlichkeit $S(t)$:

$$S(t) = \exp \{-H(t)\}.$$

Die Überlebensfunktion läßt sich also aufgrund der Rate berechnen. Für eine zeitkonstante Rate $r(t) = \alpha$ ergibt sich z.B. $S(t) = \exp(-\alpha t)$. In diesem Fall entspricht die Verteilung der Wartezeiten einer Exponentialverteilung.

etwa zur Normalverteilung) insbesondere zur Analyse von Zeitdauern qualifizieren:

- Sie beschreiben positive Zufallsvariable, denn Zeitdauern können keine negativen Werte annehmen,
- und sie können schiefe Verteilungen modellieren, denn häufig treten viele kurze und wenige lange Zeitdauern auf.

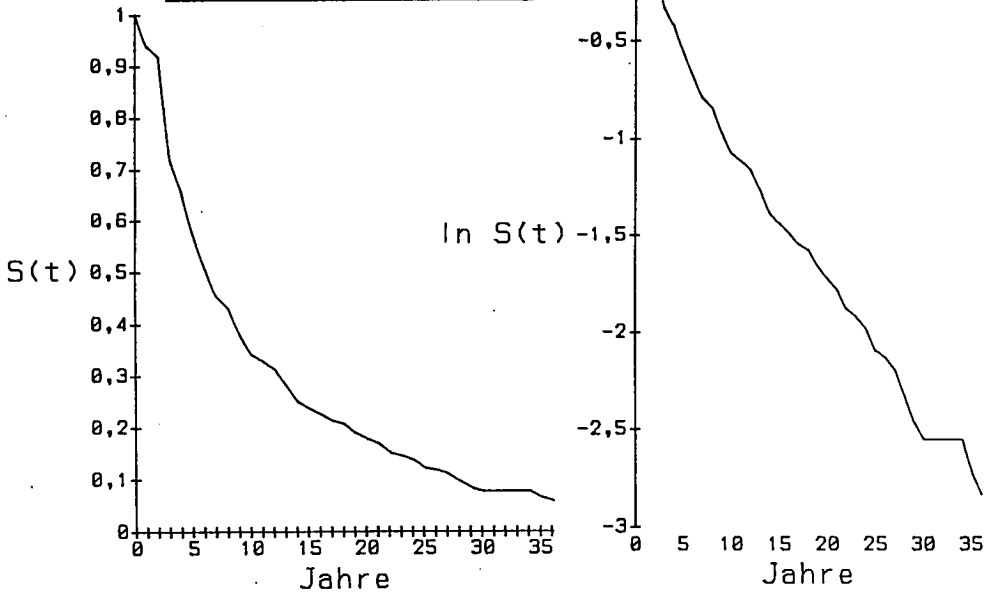
Wir wollen uns jedoch nicht mit den statistischen Details beschäftigen, sondern zur Vorbereitung und zum besseren Verständnis der eigentlichen Auswertung im folgenden die wesentlichen Analyseinstrumente anwendungsorientiert vorstellen. Wir haben sie ganz grob in eher deskriptiv-explorative und eher konfirmatorische Methoden unterteilt.

3.3 Deskriptive Methoden

Für deskriptive Zwecke ist es häufig einfacher, die Überlebensfunktion und nicht die Rate zu schätzen. Hierzu gibt es zwei nichtparametrische Verfahren: Einerseits den Ansatz von Kaplan und Meier, der sortierte Individualdaten verwendet,¹⁷ und andererseits den (verallgemeinerten) Sterbetafelansatz, der von aggregierten Rohdaten ausgeht. Abbildung 1 zeigt das Ergebnis der Sterbetafelschätzung für alle 2774 Abgeordneten. An Hand dieser Abbildung läßt sich der Median (7,14 Jahre) oder jedes andere Quantil der Verteilung ablesen. Aufschlußreich ist auch eine logarithmische Transformation $\ln S(t)$ der Überlebensfunktion: Wenn diese Funktion näherungsweise einer Geraden entspricht, dann kann man schlußfolgern, daß sich die Rate nicht mit der Zeit verändert.¹⁸ Wohlgermerkt handelt es sich bei der Zeit t um die Mandatsdauer und nicht um die historische Zeit. Der näherungsweise lineare Verlauf der Funktion $\ln S(t)$ in Abbildung 1 weist also darauf hin, daß das Ende einer Abgeordnetenkarriere jederzeit gleich wahrscheinlich ist. Die Abbildung sagt jedoch nichts darüber aus, wie sich die Mandatsdauern im historischen Verlauf entwickeln. Dazu betrachten wir Abbildung 2, in der die Überlebensfunktion $S(t)$, genauer gesagt $\ln S(t)$, für verschiedene Subgruppen

¹⁷ E. L. Kaplan und P. Meier, Nonparametric estimation from incomplete observations, *Journal of the American Statistical Association* 53 (1958), S. 457-481.

¹⁸ Hintergrund dieser Überlegung sind die Zusammenhänge zwischen Überlebensfunktion und Rate (vgl. Anm. 16). Sei $r(t) = \text{zeitkonstant}$, dann gilt für die Überlebensfunktion $S(t) = \exp(-\alpha t)$ bzw. für den natürlichen Logarithmus $\ln S(t) = -\alpha t$. Der natürliche Logarithmus der Überlebensfunktion muß also bei einer zeitkonstanten Rate eine Gerade mit Steigung $-\alpha$ ergeben, die den Nullpunkt des Koordinatenkreuzes schneidet. Um kompliziertere Funktionen der Rate zu testen, werden andere Transformationen betrachtet: z.B. die Log-Log-Transformation bei einer Weibull-Rate.

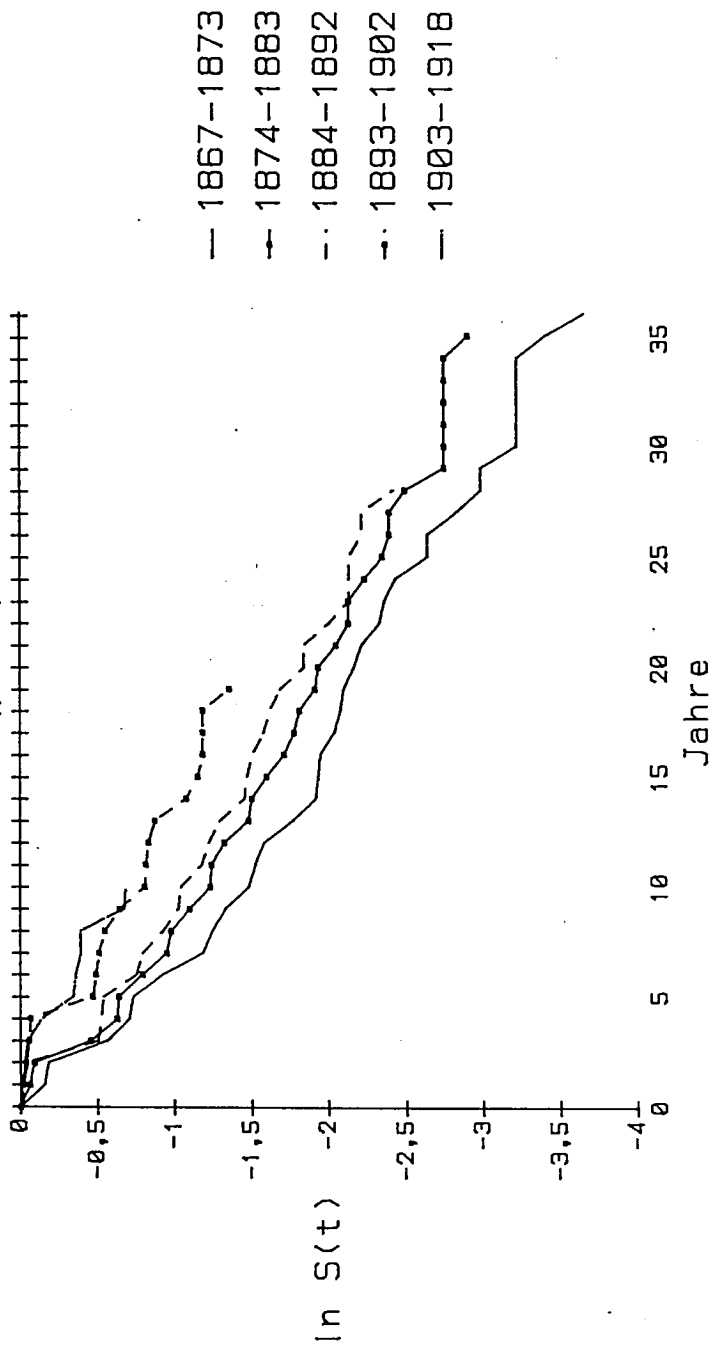
Abbildung 1:
Überlebensfunktion

von Abgeordneten dargestellt wird, die ihre Mandate in jeweils unterschiedlichen Zeitperioden begonnen haben.

Die subgruppenspezifischen Überlebensfunktionen verlaufen wiederum näherungsweise linear, was erneut ein Hinweis darauf ist, daß sich die Abgangsrate auf der individuellen Ebene nicht im Zeitablauf verändert. Deutlich ist aber zu erkennen, daß das Niveau der Überlebenswahrscheinlichkeit im historischen Zeitablauf zunimmt, d.h. die Mandatsdauer ist in den letzten Reichstagen durchschnittlich länger als in den ersten. Zusammengefaßt könnte man also als Ergebnis der deskriptiven Analyse formulieren: Auf der individuellen Ebene zeigen sich keine Anzeichen einer zunehmenden Oligarchisierung, in dem Sinne, daß mit zunehmender Mandatsdauer der Verbleib im Parlament immer wahrscheinlicher wird. Dagegen gibt es im historischen Verlauf deutliche Anzeichen für eine zunehmende Professionalisierung der Reichstage insgesamt, die sich in einer durchschnittlich längeren Mandatsdauer für die späteren Reichstage des Kaiserreichs bemerkbar macht.

Aus Abbildung 2 wird auch deutlich, wie man erste Tests möglicher Erklärungsfaktoren durchführen kann: Dazu wird das Datenmaterial disaggregiert und es wird untersucht, ob die geschätzten Überlebensfunktionen zwischen den Gruppen variieren, die sich bezüglich der in Frage stehenden

Abbildung 2: Ueberlebensfunktion nach Zeitperioden



Merkmale unterscheiden. Eine Möglichkeit, solche Gruppenunterschiede zu testen, ist die Berechnung von Konfidenzintervallen für die Überlebensfunktionen. Damit sind jedoch nur zeitpunktspezifische Vergleiche möglich. Man benötigt daher Testverfahren, die den ganzen Untersuchungszeitraum berücksichtigen. Hierzu verwendet man Verallgemeinerungen bekannter Rangtests. Ein verallgemeinerter Wilcoxon- und ein verallgemeinerter Savage-Test (Log-Rang-Test) sind die am häufigsten angewendeten Tests. Es ist natürlich auch möglich, spezielle parametrische Verteilungsmodelle (z.B. exponentiell verteilte Wartezeiten) mit Hilfe eines Likelihood-Verhältnis-Tests zu untersuchen. Diese Analysestrategie soll hier jedoch nicht weiter verfolgt werden. Wir verwenden gleich einen allgemeineren Ansatz.

3.4 Regressionsmodelle

Nachdem die aus theoretischen Überlegungen abgeleiteten Hypothesen diesen ersten empirischen Test überstanden haben, kann man die Struktureffekte der erklärenden Merkmale und die zeitlichen Veränderungen des Prozesses in einem multivariaten Modell testen (Konfirmatorische Datenanalyse). Dabei haben Regressionsmodelle für Übergangsraten (proportional hazards model) in den Sozialwissenschaften die weiteste Verbreitung gefunden, weil die Rate auch gerade unter inhaltlichen Gesichtspunkten von zentraler Bedeutung ist (vgl. Abschnitt 3.2). Es liegt daher nahe, ein Regressionsmodell für die Rate zu formulieren und diese Annahmen empirisch zu überprüfen. Da eine Rate keine negativen Werte annehmen kann, verwendet man üblicherweise ein log-lineares Modell:

$$r_i(t) = \alpha(t) \exp(x_i \beta), \quad i = 1, \dots, N.$$

Dabei ist x_i der Zeilenvektor der p Kovariaten der Untersuchungseinheit i (inkl. einer Regressionskonstanten), β der entsprechende p -dimensionale Spaltenvektor der Regressionskoeffizienten und $\alpha(t)$ eine zu spezifizierende Funktion der Zeit, die die Veränderungen des untersuchten Prozesses im Zeitablauf beschreibt.

Eine andere, in den Sozialwissenschaften weniger verbreitete Formulierung, betrachtet die logarithmierten Wartezeiten als Zielvariable (accelerated failure time model). Beide Modellklassen sind für bestimmte Wartezeitverteilungen (z.B. Exponential- oder Weibull-Verteilung) direkt ineinander überführbar. Die Parameterschätzungen des einen Modells sind daher aus denen des anderen berechenbar und umgekehrt. Die Wahl der einen oder der

anderen Zielvariable ist im Prinzip eine Frage der persönlichen Vorlieben bzw. der verfügbaren Software. Wir verwenden im folgenden ein Regressionsmodell für Raten, werden jedoch die Ergebnisse aus Gründen der Anschaulichkeit häufig in Bezug auf die Mandatsdauern interpretieren.

Das bevorzugte Schätzverfahren dieser Modellklasse ist Maximum- bzw. Partial Likelihood. Das hat vor allen Dingen den Vorteil, daß bei ML- bzw. PL-Schätzungen die gesamte Information des Datenmaterials, zensierte und unzensierte Beobachtungen, berücksichtigt werden kann. Die einzelnen Modelle unterscheiden sich im wesentlichen nur durch die Form der Zeitabhängigkeit $\alpha(t)$. Die Palette reicht von Cox's Regressionsmodell,¹⁹ das die Veränderungen im Zeitablauf nicht näher spezifiziert, über zeitkonstante Raten (exponentiell verteilte Wartezeiten) bis hin zu komplizierten Funktionen der Zeit (u.a. Gompertz- oder Weibull-verteilte Wartezeiten), mit denen sich zu- und abnehmende Raten modellieren lassen.

In unserem Fall besteht der Kovariatenvektor x_i aus den Merkmalen "Beruf", "Fraktionszugehörigkeit", "Wohnort", "politische Funktionen", "Mitgliedschaft in einzelstaatlichen Parlamenten" und "Verbandsaffiliation". Da dies alles diskrete Merkmale sind, müssen sie natürlich noch in entsprechende Dummy-Variablen transformiert werden (s.unten). Auf Grund unserer theoretischen Überlegungen und der deskriptiven Analyse haben wir weiterhin Hinweise darauf, daß sich die Mandatsdauer im Verlauf des Kaiserreichs verlängert hat oder, anders ausgedrückt, die Abgangsrate verringert hat. Wir wissen aber relativ wenig über den konkreten funktionalen Verlauf dieser Abnahme. Aus diesem Grund haben wir eine Funktion $\alpha(t)$ gewählt, die über den Verlauf relativ wenige Annahmen macht. Genauer gesagt, haben wir den gesamten Untersuchungszeitraum 1867 - 1918 in 5 Perioden aufgeteilt, innerhalb derer die Abgangsrate als konstant angenommen wird, während sie zwischen den Perioden variieren darf. Technisch funktioniert dies durch das Hinzufügen weiterer Dummy-Variablen zum Kovariatenvektor x_i , die jeweils angeben, in welcher Periode das Reichstagsmandat zuerst angenommen wurde. Unser Regressionsmodell reduziert sich daher auf folgende Funktionsgleichung:

$$r_i(t) = \exp(x_i\beta), i = 1, \dots, N.$$

Der unbekannte Verlauf der Rate wird also durch eine Art Treppenfunktion angenähert und muß nicht a priori durch eine bestimmte Funktion fest-

¹⁹ D.R. Cox, Regression models and life tables, Journal of the Royal Statistical Society, B, 34 (1972), S. 187-220.

gelegt werden. Theoretisch ist es denkbar, daß die Rate zunächst steigt, dann sinkt, um in den letzten Perioden wieder zuzunehmen. Der Preis dieser Flexibilität ist natürlich eine gewisse Willkür bei der Einteilung der Perioden und die Annahme zeitkonstanter Raten innerhalb der Perioden, welche jedoch bei einer kurzen Periodenlänge nicht besonders schwer wiegen dürfte. Bevor wir die Ergebnisse dieses multivariaten Modells unter inhaltlichen Gesichtspunkten diskutieren, seien am Beispiel der Tabelle 2 kurz die wesentlichen Ableitungsschritte erläutert.

Insgesamt beträgt die Rate, mit der Reichstagsmandate regulär beendet werden, 0,08888. Weniger technisch formuliert kann man sagen, ein Reichstagsmandat dauert durchschnittlich $1/0,08888$ oder 11,25 Jahre.²⁰ Die Verbesserung des diskutierten multivariaten Modells gegenüber diesem simplen Modell einer für alle Abgeordneten und Zeitpunkte konstanten Rate läßt sich mit Hilfe eines Likelihood-Verhältnis-Testes prüfen. Dazu vergleicht man die beiden maximierten Log-Likelihood-Werte, deren doppelte Differenz näherungsweise Chi-Quadrat-verteilt ist:

$$\begin{aligned} LR &= -2 (\text{Log-Likelihood a} - \text{Log-Likelihood A}) \\ &= -2 (-6020,026 - -5662,21) = 715,86. \end{aligned}$$

Die Anzahl der Freiheitsgrade entspricht der Anzahl der Parameter, die in dem weitergehenden Modell A verglichen mit dem eingeschränkten Modell a zusätzlich geschätzt werden: In diesem Fall also $df = 25$. Einige Autoren drücken diese Verbesserung auch prozentual aus [$1 - (-5662,21 / -6020,026) = 5,95\%$] und interpretieren das Ergebnis ähnlich wie das Bestimmtheitsmaß in einem klassischen linearen Regressionsmodell. Es handelt sich hierbei jedoch nicht um den Anteil erklärter Variation (daher die Bezeichnung Pseudo- R^2), zudem sind die berechneten Zahlen in der Regel sehr klein (wegen der hohen Basis der Prozentuierung) und suggerieren daher einen scheinbar schlechten Modellfit. Nach Maßgabe des Likelihood-Verhältnis-Testes ist jedoch das Modell signifikant besser als das simple Modell, das keine individuellen und zeitlichen Unterschiede der Mandatsdauern annimmt.

Um nun den Effekt der einzelnen Kovariaten einschließlich der Perioden zu testen, wurden diese Merkmale in entsprechende 1/0-kodierte Dummy-Variablen transformiert. Damit wird der Effekt jeder Ausprägung als Abwei-

²⁰ Man beachte, wie durch Berücksichtigung der zensierten Beobachtungen der Durchschnittswert gegenüber dem unkorrigierten Mittelwert in Tabelle 1 (7,16) zunimmt. Die Zahlen für das simple Modell $r(t) = \alpha$ sind nicht in Tabelle 2 enthalten.

Tabelle 2: Determinanten der Mandatsdauer

Variable	Schätzer	S.E.	P	exp b	t
Konstante:	-2,43	0,15	*	0,09	
Hauptberuf:					
Funktionäre/Journalisten	-0,29	0,12	*	0,75	-25%
Geistliche	-0,14	0,13		0,87	-13%
Freie Berufe	-0,04	0,10		0,96	-4%
Staatsdienst	0,02	0,07		1,02	2%
Rest	0,03	0,12		1,03	3%
Industrielle/Großkaufleute	0,03	0,09		1,03	3%
Handwerk/Kleinhandel	0,30	0,11	*	1,35	35%
Fraktionszugehörigkeit:					
Sozialdemokraten	-0,63	0,16	*	0,53	-47%
Zentrum	-0,20	0,08	*	0,82	-18%
Rest	0,07	0,08		1,08	8%
Rechtsliberale	0,24	0,07	*	1,27	27%
Linksliberale	0,27	0,09	*	1,31	31%
Wohnort:					
nicht im Wahlkreis	-0,05	0,05		0,95	-5%
Politische Ämter:					
nur lokale	-0,19	0,07	*	0,83	-17%
lokale und regionale Ämter	-0,12	0,09		0,89	-11%
nur regionale	-0,11	0,07		0,89	-11%
Einzelstaatl. Parlamente:					
Preuss. Abgeordnetenhaus	-0,66	0,06	*	0,52	-48%
andere einzel. Parlamente	-0,51	0,06	*	0,60	-40%
Verbandsaffiliation:					
politischer Verband	-0,53	0,10	*	0,59	-41%
nicht-pol. Verband	-0,24	0,05	*	0,78	-22%
Alter	0,019	0,002	*	1,02	2%
Beginn des ersten Mandats:					
1874-1883	-0,21	0,07	*	0,81	-19%
1884-1892	-0,27	0,08	*	0,76	-24%
1893-1902	-0,55	0,08	*	0,58	-42%
1903-1918	-0,80	0,08	*	0,45	-55%

Vergleichsgruppe: Gutsbesitzer, Konservative, Wohnort im Wahlkreis, keine politischen Ämter, keine Mitgliedschaft in einzelstaatlichen Parlamenten oder Verbänden, Mandatsbeginn 1867-73

Signifikanz $P < 0.05$: *

Modell

Log-Likelihood	-5662,21
Freiheitsgrade Df	2718
Pseudo-R ²	5,95%
Likelihood-Verhältnis	715,86
Df Likelihood-Verhältnis	25

chung von einer Vergleichsgruppe ausgedrückt. In diesem Fall besteht die Vergleichsgruppe aus den Mitgliedern der konservativen Fraktionen mit Hauptberuf Gutsbesitzer, die ihren Wohnsitz im Wahlkreis haben, keine weiteren politischen Funktionen ausüben, auch nicht Mitglied eines anderen Parlaments oder Verbands sind und ihre Reichstagskarrieren in den Jahren 1867-73 begonnen haben.²¹

Da es sich um ein log-lineares Modell handelt, können die Effekte nicht so einfach interpretiert werden wie in einem linearen Regressionsmodell. Zwar gibt das Vorzeichen des Schätzers b noch Auskunft über die Richtung des Effektes und der Absolutbetrag ist ein ungefähres Maß der Stärke des Effektes. Wenn es jedoch darum geht, die exakte Veränderung der Rate anzugeben, ist es häufig sinnvoller, folgende Umformung der Schätzgleichung zu betrachten:

$$r_i(t) = a_1 a_2^{x_{i2}} \dots a_p^{x_{ip}} \quad \text{mit } a_j = \exp b_j$$

Die Umrechnung $a_j = \exp b_j$ bezeichnet man auch als Antilogarithmus des Schätzers. a_j ist der Antilogarithmus der Regressionskonstanten. Bei 1/0-kodierten Dummy-Variablen entspricht er genau der Rate der Vergleichsgruppe im Alter von Null Jahren. Die geschätzte Rate dieser (fiktiven) Gruppe beträgt also $\exp(-2,43) = 0,09$, d.h. die durchschnittliche Mandatsdauer beträgt $1/0,09 = 11,32$ Jahre. Alle Effekte der einzelnen Merkmale werden nun relativ zur Vergleichsgruppe ausgedrückt. Der Antilogarithmus a_j entspricht der Zahl, mit der die Rate der Vergleichsgruppe multipliziert werden muß, wenn sich das entsprechende Merkmal x_j um eine Einheit erhöht. Es handelt sich also um ein multiplikatives Modell: Ein Antilogarithmus von 1 besagt, daß das entsprechende Merkmal keinen Einfluß hat, während Werte kleiner als 1 (größer als 1) negativen (positiven) Effekten entsprechen.

Wir wollen die Interpretationsmöglichkeiten an Hand des Merkmals "Fraktionszugehörigkeit" illustrieren. Schon aus den log-linearen Effekten b_j wird deutlich, daß die Sozialdemokraten von der Vergleichsgruppe in negativer Richtung (geringere Rate) und die Linksliberalen in positiver Richtung (höhere Rate) am weitesten abweichen. Die Multiplikatoren a_j (Antilogarithmen) lassen sich bei Dummy-Variablen direkt als prozentuale Abweichung

²¹ Die Wahl dieser Vergleichsgruppe ist in gewisser Weise beliebig, denn es ist egal, ob man nun die Abweichung der Abgangsrate der Sozialdemokraten von der Rate der Konservativen berechnet oder umgekehrt. Im ersten Fall sind die Konservativen die Vergleichsgruppe und im zweiten Fall die Sozialdemokraten. Der Gesamteinfluß des Merkmals "Fraktionszugehörigkeit" bleibt davon unberührt. Technisch ergibt sich die Vergleichsgruppe aus den Personen, bei denen alle Dummies den Wert Null haben.

interpretieren: Danach ist die Rate der Sozialdemokraten um $1 - 0,53 = 47\%$ geringer als die der Vergleichsgruppe, die der Linksliberalen ist um $1,31 - 1 = 31\%$ höher. Anders ausgedrückt: die durchschnittliche Mandatsdauer der sozialdemokratischen Abgeordneten ist eher lang, die der Linksliberalen eher kurz.

Ähnlich ist die Argumentation für die metrische Variable "Alter": Mit jedem zusätzlichem Lebensjahr erhöht sich die Rate der Vergleichsgruppe um den Faktor 1,02 oder um 2%. Bei einem Durchschnittsalter der gesamten Untersuchungsgruppe von 47,89 Jahren würde also die Rate der Vergleichsgruppe folgenden Wert haben:

$$\exp(-2,43) * \exp(0,019)^{47,89} = 0,09 * 1,02^{47,89} = 0,22$$

Legt man also das Durchschnittsalter zugrunde, ergibt sich eine durchschnittliche Mandatsdauer der Vergleichsgruppe von $1/0,22 = 4,52$ Jahren.

Schließlich läßt sich die Signifikanz der einzelnen Effekte mit den geschätzten (asymptotischen) Standardfehlern prüfen (Spalte S.E.). Der Quotient (Schätzer/ Standardfehler) ist näherungsweise t-verteilt, d.h. alle Werte, die größer als 2 sind, sind mindestens auf dem 5%-Niveau signifikant.

3.5 Ausblick: Alternative Modelle

Bevor wir zu den inhaltlichen Ergebnissen kommen, möchten wir abschließend noch einige weiterführende Hinweise auf mögliche alternative Spezifikationen des untersuchten Prozesses geben. So könnte man z.B. einwenden, daß Abgeordnete jeweils für eine Wahlperiode gewählt werden und es daher nicht sinnvoll ist, von einem kontinuierlichen Merkmal "Mandatsdauer" auszugehen, das faktisch nur ein Vielfaches der Wahlperiode des Reichstags sein kann. Stattdessen wäre es vielleicht sinnvoller, eine Folge von Dummy-Variablen zu modellieren, die jeweils angeben, ob die Person Mitglied des Reichstags war oder nicht: 0011010000000000 würde z.B. bedeuten, daß die Person Mitglied des dritten (1871), vierten (1874) und sechsten Reichstags (1878) war.

Methodisch gesehen handelt es sich um einen zeitdiskreten Prozeß, bei dem Veränderungen nicht jederzeit, sondern nur nach Ablauf der Wahlperiode stattfinden können. Die meisten Methoden der Verlaufsdatenanalyse wurden zunächst für kontinuierliche Daten entwickelt, aber wie Allison zeigt, kann man bekannte statische Verfahren für diskrete Daten (z.B. logistische Regression) verwenden, um dynamische Modelle für diskrete Wartezeiten zu

testen.²² Wahrscheinlich werden die Ergebnisse eines solchen zeitdiskreten Modells nicht wesentlich von den hier diskutierten zeitkontinuierlichen Modellen abweichen. Eine entsprechende Auswertung ist in Vorbereitung.

Für die Wahl des zeitkontinuierlichen Modells sprach, daß das zeitdiskrete Modell Abweichungen von dem periodischen Wahlzyklus nur sehr schlecht handhaben kann. Diese treten dann auf, wenn eine Person nachträglich in einen Reichstag eintritt oder frühzeitig wieder austritt. Nachträgliche Eintritte und frühzeitige Austritte müssten dann wie reguläre Mitgliedschaften über die gesamte Wahlperiode behandelt werden.

4. Zusammenfassung und Interpretation der Ergebnisse

Ausgehend von der Hypothese einer zunehmenden "Professionalisierung" und "Oligarchisierung" des Reichstages interessierte zunächst die Entwicklung der Mandatsdauern im historischen Zeitablauf. Im multivariaten Modell unter Kontrolle mehrerer anderer Merkmale (vgl. Tabelle 2) zeigt sich eindeutig, daß die Abgangsraten über alle fünf Zeitperioden mehr oder weniger kontinuierlich sanken. Verglichen mit den Mandaten, die in den ersten drei Reichstagen begannen (Periode 1867- 1873), ist die Abgangsrate in der folgenden Periode (1874- 1883) um 19% und in der letzten Periode 1903 bis 1918 sogar um 55% geringer (1884 bis 1892: - 24%, 1893 bis 1902: - 42%). Alle vier Effekte sind hoch signifikant. Unabhängig von allen anderen Erklärungsfaktoren kann man also davon ausgehen, daß die durchschnittliche Mandatsdauer im Zeitablauf erheblich zunahm. Anders formuliert: die Mandatsverhaftung der Abgeordneten verstärkte sich in direkter Abhängigkeit vom "Systemalter" des Reichstags. Dieser Befund unterstützt Michels Oligarchisierungsthese auf der Aggregatebene: Je länger Führungsgruppen etabliert sind, desto niedriger ist ihre Zirkulationsrate.

Zu beachten ist bei dieser Deutung allerdings, daß das Wiederwahlrisiko bei jenen Abgeordneten, deren Mandate in den ersten Reichstagen begannen, größer als in den letzten Jahrzehnten des Kaiserreichs war, da die durchschnittliche Dauer der Legislaturperioden zunahm. Institutionelle Veränderungen unterstützten deshalb vermutlich den Trend zur Mandatsverlängerung.²³

Andererseits haben wir zeigen können, daß sich auf der individuellen

²² P.D. Allison, *Discrete-time Methods for the Analysis of Event-Histories*, in: S. Leinhardt, (Hrsg.), *Sociological Methodology 1982* (San Francisco 1982), S. 61-98.

²³ Zwischen 1867 und 1887 traten 8 Reichstage, zwischen 1887 und 1907 traten 5 Reichstage zusammen.

Ebene die Abgangsrate im Zeitablauf nicht veränderte. Wir vermuten hier gegenläufige Wirkungen wachsender Erfahrungen und zunehmenden physischen Alters. Tatsächlich wird die Mandatsverhaftung auf natürliche Weise durch die Lebenserwartung begrenzt. Wie erwartet hat daher das Lebensalter der Abgeordneten einen signifikant positiven Effekt auf die Abgangsrate beziehungsweise einen negativen Effekt auf die Mandatsdauer.

Webers Abkömmlichkeitsthese bestätigte sich nur für zwei Kategorien unserer Berufsklassifikation: hauptberufliche Partei- und Verbandsfunktionäre, die auch in einem ökonomischen Sinne "aus der Politik einen Beruf machen" konnten, hatten eine gegenüber der Vergleichsgruppe (Gutsbesitzer) signifikant geringere Abgangsrate, während Kleingewerbetreibende, die extrem betriebsgebunden waren und kaum Delegationschancen hatten, eine signifikant geringere Mandatsverhaftung aufwiesen. Dagegen beobachteten wir im Fall der Industriellen und Großkaufleute, bei denen Max Weber eine besonders geringe Abkömmlichkeit unterstellt hatte, keine signifikante Abweichung der Mandatsverhaftung von der Vergleichsgruppe. Vermutlich unterschätzte Weber die Chancen zur Delegation von Unternehmerfunktionen und die Entlastungswirkung, die von der Verfügung über einen "Apparat" ausging, der auch für politische Ambitionen genutzt werden konnte.

Auch der Einfluß regionaler Bindungen auf die Mandatsverhaftung war gering. Zwar hatten Abgeordnete, die nicht in ihrem Wahlkreis wohnten, eine etwas geringere (- 5%) Abgangsrate als Abgeordnete, die in ihrem Wahlkreis ansässig waren, jedoch ist der Unterschied zwischen diesen beiden Gruppen nicht signifikant. Es scheint, daß sich der kostenmindernde Effekt einer leichteren Anreise und der stimmenzehrende Effekt einer mangelnden Präsenz vor Ort im Wahlkreis gegeneinander aufhoben und damit die Lage des Wohnorts des Abgeordneten faktisch ohne Einfluß war.

Insgesamt ist der Effekt jener Variablen, die sich einem abkömmlichkeitstheoretischen Modell zuordnen lassen, gering. Dafür können zwei Gründe maßgebend sein: zum einen ist es möglich, daß die selektiven Mechanismen der Abkömmlichkeit bereits im Vorfeld der Kandidatenauswahl wirksam wurden, daß also tendenziell nur solche Personen zur Kandidatur bereit waren oder aufgefordert wurden, die "privaten" Beruf und parlamentarisches Mandat miteinander verbinden konnten oder - wie etwa bei den SPD-Abgeordneten - eine gewisse Kompensation durch "Partei-Diäten" erwarten durften.²⁴ Eine andere Erklärung ist die im Kaiserreich eher geringe Bean-

²⁴ Vgl. Theodor Eschenburg, *Der Sold des Politikers*, Stuttgart 1959, S. 56 u. Wilhelm u. Wolfgang Treue, *Parlamentarismus in Deutschland. Entstehung und Entwicklung* (2. Aufl., Bonn 1963) S. 21 f.

spruchung durch das Mandat. Tatsächlich berichten zeitgenössische Beobachter immer wieder von der geringen Teilnahme an den Sitzungen des Reichstages, der oft nur deshalb handlungsfähig blieb, weil stillschweigend darauf verzichtet wurde, die Beschlußfähigkeit formal festzustellen.²⁵ Auszählungen der Anwesenheit bei namentlichen Abstimmungen bestätigen solche Beobachtungen, die darauf hindeuten, daß das Konstrukt der Mandatsverhaftung auf zweierlei Weise operationalisiert werden kann: zum einen durch die Dauer, zum anderen durch die Intensität, mit der das Mandat ausgeübt wird.

Überzeugender als Max Webers Abkömmlichkeits-Argument wird Michels Oligarchisierungsthese im Hinblick auf horizontale und vertikale Ämterkumulation durch unsere Daten unterstützt. Ein Ergebnis des multivariaten Modells ist, daß Abgeordnete, die Mitglied des preußischen Abgeordnetenhauses waren, signifikant längere Mandatszeiten hatten, was sich in einer um 48% geringeren Abgangsrate ausdrückt. Das gleiche gilt für Mitgliedschaften in anderen einzelstaatlichen Parlamenten, die die Rate um 40% verringerten. Man kann sogar sagen: neben den Periodeneffekten hatten parallele Kammermitgliedschaften den größten Einfluß auf die Mandatsdauer. Ähnliches gilt auch für die vertikale Kumulation von Funktionen in politischen Vereinen und Verbänden. Die Ergebnisse unseres multivariaten Modells zeigen deutlich einen stabilisierenden Einfluß von Verbandsaffiliationen. Insbesondere für die Abgeordneten, die Mitglied eines politischen Verbandes waren, ist eine signifikant niedrigere Abgangsrate festzustellen (- 41 %, dagegen bei nicht-politischen Verbänden nur -22%). Bei den Funktionen in der kommunalen Selbstverwaltung zeigt sich dagegen nur ein signifikanter Effekt für die Ämter auf lokaler Ebene (Abgangsrate - 17%), während Ämter auf regionaler Ebene tendenziell auch die Abgangsrate verringerten (- 11 %), dieser Effekt jedoch nicht groß genug ist, um unter statistischen Kriterien als signifikant gelten zu können.

Betrachtet man schließlich die Abhängigkeit der Mandatsverhaftung von der Fraktionszugehörigkeit, dann zeigt sich, daß Sozialdemokraten und Zentrumsabgeordnete deutlich über, Rechts- und Linksliberale deutlich unter der Vergleichsgruppe (Konservative) liegen. Insbesondere die überdurchschnittlich lange Mandatsdauer der Sozialdemokraten (ausgedrückt in einer signifikant niedrigeren Abgangsrate) ist bemerkenswert.²⁶ Dieses Ergebnis läßt sich

²⁵ Peter Molt, *Der Reichstag*, S. 48.

²⁶ Gerhard A. Ritter u. Mitarb. v. Merith Niehuss, *Wahlgeschichtliches Arbeitsbuch. Materialien zur Statistik des Kaiserreichs* (München 1980), Abb. 7, S. 55.

gut mit der Entwicklung des "Unterstützungsmarktes" in der Wählerschaft für die Sozialdemokraten erklären, deren Stimmenanteil zwischen 1871 und 1912 von 3,1% auf 34,8% und deren Sitzanteil im gleichen Zeitraum von 0,5% auf 27,7% zunahm. Ähnliches gilt für das Zentrum, das seit der Konsolidierung seines Wählerstammes im Jahr 1874 mit bemerkenswerter Konstanz zwischen 23% und 27% der Reichstagsabgeordneten stellte. Demgegenüber nahmen die Mandatsanteile der Liberalen bei zum Teil heftigen Fluktuationen langfristig ab. Dies waren keine günstigen Bedingungen für die Verstetigung parlamentarischer Karrieren liberaler Politiker. Möglicherweise war die geringe und relativ sinkende Mandatsverhaftung der Liberalen, neben der verfassungstechnisch erzwungenen Trennung von Reichsregierung und Reichstag, ein weiterer Grund für den von Max Weber beklagten Qualitätsverfall bürgerlicher Politik und die "politische Nichtigkeit des Parlaments und der Parteipolitiker" im ausgehenden Kaiserreich.²⁷ Noch in der Gründungsphase des Reiches hatten zumindest die Nationalliberalen eine im Verhältnis zur Vergleichsgruppe stärkere Mandatsverhaftung gehabt.

Mit dieser Beobachtung weiten wir die Betrachtung zu einem Zeitvergleich aus, in dem der Effekt einzelner Merkmale auf die Mandatsverhaftung nach verschiedenen historischen Perioden differenziert wird (vgl. Tab. 3a - d). Ohne daß wir hier auf Einzelheiten eingehen können, soll nur auf einen generellen Trend hingewiesen werden: Der Einfluß institutioneller Bindungen der Abgeordneten, seien es nun Funktionen in der kommunalen Selbstverwaltung, Verbands- und Vereinsaffiliationen oder Mitgliedschaften in einzelstaatlichen Parlamenten gewesen, nimmt im Zeitverlauf ab. Nach 1903 hatte keine der Positionen im Geflecht vertikaler oder horizontaler Ämterkumulation mehr eine signifikante Wirkung auf die Mandatsverhaftung. Dieser Befund darf nun keineswegs als ein Hinweis auf eine abnehmende "Oligarchisierung" mißdeutet werden, denn tatsächlich nahm ja, wie die durchgängig signifikanten Periodeneffekte zeigen, die Mandatsverhaftung im Kaiserreich zu. Wir interpretieren diesen Befund dagegen als Ausdruck eines Bedeutungsgewinns der Parteien, die die Funktion der Kandidatenselektion und -unterstützung zunehmend monopolisierten, während die parteiunabhängigen Unterstützungsinstanzen und -zirkel zurücktraten. Gefördert wurde diese Tendenz durch die Gewährung von Diäten für die Reichstagsmitglieder seit 1906. Somit würde sich auch dieser historische Wandel in den allgemeinen Trend einer "Oligarchisierung" der "mannigfaltigen Formen der Demokratie" einpassen, den R. Michels konstatierte, und der sich unter den Stichworten der

²⁷ Max Weber, "Parlament und Regierung im neugeordneten Deutschland", in: ders., Gesammelte politische Schriften (zuerst 1918), S. 313.

"Institutionalisierung" und "Professionalisierung" mit nur geringfügig erweitertem Begriffsgehalt auch in aktuellen Theorien politischer Modernisierung wiederfindet.²⁸

Tabelle 3: Historische Veränderungen einzelner Variablen

Tabelle 3a: Regionalstaatliche Parlamente

Variable	1867	1874	1884	1893	1903
Periode		-0,29* (0,09)	-0,34* (0,10)	-0,77* (0,11)	-1,08* (0,11)
Preuss. Abgeordnetenhaus	-0,78* (0,13)	-0,78* (0,13)	-0,83* (0,14)	-0,48* (0,16)	-0,27 (0,15)
andere einzel. Parlamente	-0,86* (0,12)	-0,55* (0,11)	-0,55* (0,14)	-0,21 (0,15)	-0,06 (0,15)

Alle anderen Variablen kontrolliert, Schätzer nicht extra aufgeführt

Log-Likelihood	-5648,7	Likelihood-Verhältnis	742,65
Freiheitsgrade Df	2710	Df Likelihood-Verhältnis	33
Pseudo-R ²	6,17%		

Vergleichsgruppe: s. Tabelle 2

Standardfehler in Klammern, Signifikanz $P < 0.05$: *

²⁸ Vgl. die Literaturangaben in Anm. 5 und 8.

Tabelle 3b: Verbandsaffiliation

Variable	1867	1874	1884	1893	1903
Periode		-0,19* (0,08)	-0,26* (0,09)	-0,65* (0,11)	-1,02* (0,12)
nicht-politischer Verband	-0,32* (0,11)	-0,41* (0,11)	-0,28* (0,11)	-0,08 (0,13)	0,04 (0,13)
politischer Verband	-0,58* (0,21)	-0,55* (0,22)	-0,97* (0,29)	-0,53* (0,23)	-0,14 (0,18)
Alle anderen Variablen kontrolliert, Schätzer nicht extra aufgeführt					
Log-Likelihood	-5655,2	Likelihood-Verhältnis			729,61
Freiheitsgrade Df	2710	Df Likelihood-Verhältnis			33
Pseudo-R ²	6,06%				

Vergleichsgruppe: s. Tabelle 2

Standardfehler in Klammern, Signifikanz $P < 0.05$: *

Tabelle 3c: Politische Ämter

Variable	1867	1874	1884	1893	1903
Periode		-0,12 (0,08)	-0,29* (0,10)	-0,59* (0,11)	-0,85* (0,11)
nur lokale	-0,24 (0,15)	-0,27* (0,14)	-0,14 (0,14)	-0,16 (0,17)	-0,05 (0,16)
nur regionale	0,07 (0,14)	-0,53* (0,16)	-0,01 (0,15)	0,08 (0,17)	-0,13 (0,19)
lokale und regionale Ämter	-0,42 (0,34)	-0,30 (0,20)	-0,06 (0,18)	-0,05 (0,18)	0,06 (0,17)
Alle anderen Variablen kontrolliert, Schätzer nicht extra aufgeführt					
Log-Likelihood	-5654,6	Likelihood-Verhältnis			730,92
Freiheitsgrade Df	2706	Df Likelihood-Verhältnis			37
Pseudo-R ²	6,07%				

Vergleichsgruppe: s. Tabelle 2

Standardfehler in Klammern, Signifikanz $P < 0.05$: *

Tabelle 3d: Fraktionszugehörigkeit

Variable	1867	1874	1884	1893	1903
Periode		-0,43* (0,15)	-0,48* (0,15)	-0,56* (0,15)	-0,85* (0,16)
Sozialdemokraten	-0,40 (0,59)	0,38 (0,33)	-0,82* (0,38)	-1,10* (0,31)	-0,62* (0,24)
Linksliberale	0,04 (0,18)	0,43* (0,16)	0,47* (0,17)	0,26 (0,20)	0,07 (0,22)
Rechtsliberale	-0,11 (0,14)	0,48* (0,14)	0,44* (0,16)	0,21 (0,19)	0,22 (0,19)
Zentrum	-0,24 (0,22)	-0,21 (0,17)	0,02 (0,16)	-0,31 (0,19)	-0,26 (0,19)
Rest	0,21 (0,13)	0,01 (0,18)	-0,09 (0,21)	-0,13 (0,21)	0,05 (0,21)

Alle anderen Variablen kontrolliert, Schätzer nicht extra aufgeführt

Log-Likelihood	-5641,7	Likelihood-Verhältnis	756,66
Freiheitsgrade Df	2698	Df Likelihood-Verhältnis	45
Pseudo-R ²	6,28%		

Vergleichsgruppe: s. Tabelle 2

Standardfehler in Klammern, Signifikanz $P < 0.05$: *

Tabelle 3e: Hauptberuf

Variable	1867	1874	1884	1893	1903
<i>Periode</i>		-0,29* (0,14)	-0,14 (0,14)	-0,34* (0,15)	-0,77* (0,16)
<i>Industrielle/Großkaufleute</i>	0,29 (0,19)	-0,07 (0,18)	-0,06 (0,16)	-0,28 (0,21)	0,36 (0,21)
<i>Handwerk/Kleinhandel</i>	0,46 (0,30)	0,66* (0,25)	0,13 (0,23)	-0,12 (0,23)	0,46* (0,23)
<i>Staatsdienst</i>	0,07 (0,13)	0,21 (0,13)	-0,16 (0,15)	-0,19 (0,18)	0,08 (0,17)
<i>Geistliche</i>	0,43 (0,31)	0,07 (0,23)	-0,67 (0,35)	-0,05 (0,26)	-0,70 (0,38)
<i>Freie Berufe</i>	-0,14 (0,20)	0,28 (0,17)	0,08 (0,21)	-0,47 (0,28)	-0,14 (0,23)
<i>Funktionäre/Journalisten</i>	0,22 (0,27)	0,17 (0,21)	-0,60* (0,32)	-0,41 (0,26)	-0,45* (0,22)
<i>Rest</i>	-0,01 (0,23)	0,09 (0,21)	0,16 (0,25)	-0,21 (0,37)	0,21 (0,38)

Alle anderen Variablen kontrolliert, Schätzer nicht extra aufgeführt

<i>Log-Likelihood</i>	-5642,5	<i>Likelihood-Verhältnis</i>	755,13
<i>Freiheitsgrade Df</i>	2690	<i>Df Likelihood-Verhältnis</i>	53
<i>Pseudo-R²</i>	6,27%		

Vergleichsgruppe: s. Tabelle 2

Standardfehler in Klammern, Signifikanz $P < 0.05$: *

Software zur Verlaufsdatenanalyse

Analysemethode	SPSS-X SPSS (Vers. 8)	RATE Version 2	BMDP Version 1981	SAS Version 5	GLIM Vers. 3 & 3.77	LIMDEP Vers. 5	GAUSS 2.0 Appl. Module
----------------	--------------------------	-------------------	----------------------	------------------	------------------------	-------------------	---------------------------

a) Beschreibung							
Schätzverfahren	LIFE	-	LIFE, K&M	LIFE, K&M	*	LIFE, K&M	*
verfügbare Plots	SURV, DEN, HAZ	-	SURV, DEN, HAZ	SURV, DEN, HAZ	*	SURV, HAZ, CUM	*
graphische Tests	LOG	-	LOG, CUM	LOG, LOG(-LOG)	*	*	*
b) Gruppenvergleiche							
Verfügbare Tests	WILCOXON	-	WILCOXON, SAVAGE	WILCOXON, SAVAGE, LRATIO	*	*	*
c) Regressionsmodelle							
Partiell	-	COX	COX	COX	COX	COX	-
parametrisch	-	EXP, GOMPERTZ	EXP, WEIBULL, LOGN, EXP, WEIBULL, LOGLOG, GAMMA	EXP, WEIBULL, LOGN, EXP, WEIBULL, LOGLOG, GAMMA	EXP, WEIBULL, LOGN, EXP, GAMMA, LOGLOG, GOMPERTZ	EXP, WEIBULL, LOGN, EXP, GAMMA, LOGLOG, GOMPERTZ	EXP, GAMMA, PARETO
sonstige	-	ERROR, PERIOD	COXTIME	COXTIME	ERROR	COXTIME	-
schnittweise Re-	-	-	STEPUP	STEPUP, SUBSET	ERROR	COXTIME	-
gression	-	-	-	-	ERROR	COXTIME	*
d) Modellevaluation							
Prognosen	-	-	Ja	Ja	*	*	*
Annahmetests	-	-	PROPRISK	PROPRISK	*	*	*
Residuenanalyse	-	-	Ja	Ja	*	*	*
Erläuterung							
*) Keine spezielle Prozedur verfügbar aber programmierbar durch benutzereigene Macros.							
LIFE	Sterbetafelschätzer	WILCOXON	verallg. M. Test	GAMMA	Gamma-Verteilung		
K&M	Kaplan-Meier-Schätzer	SAVAGE	verallg. S. Test	LOGLOG	log-logistische Vert.		
SURV	Überlebensfunktion	LRATIO	Likelihood Verhältnis Test	ERROR	Modell mit Fehlerterm		
DEN	Dichtefunktion	COX	COX's Regressionsmodell	PERIOD	m. periodenspez. Rate		
HAZ	Risikofunktion	EXP	Exponentialverteilung	COXTIME	zeitabh. Kovarianzen		
LOG	In(SURV)	GOMPERTZ	Gompertz-Verteilung	STEPUP	schriftw. Einschlag		
CUM	kumulierte HAZ	WEIBULL	Weibull-Verteilung	LOGN	Lognormal-Verteilung		
				PROPRISK	proportionale Risiken		

Grundlagen, Konzepte und neue Methoden der Zeitreihenanalyse

von Winfried Stier

Einleitung

Im folgenden sollen grundlegende Konzepte und Werkzeuge der Zeitreihenanalyse in möglichst einfacher Form dargestellt werden. Deshalb wird bewußt auf mathematische Ableitungen und Feinheiten verzichtet. Formeln werden nur insoweit eingeführt, als sie zum Verständnis unverzichtbar sind. Die Darstellung ist für Leser gedacht, die keine oder nur rudimentäre Vorkenntnisse in Zeitreihenanalyse besitzen. Selbstverständlich kann hier kein Gesamtüberblick über dieses mittlerweile riesige Teilgebiet der Statistik gegeben werden. Was jeweils über die Grundlagen hinaus zur Darstellung kommt, ist mehr oder weniger das Resultat subjektiver Auswahl des Autors und/oder der Herausgeber.

I. Traditionelle Zeitreihenmodelle

1. Grundlagen

Die Zeitreihenanalyse spielt seit jeher in der empirischen Wirtschaftsforschung eine bedeutende Rolle. Von daher stammt auch das sog. "Komponentenmodell", das für traditionelle Zeitreihenanalysen häufig die Grundlage bildet. Dabei wird unter einer Zeitreihe eine Folge von zeitlich geordneten, äquidistanten Beobachtungswerten eines mindestens auf Intervallskalenniveau gemessenen Merkmals verstanden. Grundlegend für das Komponentenmodell ist die Vorstellung, daß sich die beobachteten Reihenwerte aus verschiedenen "Bewegungskomponenten" zusammensetzen. Üblicherweise werden vier derartige Komponenten unterschieden:

a) Eine "Trendkomponente" T_t , deren Verlauf als durch langfristig wirkende Ursachen bedingt angesehen wird. Oft wird unterstellt, daß sie monoton wächst (z.B. auf Grund des technischen Fortschritts) oder monoton fällt (als Folge z.B. eines Bevölkerungsrückgangs).

b) Eine "Zyklische Komponente" Z_t , deren Verlauf den Konjunkturzyklus reflektiert und für die deshalb eine "wellenförmige" Bewegung postuliert wird.

c) Eine "Saisonkomponente" S_t , deren Verlauf auf jahreszeitliche und institutionelle Ursachen zurückgeführt wird. Für sie wird ebenfalls ein "wellenförmiger" Verlauf angenommen.

d) Und schließlich eine "irreguläre Komponente" U_t , deren Verlauf nicht auf die bei den anderen Komponenten aufgeführten Ursachenkomplexen zurückgeführt werden kann. Es wird angenommen, daß die U_t -Werte relativ (d.h. im Vergleich zu den T_t -, Z_t -, und den S_t -Werten) klein sind und "regellos", um den Wert Null schwanken. Die U_t sind Residualgrößen, die den Charakter von "Zufallsschwankungen" aufweisen sollen.

Häufig werden die Komponenten T_t und Z_t nicht getrennt betrachtet, sondern zur sog. "glatten Komponente", G_t , zusammengefaßt. Dies hängt nicht zuletzt damit zusammen, daß eine Trennung dieser beiden Komponenten mit traditionellen zeitreihenanalytischen Werkzeugen als äußerst problematisch anzusehen ist.

Die eben skizzierte inhaltliche Deutung der Komponenten reicht natürlich für eine Analyse nicht aus. Dazu ist mindestens eine Vorstellung darüber erforderlich, wie die einzelnen Komponenten im Zeitablauf zusammenwirken. Das einfachste Modell postuliert eine additive Überlagerung dieser Komponenten. Für die Zeitreihe X_t gilt mit diesem Postulat somit:

$$X_t = T_t + Z_t + S_t + U_t, \quad t = 1, 2, \dots, T$$

Der Index t bezieht sich dabei auf vorgegebene Zeitperioden, für die das Beobachtungsmaterial vorliegt. Diese sind meistens Monate, Quartale, Halbjahre. Aber auch Tageswerte (z.B. bei Börsenkursen) oder Wochenwerte finden sich in der Praxis. Für Jahreswerte hat die Komponente S_t natürlich keinen Sinn, da saisonale Schwankungen per Definition nur bei unterjährig Daten existieren. Hier sind nur drei Komponenten sinnvoll.

Statt einer additiven Verknüpfung wird häufig von einer multiplikativen Komponentenverknüpfung ausgegangen, so z.B. bei einem Saisonbereinungsverfahren, das als "Census-X-11", bekannt ist. (Auf derartige Verfahren wird gleich noch näher einzugehen sein). Formal kann das multiplikative Modell durch Logarithmierung auf ein additives zurückgeführt werden. Neben den "reinen Typen" sind auch Mischformen denkbar, so z.B. wenn angenommen wird, daß sich die saisonale Komponente proportional zur glatten Komponente entwickelt.

2. Probleme der Zeitreihenanalyse

An obigem traditionellen Komponentenmodell lassen sich auf einfache Weise eine Reihe von Fragestellungen und Probleme diskutieren, die prinzipiell auch bei komplizierteren Zeitreihenansätzen in gleicher oder ganz ähnlicher Weise auftreten. Ein wichtiges praktisches Problem ist die sog. "Saisonbereinigung". Dabei soll die Komponente S_t identifiziert und eliminiert werden. Es ist unmittelbar einleuchtend, daß saisonbereinigte Reihen in der Praxis eine wichtige Rolle spielen. Ist z.B. eine Zunahme der Zahl der Arbeitslosen ein Indiz für einen sich tendenziell verschlechternden Arbeitsmarkt oder ist diese nur jahreszeitlich bedingt, also kurzfristiger Natur? Um diese und ähnliche Fragen beantworten zu können, muß eine Reihe um die Saisonkomponente "bereinigt" werden. Dies geschieht in der Hoffnung, daß mit Hilfe der bereinigten Reihe die Frage beantwortet werden kann, wie die Entwicklung verlaufen wäre, wenn eben keine jahreszeitlichen Einflüsse wirksam gewesen wären.

Eine andere Frage wäre z.B. die nach der "langfristigen" Entwicklung einer Reihe, d.h., wie würde sich eine Reihe ohne alle kurzfristigen Einflüsse entwickeln. Zur Beantwortung dieser Frage muß neben der saisonalen auch die irreguläre Komponente eliminiert werden. Diese Frage ist natürlich auch sinnvoll für Jahresreihen oder für Reihen mit mehrjährigem Rhythmus (z.B. Fünfjahresreihen), wie sie etwa in der historischen Forschung häufig vorkommen. Hier interessiert man sich z.B. für einen "säkularen" Trend. Oder dafür, ob eine Reihe eine zyklische Komponente enthält und falls ja, welche Zykluslänge vorliegt. Offensichtlich ist das letztere Problem wesentlich komplizierter als die vorher erwähnten, denn die Isolation einer zyklischen Komponente setzt eine Elimination aller übrigen Komponenten voraus. Wesentlich ist dabei, daß deren Elimination in einer Weise erfolgt, daß die zyklische Komponente nicht tangiert wird. Auf diesen Problembereich wird noch zurückzukommen sein.

Wie diese Ausführungen zeigen, besteht ein wichtiges Ziel der Zeitreihenanalyse darin, gewisse Komponenten einer Reihe zu isolieren oder zu eliminieren. Die Frage stellt sich deshalb, welche Werkzeuge die traditionelle Zeitreihenanalyse zur Verfügung stellt, um die angesprochenen Probleme zu lösen.

3. Werkzeuge der traditionellen Zeitreihenanalyse

In diesem Abschnitt sollen einige Werkzeuge der traditionellen Zeitreihenanalyse besprochen werden. Diese spielen durchaus auch heute noch eine bedeutende Rolle in der Praxis, allerdings häufig in modifizierter und/oder verfeinerter Form. Ihr Einsatz läßt sich am einfachsten im Kontext eines konkreten zeitreihenanalytischen Problems zeigen. Insbesondere können dabei sowohl theoretische als auch praktische Probleme, die mit ihrem Einsatz verknüpft sind, anschaulich diskutiert werden.

Ausgegangen werden soll vom oben schon ansatzweise diskutierten Problem der Saisonbereinigung, wobei das traditionelle, additive Komponentenmodell zugrundegelegt wird. Ohne Beschränkung der Allgemeinheit wird von monatlichen Reihen ausgegangen. Dieses Komponentenmodell allein reicht jedoch nicht aus, um die gestellte Aufgabe zu lösen. Vielmehr ist es erforderlich, für die einzelnen Komponenten bestimmte Verlaufshypothesen zu formulieren:

Es sei angenommen, daß

$$a) \quad S_t = S_{t+12}, \quad t=1,2,\dots,T$$

ist, d.h. eine streng periodische Saisonfigur wird unterstellt.

b) die glatte Komponente innerhalb eines Zeitraumes von 13 Monaten durch eine lineare Funktion der Zeit hinreichend genau approximiert werden kann.

Unter diesen beiden Voraussetzungen und der weiteren, daß die Saisonfigur auf Null normiert wird, d.h., daß gilt $S_1 + S_2 + \dots + S_{12} = 0$, kann die glatte Komponente mit Hilfe eines gleitenden symmetrischen 12-Monats-Durchschnitts "geschätzt" werden:

$$\hat{G} = \frac{1}{12} \left[\frac{1}{2} X_{t-1} + \frac{1}{2} X_{t+1} + \sum_{i=t-5}^{t+5} X_i \right]$$

Die Normierung der Saisonfigur auf die Summe Null bedeutet keine Einschränkung der Allgemeinheit des Modellansatzes, da ja im Fall einer von Null verschiedenen Summe eine Konstante von den S-Werten subtrahiert und der glatten Komponente zugeschlagen werden könnte.

Problematisch sind allerdings die Postulate a) und b). Abgesehen davon, daß die postulierte strenge Periodizität von S_t völlig unrealistisch ist (wie man durch Betrachten konkreter Reihen leicht erkennt), bleibt anzumerken, daß das benutzte Werkzeug zur Isolation der glatten Komponente eine ganz bestimmte Verlaufsform für G_t impliziert. Es kann nämlich gezeigt werden, daß dieser gleitende Durchschnitt die glatte Komponente nur dann fehlerfrei erfaßt, wenn diese tatsächlich linear im angegebenen "Stützbereich" von 13 Monaten verläuft.

Dieses Beispiel zeigt sehr anschaulich, daß das verwendete Instrumentarium (gleitende Durchschnitte) eigentlich eine dazu "passende" Definition der zu erfassenden Komponente voraussetzt. Es wird zwar keine "globale" Linearität (d.h. für alle Zeitpunkte) vorausgesetzt, wohl aber eine (weniger rigide) "lokale" Linearität (d.h. für jeweils 13 Monate). Verläuft nun G_t nicht linear innerhalb der 13 Monate-Stützbereiche, dann kann die glatte Komponente offensichtlich nicht "sauber" isoliert werden. Die Fehler sind umso gravierender, je weniger die Verlaufshypothese zutrifft. Deswegen wurde in b) schon vorsichtig von einer "hinreichenden" Genauigkeit gesprochen, ohne daß diese allerdings präzisiert wurde.

Methodologisch unbefriedigend ist insbesondere der Sachverhalt, daß bei diesem Verfahren keine operationale Definition der glatten Komponente a priori gegeben wird, diese sich vielmehr a posteriori aus dem verwendeten Analysewerkzeug ergibt. Eigentlich müßte es ja gerade umgekehrt sein: das Analysewerkzeug müßte so konstruiert werden, daß es in "optimaler" Weise der auf substanzwissenschaftlichen Überlegungen beruhenden Komponentendefinition gerecht wird. Diese Problematik wird unten nochmals zu diskutieren sein.

Im Vorgriff auf spätere Darlegungen sei hier schon erwähnt, daß gleitende Durchschnitte als "Filter" interpretiert werden können. Mit Hilfe filtertheoretischer Methoden kann gezeigt werden, daß symmetrische gleitende Durchschnitte die Eigenschaft der sog. "Nullphasigkeit" haben, d.h. die ausgefilterten Werte werden den richtigen Zeitpunkten zugeordnet, so daß sich keine Verzögerungen zwischen der Originalreihe und der glatten Komponente ergeben. Dabei ist jedoch auch sogleich auf einen praktisch gravierenden Nachteil gleitender Durchschnitte aufmerksam zu machen: sowohl am Reihenanfang als auch am Reihende gehen Werte verloren (bei obigem gleitenden Durchschnitt je 6). Dies impliziert, daß die glatte Komponente insbesondere nicht bis zum "aktuellen Rand" einer Reihe bestimmt werden kann. Da sich das Interesse aber in der Praxis gerade häufig auf diesen

Randbereich konzentriert (Verwendung bereinigter Reihen zu Diagnosezwecken), ist dieser Sachverhalt besonders schwerwiegend. Man könnte deshalb auf die Idee kommen, asymmetrische gleitende Durchschnitte zu verwenden, die außer dem jeweiligen aktuellen Reihenwert nur vergangene Reihenwerte verarbeiten. Formal kommt man damit bis zum aktuellen Rand, jedoch weisen solche gleitende Durchschnitte eine (negative) Phase auf, d.h. der letzte gefilterte Wert ist nicht dem letzten Zeitpunkt zuzuordnen, sondern einem früheren. Je nach Konstruktion eines asymmetrischen gleitenden Durchschnitts ist das Ausmaß dieser Phasenverschiebung mehr oder weniger praktisch bedeutsam. Die sich hier abzeichnende "Randproblematik" tritt nicht nur bei traditionellen Werkzeugen der Zeitreihenanalyse auf.

Die saisonbereinigte Reihe ergibt sich schließlich mit Hilfe einiger einfacher Operationen. Da \hat{G}_t eine "Schätzung" für G_t ist, kann geschrieben werden:

$$X_t - \hat{G}_t \approx S_t + U_t.$$

Die $S_t + U_t$ - Werte werden nun über gleichnamige Monate gemittelt. Dazu ist eine Doppelindizierung zweckmäßig: X_{ij} sei der Reihenwert des j -ten Monats im i -ten Jahr und $S_{ij} = S_j$ auf Grund der postulierten Periodizität. Dann ist:

$$\bar{S}_j = \frac{1}{n_j} \sum_i (X_{ij} - \hat{G}_{ij}) \approx S_{ij} + \frac{1}{n_j} \sum_i U_{ij}, \quad j = 1, 2, \dots, 12$$

n_j bedeutet die Anzahl der Jahre für die S_j berechenbar ist. Die Normierung der 12 Werte S_1, S_2, \dots, S_{12} erfolgt durch Subtraktion ihres arithmetischen Mittels:

$$\hat{S}_j = \bar{S}_j - \frac{1}{12} \sum_{j=1}^{12} \bar{S}_j, \quad j = 1, 2, \dots, 12$$

Die saisonbereinigte Reihe ergibt sich schließlich durch Bildung der Differenzen $X_{ij} - \hat{S}_j$.

Schließlich sei noch angemerkt, daß dieses einfache Verfahren in den dreißiger Jahren vom Federal Reserve Board (USA) und vom Londoner Wirtschaftsdienst für praktische Analysen verwendet wurde.

Später entwickelte und heute gebräuchliche Saisonbereinigungsverfahren stellen natürlich vergleichsweise wesentlich raffiniertere und kompliziertere

Konstruktionen dar, obwohl sie in der Regel immer noch auf dem ursprünglichen Komponentenmodell beruhen. Diese Konstruktionen unterscheiden sich vom obigen einfachen (und ähnlichen hier nicht besprochenen) Verfahren in verschiedener Hinsicht:

a) Zum einen werden a priori bestimmte Verlaufshypothesen für die einzelnen Komponenten eingeführt, die als flexibler bezeichnet werden können und für die dann jeweils geeignete Schätzverfahren verwendet oder entwickelt werden. Wenn hier von Schätzverfahren gesprochen wird, dann ist damit gemeint, daß es sich um Prozeduren im Sinne der statistischen Schätztheorie handelt. Dies setzt voraus, daß für das traditionelle Zeitreihenmodell eine "geeignete" stochastische Spezifikation formuliert wird, was in der Regel dadurch erfolgt, daß U_t nunmehr als eine Folge von Zufallsvariablen (d.h. als stochastischer Prozeß, dazu siehe unten) interpretiert wird. Sowohl die glatte als auch die saisonale Komponente wird nach wie vor als deterministische Funktion der Zeit aufgefaßt (über Ausnahmen siehe unten).

b) Zum anderen wird neben gleitenden Durchschnitten vor allem die Methode der Kleinsten-Quadrate verwendet. Diese Methode ist eine in der Statistik generell häufig angewandte Schätzprozedur. Sie beruht auf einem einfachen Grundgedanken, der an Hand des folgenden bewußt simplen Zeitreihenmodelles illustriert werden soll. Angenommen, es gelte das Zwei-Komponentenmodell:

$$X_t = G_t + U_t, \quad t=1,2, \dots, T$$

und für die glatte Komponente werde eine lineare Verlaufsform angenommen, d.h. $G_t = a + bt$, $t=1,2, \dots, T$, wobei die unbekannt Parameter a und b aus der vorliegenden Zeitreihe zu bestimmen sind. Die irreguläre Komponente soll zunächst obigen - etwas vage formulierten - Annahmen genügen. Nach dem Kleinst-Quadrate-Prinzip werden nun die Parameter so bestimmt, daß die Summe der "Abweichungsquadrate"

$$\sum_{t=1}^T (X_t - G_t)^2 = \sum_{t=1}^T (X_t - \hat{a} - \hat{b}t)^2, \quad t = 1, 2, \dots, T$$

minimiert wird bezüglich \hat{a} und \hat{b} . \hat{a} und \hat{b} sind die Kleinst-Quadrate-"Schätzungen" von a und b . Sie lassen sich technisch durch partielle Differentiation bestimmen.

Hier ist nun eine Zwischenbemerkung zu machen. Bei dieser Kleinst-Quadrate-Prozedur ist von einem sog. "deskriptiven" Zeitreihenmodell ausge-

gangen worden, d.h. U_t wurde einfach als eine "Rest"-Komponente mit gewissen qualitativ umschriebenen Eigenschaften angesehen, die sie als "zufällig" qualifizieren soll. Präzisiert man jedoch "zufällig" im Sinne der Wahrscheinlichkeitstheorie, dann erhalten die U_t den Status von Zufallsvariablen (genauer: einer Folge von Zufallsvariablen, vgl. unten) mit genau definierten Eigenschaften, was man mit "stochastischer Spezifikation" bezeichnet. Im einfachsten Fall besteht diese in den drei Annahmen:

- aa) die Mittelwerte (genauer: Erwartungswerte) der U_t sind gleich Null für alle $t=1,2,\dots,T$
- bb) die Varianzen der U_t sind konstant für alle $t=1,2,\dots,T$
- cc) die U_t sind unkorreliert.

Mit dieser stochastischen Spezifikation sind \hat{a} und \hat{b} Schätzfunktionen im Sinne der statistischen Schätztheorie, d.h. Zufallsvariablen. Die konkreten numerischen Werte für \hat{a} und \hat{b} werden als Realisationen dieser Zufallsvariablen betrachtet. Erst auf dieser Ebene können dann Aussagen über etwaige Optimalitätseigenschaften von Schätzverfahren gemacht werden. So kann z.B. gezeigt werden, daß \hat{a} und \hat{b} bei der angenommenen stochastischen Spezifikation BLUE (Best Linear Unbiased Estimates) sind, d.h. die unbekannt Parameter a und b werden "unverzerrt" mit kleinster Varianz geschätzt, wobei die Schätzfunktionen linear von den Beobachtungswerten X_t abhängen.

Schließlich sei noch angemerkt, daß zwischen gleitenden Durchschnitten und Kleinst-Quadrate-Schätzungen manchmal bestimmte Zusammenhänge bestehen. Bei gewissen Schätzproblemen stellen sich beide Ansätze als formal äquivalent heraus, d.h. eine Kleinst-Quadrate-Lösung läßt sich als gleitender Durchschnitt darstellen.

Eine Skizzierung einiger praktischer Verfahren soll das Gesagte illustrieren. Beim "Berliner Verfahren" z.B., dem offiziellen Saisonbereinungsverfahren des Statistischen Bundesamtes, wird davon ausgegangen, daß sich die glatte Komponente als Polynom dritter Ordnung der Zeit darstellen läßt. Eine Trennung von Trend und Zyklus erfolgt also nicht. Daß kein geringerer Polynomgrad gewählt wird, hängt damit zusammen, daß sonst keine konjunkturellen Wendepunkte modelliert werden könnten. Die Wahl des Polynomgrades ist also mindestens teilweise substanzwissenschaftlich motiviert. Weiter wird angenommen, daß die saisonale Komponente als trigonometrisches Polynom der Zeit formulierbar ist mit den "saisonalen Frequenzen" $\lambda_k = 2\pi k/12$, $k=1,\dots,6$. Für die irreguläre Komponente wird ein schwach

stationärer Prozeß (zu diesen Prozessen siehe unten) angenommen mit Erwartungswert Null. Für eine Zeitreihe kann dann geschrieben werden:

$$X_t = \sum_{j=1}^3 a_j t^j + \sum_{k=1}^6 (b_k \cos \lambda_k t + c_k \sin \lambda_k t) + u_t$$

Charakteristisch für das Verfahren ist nun, daß die Gültigkeit dieses Modelles nicht für alle Zeitpunkte $t = 1, 2, \dots, T$ angenommen wird, sondern nur für die Länge gewisser Stützbereiche welche das Intervall $[1, T]$ gleitend überdecken (sog. "gleitende Regression"). Die unbekannt Parameter des Modelles werden nun mit Hilfe der Methode der Kleinsten Quadrate bestimmt. Es kann gezeigt werden, daß die Kleinst-Quadrate-Lösung als gleitender Durchschnitt dargestellt werden kann (wobei die "Gewichte" ein komplizierteres Verlaufsmuster aufweisen als beim einfachen 12-er Durchschnitt), d.h. wiederum als spezieller Filter. Die Methode der kleinsten Quadrate kann in diesem Zusammenhang deshalb auch als eine - wenn auch sehr "indirekte" - Methode des Filter-Designs interpretiert werden (vgl. dazu die Ausführungen zur Filtertheorie unten).

Ein Bereinigungsverfahren, das im wesentlichen aus einer Abfolge von gleitenden Durchschnitten besteht, ist das Census X-11 -Verfahren, das z.B. von der Deutschen Bundesbank offiziell verwendet wird. Neben dem schon bekannten 12-er Durchschnitt werden 5-er, 9-er, 23-er usw. Durchschnitte in verschiedenen Phasen des iterativen Bereinigungsprozesses verwendet.

Neuere Verfahren, wie z.B. von Schlicht und Pauly-Schlicht nehmen für die glatte und die saisonale Komponente keine deterministische Funktion der Zeit mehr an, sondern wählen auch für diese Komponenten einen stochastischen Ansatz. Grundlegend ist aber auch hier das traditionelle Komponentenmodell.

Es kann hier nicht darum gehen, einen Überblick über Saisonbereinigungsverfahren zu geben. Der daran interessierte Leser sei dafür auf Stier (1985) verwiesen. Die angeführten Beispiele zeigen allerdings die Grundproblematik des traditionellen Komponentenmodelles sehr deutlich:

a) Für die einzelnen Komponenten muß eine ganz bestimmte Verlaufsform postuliert werden, egal ob der Ansatz deterministisch oder stochastisch ist.

b) Da es eine Vielzahl von denkbaren Verlaufsformen gibt, stellt sich die Frage nach der jeweils "geeigneten". Diese Frage ist aber von den Daten her

grundsätzlich nicht zu entscheiden, denn offensichtlich "funktionieren" die erwähnten (und auch die nicht erwähnten) Bereinigungsverfahren auf "beliebigen" Zeitreihen. Ein wertender Vergleich der einzelnen Verfahren dahingehend, welches "am besten" bereinigt, ist sehr problematisch, da die zu verwendenden Vergleichskriterien selbst nicht unumstritten sind.

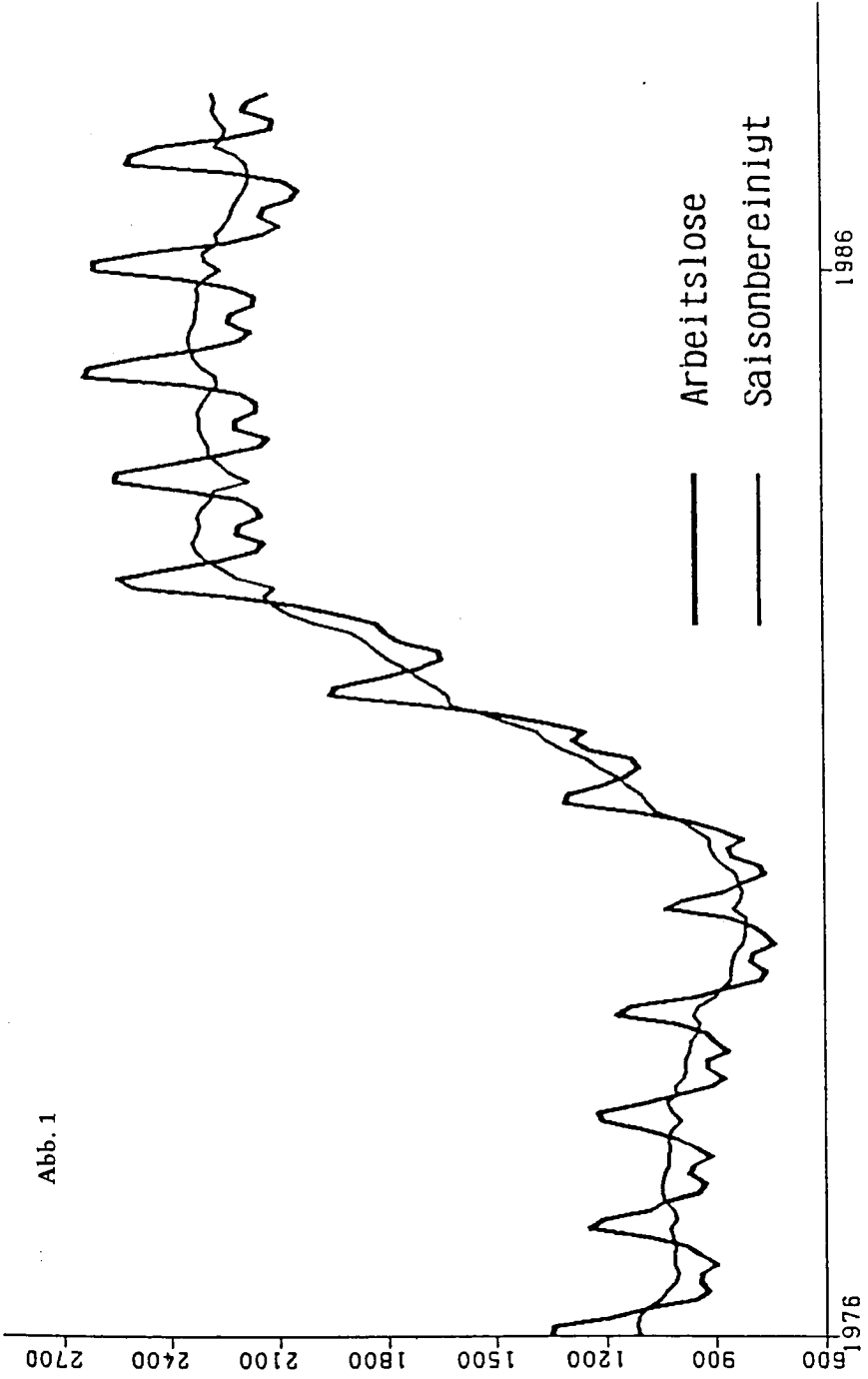
Schließlich seien hier die Resultate einer praktischen Saisonbereinigung graphisch wiedergegeben. Die monatliche Reihe "Arbeitslose in der BRD" (1976-1987) wurde mit Census X-11 (multiplikativ) bereinigt. Die beiden nachstehenden Abbildungen zeigen die saisonbereinigte Reihe (Abb. 1) und die glatte Komponente (Abb. 2), jeweils zusammen mit der Originalreihe.

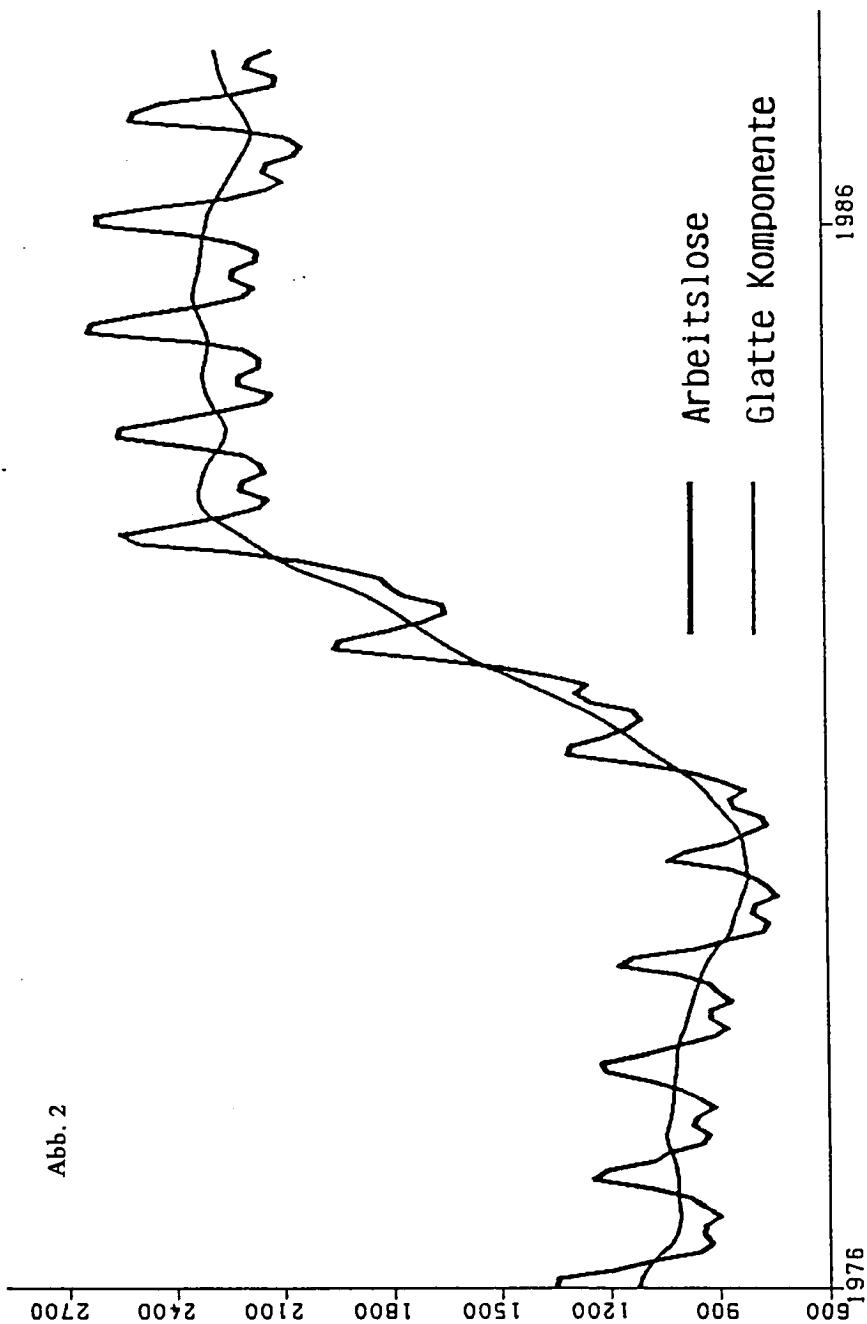
4. Weitere Bemerkungen zum traditionellen Komponentenmodell

Die Ausführungen zum traditionellen Komponentenmodell in der Zeitreihenanalyse, die oben im Kontext der Saisonbereinigung gemacht wurden, gelten auch - unter jeweils zu beachtenden Modifikationen, Einschränkungen bzw. Erweiterungen - für andere Problemstellungen. Insbesondere treten auch hier die eben erwähnten Probleme a) und b) auf. Diese kann man zusammenfassend dahingehend charakterisieren, daß bei diesem Ansatz mit nicht-testbaren Hypothesen gearbeitet wird, was methodologisch ein bemerkenswertes Faktum ist, das jedoch häufig übersehen wird. Um die damit verbundenen Probleme nochmals zu verdeutlichen, sei hier abschließend kurz auf einige Aspekte der sog. "Langen-Wellen"-Forschung eingegangen, die meistens auf der Basis des traditionellen Komponentenmodelles betrieben wird.

Das zu lösende Grundproblem kann kurz so umrissen werden: Für viele historische Zeitreihen wird die Hypothese der Existenz langfristiger zyklischer Bewegungsabläufe als sinnvoll angesehen. Darunter werden quasi-periodische Abläufe verstanden mit einer Dauer von z.B. etwa 50 Jahren, um den längsten (nach Kondratieff benannten) Zyklus beispielhaft herauszugreifen. Daneben werden von einzelnen Forschern kürzere Zyklen von etwa 20 Jahren usw. postuliert. Einzelheiten interessieren hier nicht. Nun ist es bei historischen Reihen praktisch nie möglich, die Existenz derartiger Zyklen unmittelbar nachzuweisen, z.B. durch eine optische Inspektion einer Reihe, da andere, nicht-zyklische Komponenten, wie vor allem der Trend sowie "noise", dominieren. Anders ausgedrückt: denkt man sich die Varianz einer Reihe auf einzelne Reihenkompenten aufgeteilt, so entfällt nur ein geringer Teil dieser Varianz auf eine solche möglicherweise vorhandene zyklische Komponente.

Abb. 1





Die Suche nach einer (oder mehreren) derartigen zyklischen Komponenten gleicht somit der Suche nach einer Nähnadel im Heuhaufen. Und sie kann auch nur dann mit Aussicht auf Erfolg unternommen werden, wenn es gelingt, die nicht-interessierenden Komponenten zu eliminieren. Dabei muß jedoch unbedingt sichergestellt werden, daß die verwendete Eliminationsprozedur weder vorhandene zyklische Komponenten tangiert (also z.B. ihre zyklischen Struktur verändert oder sie ganz oder teilweise eliminiert) noch Zyklen erzeugt. In der Regel besteht der erste Schritt darin, einen vorhandenen Trend zu eliminieren. Dabei stellt sich die Frage nach einer "geeigneten" Trendfunktion, wobei die oben diskutierten Probleme auftreten. Die Wahl einer speziellen Trendfunktion ist jedoch präjudizierend, denn der Verlauf der trendfreien Reihe hängt ganz entscheidend von dieser Wahl ab. Es leuchtet unmittelbar ein, daß die Suche nach zyklischen Komponenten in einer trendfreien Reihe ganz verschiedene Resultate erbringt, je nachdem, welcher Trend eliminiert wurde. Auf der Basis des klassischen Komponentenmodelles erscheint daher eine objektive Suche nach "langen Wellen" als ein aussichtsloses Unterfangen.

Wenn man von der Trendproblematik einmal absieht, verbleibt das weitere Problem der Elimination von kurzfristigen Schwankungen. Üblicherweise werden dazu gleitende Durchschnitte verwendet. Aber dabei stellt sich das Problem nach der Länge dieser Durchschnitte. Der Verlauf der geglätteten Reihe, die ja die zyklische Komponente in nuce darstellen soll, hängt nämlich davon ab. Außerdem ist bei Verwendung von gleitenden Durchschnitten, insbesondere wenn sie iterativ angewendet werden, darauf zu achten, daß durch sie keine Zyklen erzeugt werden. Es kann gezeigt werden, daß man durch Glätten von "noise" zyklische Reihen erzeugen kann (sog. Slutzy-Effekt). Weiter unten wird skizziert, wie man durch den Entwurf geeigneter Filter die erwähnten Probleme lösen bzw. umgehen kann.

II. Zeitreihenmodelle auf der Basis von stochastischen Prozessen

1. Grundlagen

Beim traditionellen Komponentenmodell wird unterstellt, daß sich eine Zeitreihe aus drei deterministischen Funktionen und einem Zufallsterm zusammensetzt. Der Zufall spielt in diesem Ansatz sozusagen die Rolle einer sachlich nicht weiter zu erklärenden Residualgröße. Bei neueren Modellkonzeptionen spielt der Zufall jedoch nicht mehr eine nur untergeordnete Nebenrolle. Vielmehr versucht man Zeitreihenmodelle zu bauen, die auf der

Theorie der stochastischen Prozesse beruhen. Unter einem stochastischen Prozeß versteht man - vereinfacht ausgedrückt - eine Folge von Zufallsvariablen, die im allgemeinen korreliert sind, also eine zeitliche Struktur aufweisen. Unter den stochastischen Prozessen sind die sog. "stationären" Prozesse mathematisch am weitesten entwickelt, insbesondere die sog. "schwach stationären" Prozesse. Diese sind einmal dadurch charakterisiert, daß sowohl ihre Mittelwert als auch ihre Varianzfunktion zeitinvariant sind. Zum anderen ist ihre Autokovarianz- bzw. Autokorrelationsfunktion nur vom "lag", d.h. der zeitlichen Distanz zweier Zufallsvariablen, abhängig und nicht von den jeweiligen absoluten Zeitpunkten.

Autokovarianz- und Autokorrelationsfunktion enthalten praktisch die gleichen Informationen über die zeitliche Struktur eines Prozesses. Während jedoch die Kovarianzfunktion vom verwendeten Maßstab abhängig und unbeschränkt ist, nimmt die Korrelationsfunktion nur Werte an, die absolut im Intervall $[0,1]$ liegen. Sie geht aus der Kovarianzfunktion durch eine geeignete Normierung hervor. Ein schwach stationärer Prozeß ist also durch die folgenden drei Postulate charakterisiert:

$$a) E(X_t) = m, t=1,2,\dots$$

$$b) \text{Var}(X_t) = \sigma^2, t=1,2,\dots$$

$$c) \text{Cov}(X_t, X_{t+s}) = E[(X_t - m)(X_{t+s} - m)] = C(s) \\ s=0,1,2,\dots, t=1,2,\dots$$

wobei s den zeitlichen Abstand je zweier Zufallsvariablen bezeichnet. Die Autokorrelationsfunktion $R(s)$ ist dabei durch

$$R(s) = C(s)/\text{Var}(X_t), s=0,1,2,\dots$$

mit $|R(s)| \leq 1, s = 0,1,2,\dots$ gegeben.

Offensichtlich ist die oben charakterisierte Zufallskomponente U_t im klassischen Komponentenmodell nichts anderes als ein schwach stationärer Prozeß mit einer speziellen Korrelationsfunktion: die paarweisen Korrelationen der $U_t, t=1,2,\dots$ sind gleich Null. Einen derartigen Prozeß nennt man auch "weißes Rauschen" ("white noise"). Dieser einfachste Prozeßtyp erweist sich für viele Analysesituationen als grundlegend.

Für kompliziertere Prozesse sind natürlich auch kompliziertere Autokorrelationsfunktionen zu erwarten.

Die Entwicklung eines Zeitreihenmodelles mit Hilfe stochastischer Prozesse impliziert die Hypothese, daß eine gegebene Reihe als eine Realisation eines derartigen (wenn auch zunächst noch unbekannt) Prozesses interpretiert werden kann. Anders ausgedrückt: man versucht, von einer vorliegenden Reihe auf den "erzeugenden" Prozeß zu schließen, d.h. den "zugrundeliegenden" Prozeß zu "identifizieren". Dabei verwendet man aus guten Gründen möglichst einfache Prozeßtypen, deren Korrelationsverhalten theoretisch bekannt ist und für deren unbekannt Parameter Schätzprozeduren verfügbar sind.

2. AR-, MA-, ARMA- und ARIMA-Prozesse

Grundlegende "Bausteine" zur Modellbildung sind die autoregressiven und die gleitenden Durchschnittsprozesse. Die Prozeßgleichung für einen autoregressiven Prozeß der Ordnung p (kurz AR(p)-Prozeß) lautet:

$$X_t - \Phi_1 X_{t-1} - \Phi_2 X_{t-2} - \dots - \Phi_p X_{t-p} = u_t, t=1,2,\dots$$

Dabei bezeichnet u_t weißes Rauschen. Damit AR(p) ein stationärer Prozeß ist, müssen die Parameter $\Phi_1, \Phi_2, \dots, \Phi_p$, bestimmten Bedingungen genügen, auf die hier allgemein nicht eingegangen werden kann. Für $p=1$ z.B. muß Φ_1 betragsmäßig kleiner als Eins sein. Man sieht leicht, daß für $|\Phi_1| > 1$, der Prozeß "explodieren" würde. Die "mildeste" Form von Nicht-Stationarität liegt hier vor für $\Phi_1 = 1$. Die beiden nachstehenden Graphiken zeigen je eine Prozeßrealisierung von 200 Werten eines AR(1)-Prozesses mit $\Phi_1 = 0.8$ und $\Phi_1 = 1$.

Interessant ist, daß sich das Niveau des Prozesses für $|\Phi_1| = 1$ dauernd ändert und tendenziell länger andauernde "trendartige" Auf- bzw. Abwärtsbewegungen zu beobachten sind. Dieser Prozeß wird auch als "random-walk" bezeichnet. Der Verlauf von Aktienkursen läßt sich z.B. häufig durch einen random-walk beschreiben.

Ein gleitender Durchschnittsprozeß der Ordnung q (auch "moving-average" - oder kurz MA(q)-Prozeß genannt) ist wie folgt definiert:

$$X_t = u_t - \Theta_1 u_{t-1} - \Theta_2 u_{t-2} - \dots - \Theta_q u_{t-q}$$

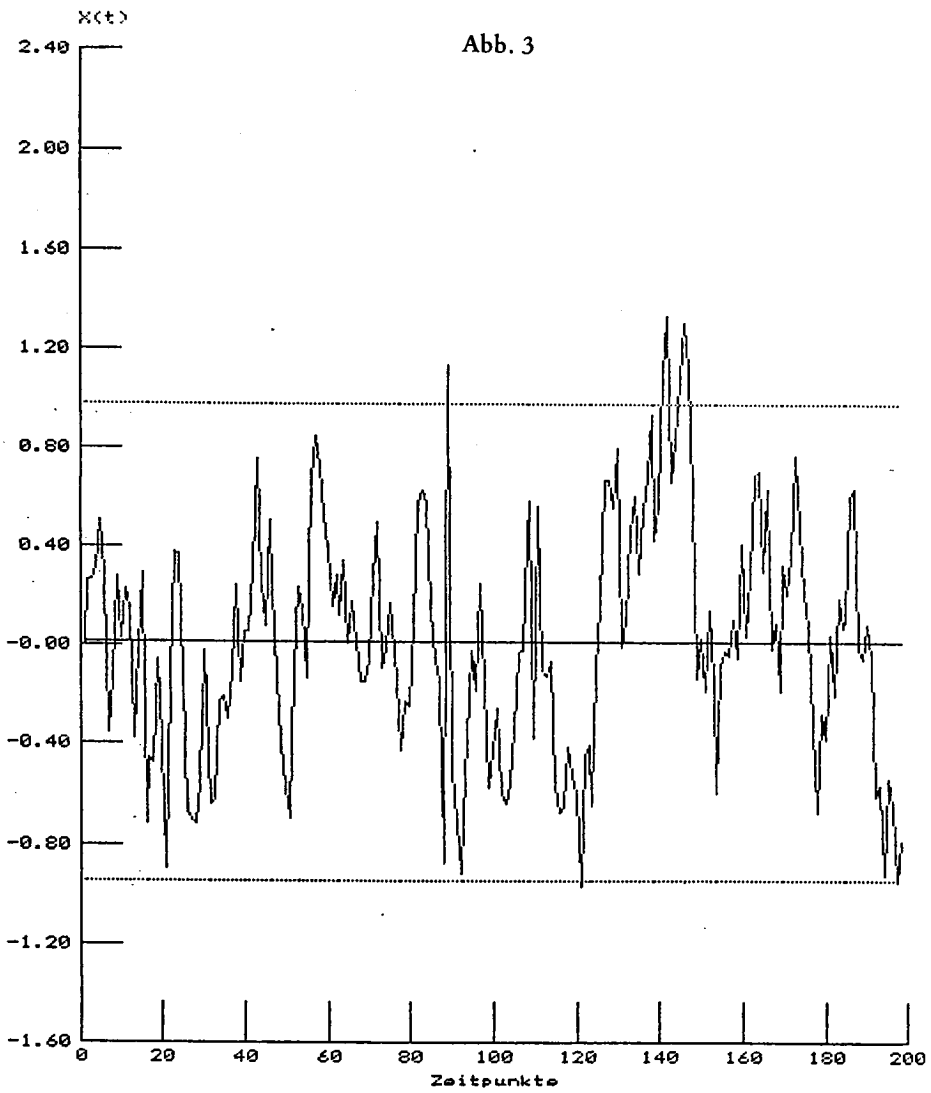
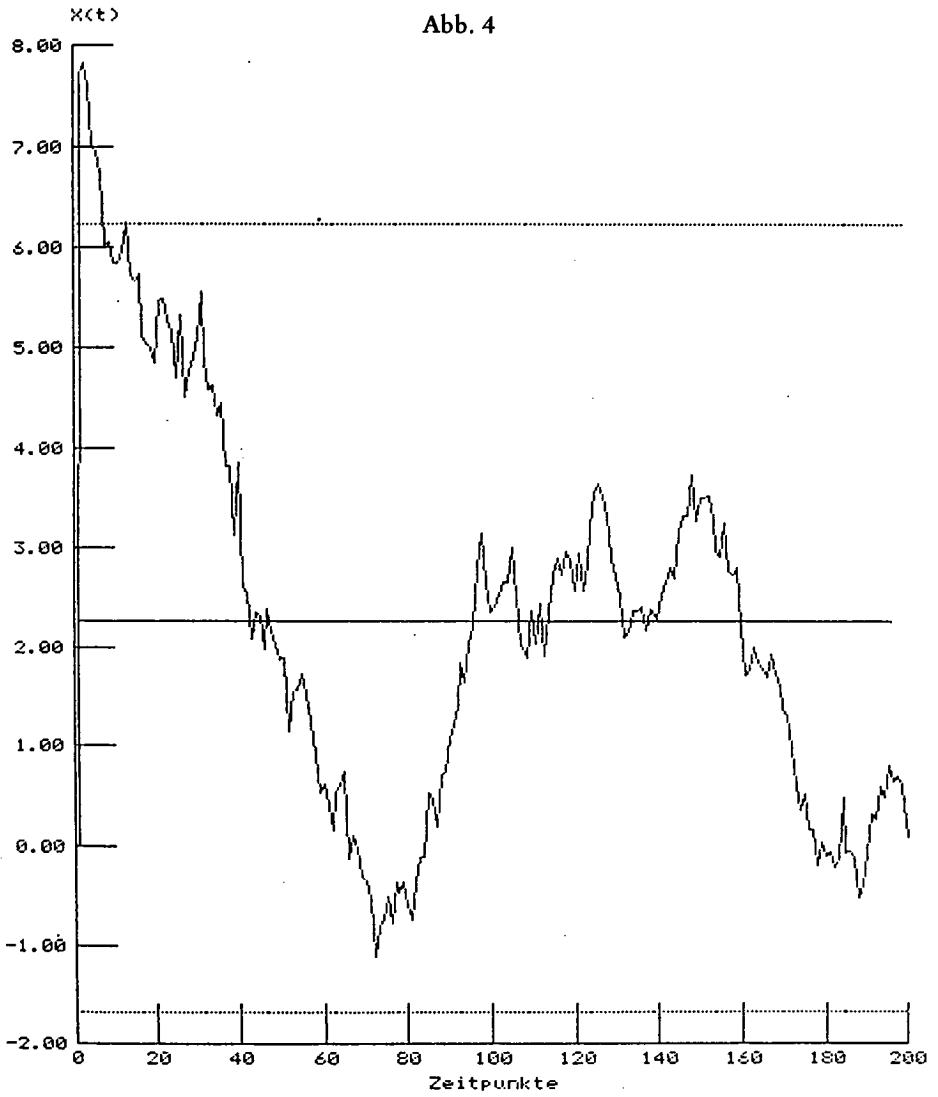
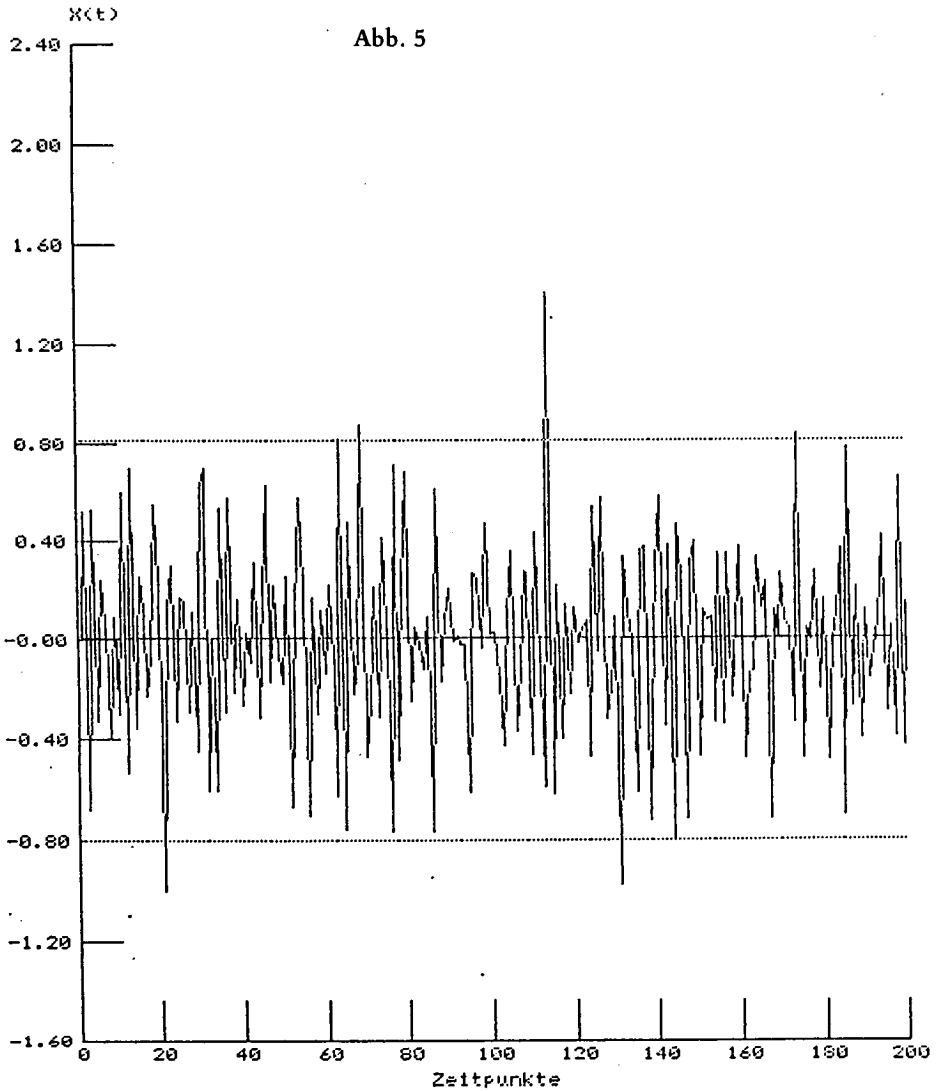


Abb. 4



wobei u_t wiederum weißes Rauschen bezeichnet. Nachstehende Abbildung zeigt eine Realisation eines MA(2)-Prozesses mit den Parametern $\Theta_1 = 0.9$ und $\Theta_2 = -0.1$.



Sowohl AR-als auch MA-Prozesse lassen sich mit Hilfe des "Verschiebungsoperators "

$$B^m(X_t) = X_{t-m}, \quad m=1,2,\dots$$

kürzer darstellen, wenn man damit die beiden " lag-Polynome"

$$\Phi(B) := 1 - \Phi_1 B - \Phi_2 B^2 - \dots - \Phi_p B^p$$

$$\Theta(B) := 1 - \Theta_1 B - \Theta_2 B^2 - \dots - \Theta_q B^q$$

einführt. Die beiden Prozesse lassen sich dann kurz schreiben als

$$\Phi(B)X_t = u_t$$

$$\text{bzw. } X_t = \Theta(B)u_t$$

Bei der Modellierung von Zeitreihen zeigt sich nun häufig, daß eine Verwendung eines "reinen" Prozeßtyps, also entweder eines AR- oder eines MA-Prozesses, zu hohen Prozeßordnungen (d. h. zu einem großen p bzw. q) führen kann. Da in der Praxis die Prozeßparameter unbekannt sind, müssen sie aus dem vorliegenden Datenmaterial (d. h. einer endlichen Reihe) geschätzt werden. Je mehr Parameter geschätzt werden müssen, umso ungenauer fällt ihre Schätzung aus, d. h. umso größer ist ihre Varianz. Es empfiehlt sich daher, Modelle mit möglichst wenigen Parametern zu verwenden (sog. Prinzip der "parsimony"). Dies kann häufig durch eine Kombination beider Prozeßtypen erreicht werden. Solche werden als ARMA-Prozesse bezeichnet. Die Prozeßgleichung eines ARMA (p, q) -Prozesses ist gegeben durch:

$$X_t - \Phi_1 X_{t-1} - \Phi_2 X_{t-2} - \dots - \Phi_p X_{t-p} = u_t - \Theta_1 u_{t-1} - \Theta_2 u_{t-2} - \dots - \Theta_q u_{t-q},$$

oder kurz durch

$$\Phi(B)X_t = \Theta(B)u_t.$$

Zur Modellierung praktischer Zeitreihen wie sie im Bereich der Ökonomie oder der historischen Forschung vorkommen, reicht allerdings auch die Klasse der ARMA-Prozesse nicht aus. Dies deshalb, weil solche Reihen in der Regel nicht mit dem Stationaritätspostulat vereinbar sind, das auch den ARMA-Prozessen zugrundeliegt. Der Hauptgrund dafür liegt darin, daß sie

meistens einen ausgeprägten Trend aufweisen und deshalb nicht um ein konstantes Niveau schwanken, wie das für stationäre Prozesse charakteristisch ist. Um dennoch mit ARMA-Prozessen arbeiten zu können, wurde von Box und Jenkins eine einfache Prozedur vorgeschlagen, die dazu dienen soll, die auf einen Trend zurückzuführende Nicht-Stationarität zu beseitigen. Diese besteht in einer Differenzenbildung:

$$Y_t := (1 - B)^d X_t, \quad d = 1, 2, \dots$$

d.h. die Reihe X_t soll durch Differenzenbildung in eine (stationäre) Reihe Y_t transformiert werden, die dann für die Modellbildung benutzt wird. Dabei gibt der Exponent "d" den "Grad" der Differenzenbildung an. Bei $d = 1$ ist $(1 - B)X_t = X_t - X_{t-1}$. Die hier gewählte vorsichtige Formulierung, daß Stationarität erreicht werden "soll", will darauf hinweisen, daß eine Differenzenbildung natürlich noch keine Garantie für ein stationäres Y_t ist. Allerdings existieren auch Tests zur Überprüfung dieser Hypothese, z.B. der Test von Dickey-Fuller, die aber praktisch eher selten verwendet werden. Hier verläßt man sich mehr auf die Erfahrung, die davon ausgeht, daß sich bei den meisten praktischen Reihen in der Regel durch höchstens zweimalige Differenzierung eine Reihe erzeugen läßt, die "hinreichend" stationär ist. Es sei hier noch kritisch angemerkt, daß die Technik der Differenzenbildung zur Trendelimination in anderen Zusammenhängen bedenklich sein kann, z.B. in der "Langen Wellen"-Forschung. Man kann nämlich zeigen, daß Differenzen eine ganz bestimmte Art von "Hochpaßfilter" sind mit einer sehr unscharfen "Amplitudenfunktion", mit der Konsequenz, daß gerade die interessierenden Bewegungskomponenten ausgefiltert werden. Näheres dazu siehe unten.

Prozesse, die sich nach d-maliger Differenzenbildung als ARMA-Prozesse darstellen lassen, werden als ARIMA(p,d,q)-Prozesse bezeichnet, die sich kurz so schreiben lassen:

$$\Phi(B) (1 - B)^d X_t = \Theta(B) u_t$$

p, d, und q bestimmen die Prozeßgleichung vollständig. Diese lautet z.B. für den ARIMA (1,1,2)-Prozeß:

$$(1 - \Phi_1 B) (1 - B) X_t = u_t - \Theta_1 u_{t-1} - \Theta_2 u_{t-2}$$

oder ausführlich:

$$X_t - (1 + \Phi_1)X_{t-1} + \Phi_1 X_{t-2} = u_t - \Theta_1 u_{t-1} - \Theta_2 u_{t-2}$$

Schließlich sei noch erwähnt, daß für Reihen, die eine saisonale Komponente enthalten, eine Erweiterung notwendig ist. Diese beruht im wesentlichen darauf, daß man sich eine Reihe in einen saisonalen Teil und in einen nicht-saisonalen Teil zerlegt denkt. Für beide Teile wird jeweils ein ARIMA-Modell entwickelt. Beide Modelle werden multiplikativ miteinander verknüpft zu einem Modell für die gesamte Reihe. Für den saisonalen Modellteil ist die Symbolik (P,D,Q) gebräuchlich. D bezeichnet dabei den Grad der saisonalen Differenzenbildung. Bei Monatsreihen z.B. werden für $D=1$ die Differenzen $X_t - X_{t-12}$ gebildet. P bzw. Q bezeichnen den Grad des saisonalen AR- bzw. MA-Polynoms. Das ganze Modell wird bezeichnet mit ARIMA $(p,d,q)*(P,D,Q)_{12}$ für Monatsreihen, mit ARIMA $(p,d,q)*(P,D,Q)_4$ für Quartalsreihen usw. Auf weitere Einzelheiten kann hier nicht näher eingegangen werden.

3. Identifikation, Schätzung und Diagnose von ARIMA-Modellen

Unter der Identifikation eines ARIMA-Modelles versteht man die Festlegung der Ordnungen p , q und d . Einzige Informationsquelle dafür ist die gegebene Zeitreihe, genauer: Kenngrößen, die auf der Basis dieser konkreten Reihe berechnet werden können. Hauptsächlich kommen dafür die Autokorrelationsfunktion und die sog. "partielle" Autokorrelationsfunktion in Frage.

Die Autokorrelationsfunktion wurde oben schon eingeführt. Es kann gezeigt werden, daß diese für einen reinen MA(q)-Prozeß ab dem lag $q+1$ identisch gleich Null ist. Für einen reinen AR(p)-Prozeß verläuft sie jedoch asymptotisch (monoton oder oszillierend) gegen Null. Somit ist diese Funktion prinzipiell geeignet, zwischen diesen beiden Prozeßtypen zu diskriminieren. In der Praxis muß sie allerdings erst geschätzt werden. Dafür existieren geeignete Schätzfunktionen, auf deren Wiedergabe hier verzichtet werden soll. Liegt nun z.B. ein MA(3)-Prozeß vor, dann müßten die geschätzten Autokorrelationen theoretisch ab dem lag 4 identisch gleich Null sein. Praktisch wird das jedoch nicht der Fall sein. Vielmehr werden die geschätzten Autokorrelationen auch ab dem lag 4 noch von Null verschieden sein. Allerdings ist zu erwarten, daß sie ab diesem lag nicht mehr "signifikant" von Null verschieden sind, d.h. für einen MA(3)-Prozeß ist genau mit 3 signifikant von Null abweichenden Autokorrelationen zu rechnen. Die Entscheidung über den vorliegenden Prozeßtyp wird deshalb mit Hilfe von speziellen Signifikanztests vollzogen. Als Prozeßordnung q wird generell diejenige lag-Zahl

identifiziert, deren zugehöriger Korrelationswert sich noch als signifikant verschieden von Null erweist.

Zur Identifikation eines reinen AR-Prozesses verwendet man die partielle Autokorrelationsfunktion. Diese läßt sich (vom Schätzaspekt her) am einfachsten durch folgende Überlegung verständlich machen. Angenommen, es liege ein reiner AR-Prozeß unbekannter Ordnung p vor. Eine naheliegende Strategie zur Bestimmung von p besteht darin, daß man eine Folge von Regressionen berechnet: zuerst X_t auf X_{t-1} , dann X_t auf X_{t-1} und X_{t-2} usw. Ist z.B. $p=4$, dann ist zu erwarten, daß der 5. und alle weiteren geschätzten Regressionskoeffizienten nicht mehr signifikant von Null verschieden sind. Die Entscheidung über Signifikanz bzw. Nichtsignifikanz dieser Regressionskoeffizienten kann mit dem in der Regressionstheorie üblichen t-Test getroffen werden. Die Folge dieser geschätzten Regressionskoeffizienten ist die geschätzte partielle Autokorrelationsfunktion. Als Ordnung p wird daher die Anzahl der von Null signifikanten Koeffizienten gewählt. Im Gegensatz zu AR-Prozessen bricht die partielle Autokorrelation für MA-Prozesse nicht abrupt ab, sondern geht asymptotisch gegen Null. Auto- und partielle Autokorrelation verhalten sich bezüglich der beiden reinen Prozeßtypen AR und MA sozusagen zueinander dual.

Wesentlich schwieriger gestaltet sich die Identifikation von ARMA-Prozessen, da sich für gemischte Prozesse weder für die Autokorrelation noch für die partielle Autokorrelation ein charakteristisches Verlaufsmuster angeben läßt. Beide Korrelationen gehen asymptotisch gegen Null. Deswegen wurden für diesen Fall zusätzliche Identifikationstechniken- bzw. Kriterien entwickelt wie z.B. die Vektorkorrelation, das AIC (Akaike)- und das BIC (Bayes)-Kriterium usw. Darauf kann hier nicht eingegangen werden.

Die Festlegung von d macht in der Regel weniger Schwierigkeiten, meistens genügt $d=1$ oder $d=2$. Generell sollte d so gewählt werden, daß die geschätzte Autokorrelationsfunktion "rasch" gegen Null abfällt (entweder monoton oder oszillierend).

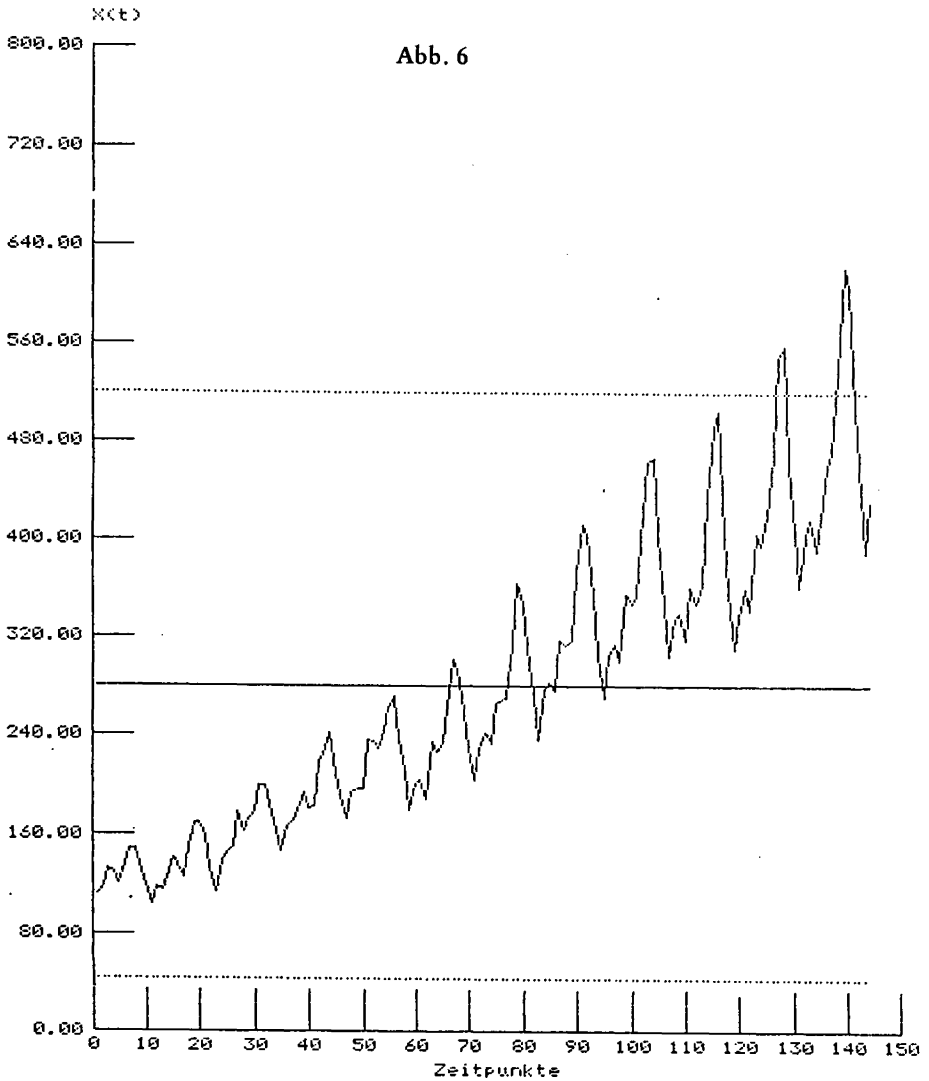
Die Identifikation eines ARIMA-Prozesses kann als ein Mustererkennungsproblem angesehen werden, das eigentlich viel Erfahrung und Fingerpitzengefühl voraussetzt. Allerdings existieren heute auch Software-Pakete (auch für PC) mit automatischen Identifikationsprozeduren wie z.B. AUTOBOX PLUS, so daß auch ein Nicht-Experte mühelos ARIMA-Modelle identifizieren kann.

Nach der Identifikation sind die unbekannt Parameter des ARMA-Prozesses zu schätzen. Dabei geht es im Prinzip darum, diese so zu bestimmen,

daß sich ein möglichst guter "fit" von Modell (genauer: den vom Modell "vorhergesagten" Reihenwerten) und den tatsächlichen Reihenwerten ergibt. Dazu wird entweder die Methode der kleinsten Quadrate verwendet oder die sog. "Maximum-Likelihood"-Methode. Die Likelihood-Funktion kann als Plausibilitätsmaß dafür interpretiert werden, daß der zugrundeliegende Prozeß gerade die vorliegende Zeitreihe erzeugt hat. Es liegt deshalb nahe, die unbekannt Parameter so zu wählen, daß diese Plausibilität maximal wird. Auf Einzelheiten muß wieder verzichtet werden. Es sei nur noch erwähnt, daß sich ein Anwender um diese ziemlich komplizierten Schätzprozeduren nicht zu kümmern braucht, da in allen einschlägigen Software-Paketen der Schätzvorgang vollständig automatisiert ist.

Als letzter Schritt der Modellentwicklung ist schließlich noch die Modell-diagnose zu nennen. Dabei geht es im wesentlichen um Residuenanalyse. Nach Modellkonzeption sollten die Residuen (die als Differenzen zwischen Modell-Reihenwerten und den tatsächlichen Reihenwerten definiert sind) als Realisation eines white-noise-Prozesses interpretierbar sein, was gleichbedeutend damit ist, daß ihre Autokorrelationen nicht signifikant von Null verschieden sein dürfen. Dies kann mit Hilfe spezieller Tests (z.B. dem Ljung-Box-Test) überprüft werden. Können die Residuen nicht in dieser Weise interpretiert werden, so ist das ein Indiz für eine fehlerhafte Modellspezifikation, was Anlaß für eine erneute Identifikation geben kann. Bei der automatischen Identifikation werden die einzelnen Schritte: Identifikation-Schätzung-Diagnose iterativ so oft durchlaufen, bis sich eine "befriedigende" Diagnose ergibt.

Abschließend sei hier noch eine grundsätzliche Bemerkung angefügt. Der oben skizzierte Ansatz unterstellt, daß sich eine gegebene Zeitreihe als eine Realisation eines ARIMA-Prozesses darstellen läßt. Praktische Reihen sind jedoch häufig "verzerrt" durch "Ausreißer" und/oder "Strukturbrüche" (wie z.B. plötzliche Niveauverschiebungen). Letztere können vorübergehend ("transient") sein oder auch (ab einem bestimmten Zeitpunkt) "permanent". Solche Effekte kann man insbesondere bei historischen Zeitreihen beobachten. Bei der ARIMA-Modellierung ist zu beachten, daß diese Effekte einen starken Einfluß auf die Schätzung der Korrelationsfunktionen haben können, mit der Konsequenz, daß die resultierende Identifikation möglicherweise überwiegend ein Resultat von Reihen-"Irregularitäten" ist. Da diese sich aber nicht unbedingt in der Zukunft wiederholen, kann die Prognosequalität des identifizierten Modelles davon negativ tangiert sein. Allerdings sind diese Effekte einer statistischen Definition und Analyse zugänglich. Mit Hilfe der sog. "Interventionsanalyse" können sie bei der Modellbildung berücksichtigt

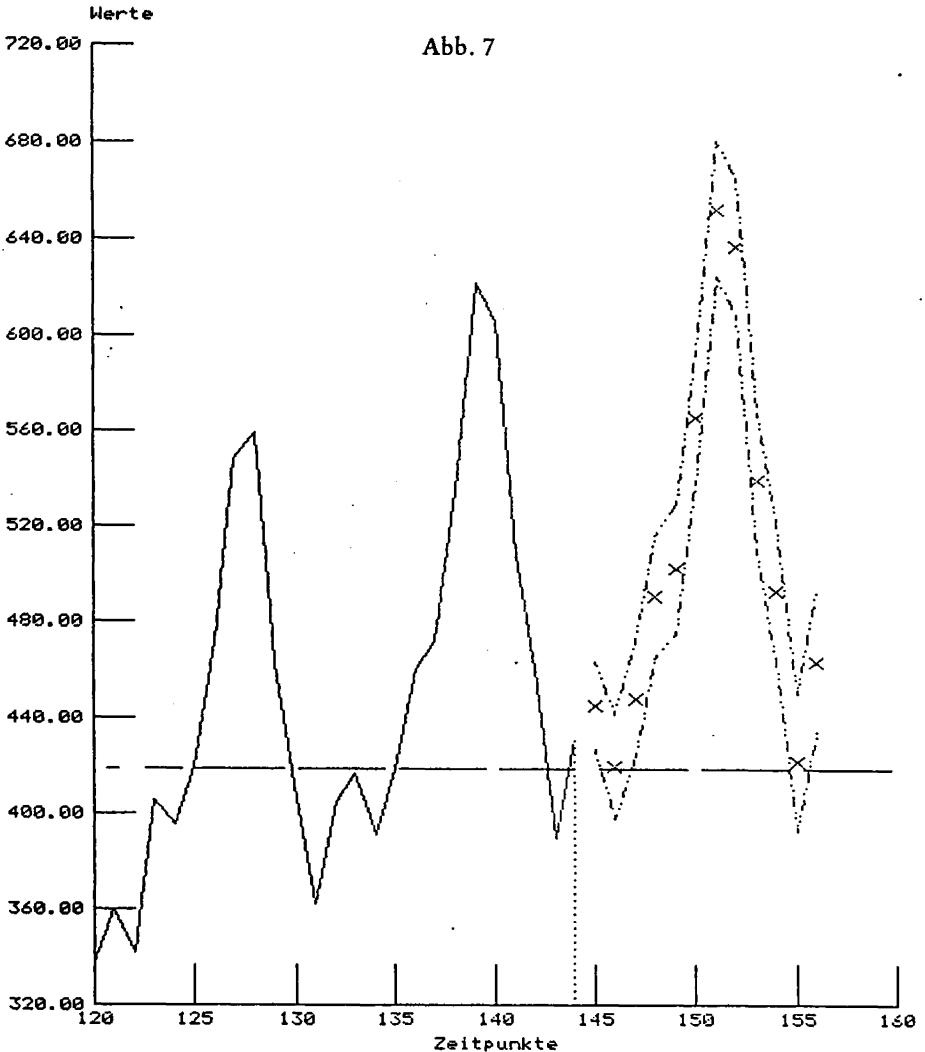


werden. Die angedeuteten Probleme können damit vermieden, zumindest aber in ihren Konsequenzen stark vermindert werden. Im AUTOBOX PLUS z.B. wird eine Interventionsanalyse routinemäßig durchgeführt.

4. Prognose mit Hilfe von ARIMA-Modellen

ARIMA-Zeitreihenmodelle werden vor allem erstellt, um Prognosen durchzuführen. Sie werden jedoch auch bei der Identifikation dynamischer

Beziehungen zwischen zwei oder mehr Zeitreihen benötigt. Aus der identifizierten und geschätzten Prozeßgleichung läßt sich relativ einfach eine Prognosegleichung ableiten. Eine eventuell vorgenommene Differenzbildung muß für die Prognose durch Summenbildung wieder rückgängig gemacht werden. Einzelheiten können wiederum nicht besprochen werden. Stattdessen soll abschließend ein Beispiel für die konkrete Modellierung und Prognose einer Reihe gegeben werden. Verwendet wird dazu die Reihe AIRLINE (Anzahl der Passagiere, monatlich, von Januar 1949 bis Dezember 1960, vgl. Box/Jenkins S.531), die nachstehend wiedergegeben ist:



Die automatische Identifikation mit AUTOBOX PLUS (mit Interventionsanalyse) ergibt nach Parameterschätzung folgendes Modell:

$$(1 - .55268B - .25479B^2)(1 - B^{12})X_t + .074999 = u_t,$$

d.h. einen ARIMA $(2,0,0)*(0,1,0)_{12}$ - Prozeß. Die Prognosewerte für 12 Monate (also bis Dezember 1961) sind in der nächsten Graphik zu finden, wobei die Reihe ab Dezember 1958 wiedergegeben ist:

Die gestrichelten Linien bedeuten die untere bzw. obere 95%- Konfidenzgrenze für die prognostizierten Werte. Diese Intervallprognosen beruhen auf der Annahme, daß der white-noise Prozeß u_t normal verteilt ist.

5. Multivariate ARIMA(Transferfunktionen)-Modelle

Bisher wurden ausschließlich univariate Zeitreihenmodelle betrachtet. Ein nicht unwesentlicher Fortschritt gegenüber dem traditionellen Komponentenmodell besteht darin, daß bei einer Modellierung auf der Grundlage stochastischer Prozesse multivariate Zusammenhänge erfaßt werden können. Eine wichtige Klasse von Modellen stellen die sog. "Transferfunktionenmodelle" dar. Dabei wird eine Reihe als abhängig von mehreren anderen betrachtet. Damit lassen sich z.B. substanzwissenschaftlich formulierte Abhängigkeiten empirisch testen. Existieren solche Abhängigkeiten, dann ist zu erwarten, daß man eine Reihe (nämlich die "abhängige") "besser" prognostizieren kann, wenn man "erklärende" Variablen hinzuzieht, als wenn diese Reihe nur "aus sich selbst" (also univariat) prognostiziert wird. Außerdem ist es möglich, dabei "lead-lag"- Beziehungen aufzuspüren. Beides kann unter bestimmten Bedingungen als Indikator für Kausalbeziehungen interpretiert werden. Diese Möglichkeiten erscheinen auch insbesondere für historische Fragestellungen interessant zu sein. Bisher sind allerdings keine einschlägigen Anwendungen bekannt geworden, was auch damit zusammenhängen mag, daß die Theorie der Transferfunktionenmodelle wesentlich komplizierter ist als diejenige der univariaten ARIMA-Modelle. Aus diesem Grund können hier auch nur einige grundlegende Zusammenhänge dargestellt werden. Dabei soll vom einfachsten Fall mit nur einer erklärenden Variablen ausgegangen werden. Die abhängige Variable sei mit Y_t und die unabhängige mit X_t bezeichnet. Das Transferfunktionsmodell hat dann folgende Gestalt:

$$Y_t = \frac{\omega(B)}{\delta(B)} X_{t-b} + N_t,$$

wobei der noise-Prozeß nicht unbedingt white-noise, sondern im allgemeinen selbst ein ARIMA(p,d,q)-Prozeß ist:

$$\Phi(B)N_t = \Theta(B)u_t$$

$\omega(B)$ und $\delta(B)$ sind Polynome in B und der Quotient $\omega(B)/\delta(B)$ bezeichnet die "Transferfunktion" zwischen der Input-Variablen X_t und der Output-Variablen Y_t . b ist die sog."reine" Verzögerung zwischen In- und Output. Diese wird nicht vorgegeben, sondern aus den Daten geschätzt.

Das gesamte Modell lautet somit:

$$Y_t = \frac{\omega(B)}{\delta(B)} X_{t-b} + \frac{\theta(B)}{\phi(B)} u_t \quad .$$

Dieses Modell kann erweitert werden durch Einbeziehung weiterer erklärender Variablen. Offensichtlich besteht eine Analogie zwischen Transferfunktions- und Regressionsmodellen. Auch hier verläuft die Modellbildung in einem dreistufigen iterativen Prozeß von Identifikation, Schätzung und Diagnose, wobei die einzelnen Schritte natürlich komplizierter als bei den univariaten Modellen sind. Das sollte aber interessierte Anwender nicht abschrecken, z.B.ist mit AUTOBOX PLUS eine automatische Transferfunktionen-Modellierung möglich.

III. Grundzüge der Filtertheorie

1. Vorbemerkungen

Filtermethoden gehören seit jeher zu den wichtigsten Werkzeugen der Zeitreihenanalyse. Daß sie auch für Fragestellungen in den Geschichtswissenschaften von Bedeutung sein können, kann man am Beispiel der "langen Wellen"-Forschung sehen. Dabei wird in starkem Maß von Filtern Gebrauch gemacht, meistens in Form von gleitenden Durchschnitten. Wie oben ausgeführt wurde, können schon einfache gleitende Durchschnitte als Filter interpretiert werden. Die Begründung dafür ergibt sich aus der Filtertheorie, deren Grundzüge nachstehend skizziert werde. Wiederum wird auf Ableitungen und Details verzichtet.

2. FIR-und IIR-Filter

Unter einem "Filter" kann allgemein ein Algorithmus zur Transformation einer Zeitreihe verstanden werden. Im Rahmen filtertheoretischer Überle-

gungen spricht man statt von Zeitreihen auch häufig von "Signalen". Formal kann der Zusammenhang zwischen der zu filternden Reihe X_t (dem "Filter-Input") und der gefilterten Reihe (dem "Filter-Output") Y_t folgendermaßen dargestellt werden:

$$Y_t = F(X_t).$$

Für theoretische Zwecke ist es oft zweckmäßig, wenn angenommen wird, daß X_t ein Signal unendlicher Länge ist. Die bei der Filterung praktischer Reihen auftretenden Probleme sind allerdings häufig auf deren Endlichkeit zurückzuführen.

Als theoretischer Hintergrund für die hier darzustellenden Filter eignet sich die Theorie der diskreten, linearen, zeitinvarianten Systeme. Dabei wird unter einem "System" jede Vorschrift verstanden, die ein Input-Signal in ein Output-Signal überführt. Linear und zeitinvariant ist ein System, wenn diese Vorschrift durch eine lineare Gleichung mit konstanten Koeffizienten darstellbar ist. So ist z.B. ein einfacher gleitender Durchschnitt mit $2m+1$ Gliedern

$$Y_t = \frac{1}{2m+1} \sum_{t=-m}^m X_t$$

ein zeitinvariantes System. Allgemein kann man für ein lineares System schreiben:

$$Y_t = \sum_{k=-\infty}^{\infty} h_k X_{t-k}, \quad t = \dots -2, -1, 0, 1, 2, \dots$$

wobei die Folge h_k die sog. "Impuls-Antwortfunktion" des Systems (Filters) ist. Diese ist der Output eines Filters, wenn als Input die sog. "diskrete Impulsfunktion" verwendet wird. Darunter ist ein Signal zu verstehen, das für $t=0$ gleich Eins und sonst gleich Null ist.

Mit Hilfe der Impuls-Antwortfunktion (IAF) können Filter grundsätzlich in zwei Klassen eingeteilt werden: solche mit endlicher IAF und solche mit unendlicher IAF. Die ersteren werden auch als FIR-Filter (Finite-Impulse-Response-Filter) und die letzteren als IIR-Filter (Infinite-Impulse-Response-Filter) bezeichnet. Offensichtlich sind gleitende Durchschnitte spezielle FIR-Filter.

Die Hauptaufgabe der Filtertheorie besteht darin, Methoden zur Konstruktion von Filtern zu entwickeln, die gewissen a priori spezifizierten Kriterien entsprechen. Dabei spielt die sog. "Transferfunktion" eines Filters die

entscheidende Rolle. Sie gibt, grob gesagt an, wie der Filter unterschiedliche Schwingungskomponenten einer Zeitreihe jeweils gewichtet (dazu unten mehr). Jede Zeitreihe läßt sich nämlich, auch wenn sie nicht "zyklisch" aussieht, als Kompositum einer Vielzahl von Schwingungen ("harmonische" Wellen) unterschiedlicher "Frequenzen" auffassen. Dabei ist die Frequenz der Reziprokwert der Schwingungsdauer ("Periode"), die wiederum die Anzahl der Zeiteinheiten darstellt, die eine komplette Schwingung benötigt. Die kleinste beobachtbare Schwingungsdauer beträgt (bei diskreten Reihen) zwei Zeiteinheiten. Die höchste Frequenz ist also $f = 1/2$. Ein linearer Trend läßt sich als eine Welle mit "unendlich langer" Periodendauer konzipieren, die entsprechende Frequenz ist also Null.

Für IIR-Filter läßt sich die Beziehung zwischen Filter-Input und Filter-Output in Form einer linearen Differenzgleichung

$$y_t + a_1 y_{t-1} + \dots + a_M y_{t-M} = b_0 x_t + b_1 x_{t-1} + \dots + b_N x_{t-N}$$

darstellen. Der Filter ist vollständig bestimmt, wenn die Koeffizienten a_1, \dots, a_M und b_1, \dots, b_M bestimmt sind (für FIR-Filter gilt $a_1 = \dots = a_M = 0$). M und N bezeichnen die Ordnung eines Filters. Seine Transferfunktion ist gegeben durch:

$$T(\lambda) = \frac{\sum_{r=0}^N b_r e^{-i\lambda r}}{\sum_{s=0}^M a_s e^{-i\lambda s}}, \quad |\lambda| \leq \pi$$

Offensichtlich hat diese Funktion für FIR-Filter nur Nullstellen, während sie für IIR-Filter sowohl Nullstellen als auch Pole besitzt. (Unter den Polen einer rationalen Funktion sind die Nullstellen des Nenners zu verstehen. An diesen Stellen nimmt die Funktion einen unendlich großen Wert an. Zum Beispiel hat die Funktion $f(z) = (z + 2)/(z-1)$ eine Nullstelle bei $z = -2$ und einen Pol bei $z = 1$.) Die Transferfunktion ist im allgemeinen komplexwertig und kann deshalb geschrieben werden als:

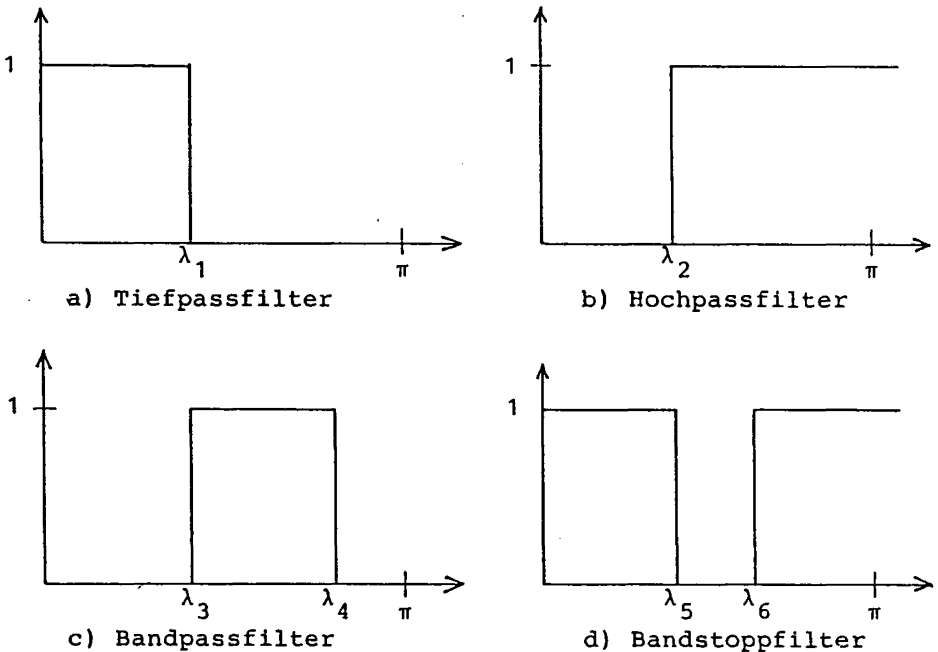
$$T(\lambda) = |T(\lambda)| e^{i\Phi(\lambda)}.$$

Dabei ist $|T(\lambda)| = A(\lambda)$ die sog. "Amplitudenfunktion" und $\Phi(\lambda)$ die sog. "Phasenfunktion" des Filters. Aus der Amplitudenfunktion lassen sich die "Ausblendeigenschaften" eines Filters ermitteln, d.h. die Frage beantwor-

ten, wie die "power" eines Signals, die einzelnen Frequenzen (bzw. Frequenzintervallen) zuzuordnen ist, vom Filter verändert (d.h. abgeschwächt oder eliminiert oder verstärkt) wird. Ein wesentlicher Aspekt beim Filter-Design besteht gerade darin, Filter zu entwerfen, die eine vorgegebene Amplitudenfunktion "realisieren", die nach substanzwissenschaftlichen Gesichtspunkten spezifiziert wurde. Dabei wird in der Regel das Ziel verfolgt, Informationen, die in einem Signal enthalten sind, zu extrahieren, so daß sie quasi "isoliert" auswertbar sind (vgl. dazu das oben skizzierte Problem der Extraktion von "langen Wellen"), oder sie auch zu eliminieren.

Grundlegend und praktisch am wichtigsten sind die vier Filtertypen: Tiefpaß-, Hochpaß-, Bandpaß- und Bandstoppfilter. Deren (ideale) Amplitudenfunktionen folgen nachstehend (dabei wird nur die positive Frequenzachse wiedergegeben, da sie symmetrisch zur Ordinate sind):

Abb. 8



Mit einem Hochpaß z.B. läßt sich eine Trendelimination durchführen, d.h. es werden alle Schwingungskomponenten mit einer Frequenz $\lambda < \lambda_2$ aus der Reihe eliminiert. Dabei ist die Frequenz λ_2 so festzulegen, daß sie mit der jeweiligen substanzwissenschaftlich fundierten Definition von "Trend" korrespondiert. Das gleiche gilt für die Frequenz λ_1 beim Tiefpaß, der einen Trend

extrahiert oder für die Frequenzen λ_3 und λ_4 beim Bandpaß, die z.B. eine "lange Welle" definieren, während die Frequenzen λ_5 und λ_6 beim Bandstopp eine zu eliminierende Komponente definieren. Man sieht, daß hier die Möglichkeit besteht, die einzelnen Filter so zu entwerfen, daß sie einzelne Komponenten einer Reihe "problembezogen" erfassen. Eine derartige Möglichkeit besteht im Rahmen des traditionellen Komponentenmodelles und mit den Werkzeugen der traditionellen Zeitreihenanalyse nicht (vgl. dazu die Ausführungen unter 1.3).

Da obige Amplitudenverläufe als "ideal" bezeichnet wurden, steht zu vermuten, daß reale Filter die dargestellten "exakten" Amplituden nicht unbedingt aufweisen. Unter "exakt" sei hier verstanden, daß die Amplituden in den Paß- bzw. Stoppbändern genau den Wert Eins bzw. Null annehmen. Eine Zielsetzung bei der Filterkonstruktion besteht deshalb in einer möglichst guten Approximation der idealen Amplitudenfunktionen.

Neben der Amplitudenfunktion ist die Phasenfunktion eine wichtige Kenngröße eines Filters. Sie liefert Informationen darüber, mit welcher Verzögerung der Filter-Output dem Filter-Input folgt. Es kann gezeigt werden, daß FIR-Filter eine lineare, (und damit als Spezialfall möglicherweise auch eine Nullphase), IIR-Filter jedoch eine nicht-lineare Phase besitzen.

Es existiert eine Vielfalt von Filter-Designmethoden sowohl für FIR-als auch für IIR-Filter, die hier jedoch nicht einmal ansatzweise besprochen werden können. Der interessierte Leser sei dazu auf Stier (1978, 1990) verwiesen. Vergleichend sei hier lediglich auf einige wichtige charakteristische Unterschiede bzw. Besonderheiten hingewiesen:

a) Mit IIR-Filter lassen sich schon bei niedriger Filterordnung sehr "scharfe" (praktisch "exakte") Amplitudenfunktionen realisieren. Dies ist mit FIR-Filter nicht möglich. Scharfe Amplituden erfordern bei diesem Filtertyp sehr hohe Ordnungen, zu deren Realisierung meistens mehrere hundert Datenpunkte erforderlich sind.

b) Exakte Nullphasen sind nur mit FIR-Filter möglich. Dabei ist aber eine "symmetrische" Implementierung Voraussetzung. "Asymmetrische" Implementierungen erzeugen eine lineare Phase (d.h. jeder Output-Wert ist um den gleichen Betrag gegenüber seinem Inputwert verschoben). Allerdings kann mit IIR-Filter manchmal eine "quasi"- Nullphase erzielt werden.

c) Zwischen der Schärfe der Amplitudenfunktion und der Größe der Phasenverschiebung besteht grundsätzlich eine Antinomie: je schärfer die Amplitude, desto größer ist die Phase.

d) IIR-Filter erzeugen absolut randstabile Filter-Outputs.

3. Filtern im Frequenzbereich

Beim Entwurf von FIR- und IIR-Filter wird sowohl im Frequenz- als auch im Zeitbereich operiert. In der Regel wird die Transferfunktion im Frequenzbereich konstruiert. Daraus werden die Filterkoeffizienten bestimmt und die Filterung im Zeitbereich durchgeführt (via Differenzgleichung). Idealerweise sollte ein Filter die folgenden drei Eigenschaften aufweisen:

- a) exakte Amplitudenfunktion
- b) exakte Nullphase
- c) absolute Randstabilität.

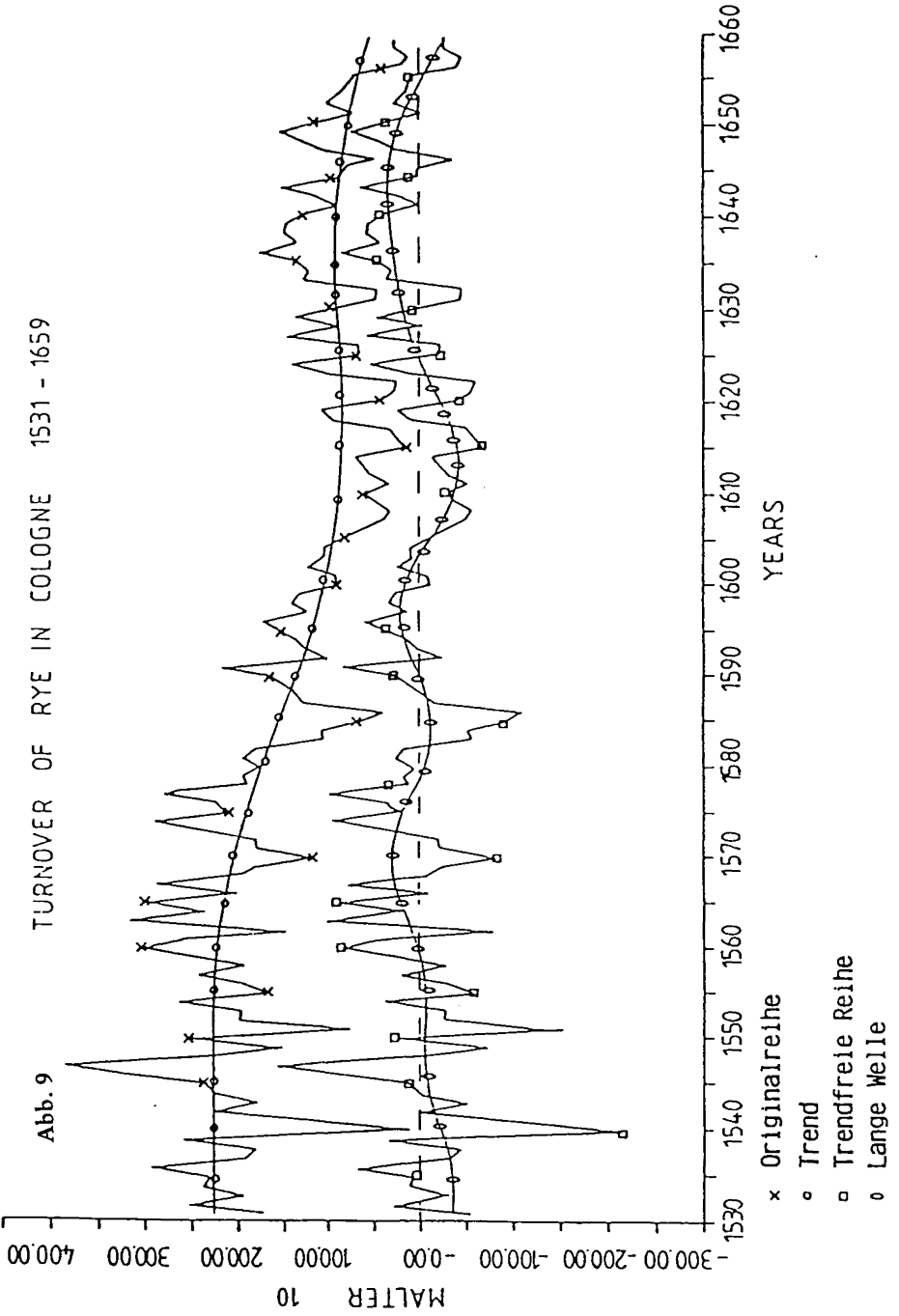
Man kann sich leicht überlegen, daß weder mit FIR- noch mit IIR-Filter alle drei Desiderata simultan erfüllbar sind. Für manche Analysen erweisen sich aber die Eigenschaften a) und b) als fast unverzichtbar. Dies trifft z.B. auf Probleme zu, wie sie bei der "langen Wellen"-Forschung auftreten. Das "Aufspüren" und gegebenenfalls Extrahieren solcher "Wellen" erfordert Filter, die Postulat a) erfüllen. Da sie als historische Abläufe interpretierbar sein müssen, kommt "Datierungen" ein besonderes Gewicht zu. Deshalb ist auch b) in diesem Kontext unverzichtbar.

Neben den oben erwähnten Filter-Design-Möglichkeiten gibt es jedoch eine dritte, bei der sowohl a) als auch b) erfüllbar sind. Diese ist dadurch charakterisiert, daß auch der Filtervorgang selber im Frequenzbereich stattfindet. Dazu ist es notwendig, die Daten in diesen Bereich zu transformieren, was mit der "diskreten Fourier-Transformation" möglich ist. Die Amplitudenfunktion läßt sich direkt im Frequenzbereich exakt vorgeben. Allerdings ist mit dieser Filtermethode kein absolut randstabiler Output zu erzielen. Auch gibt es dabei ein "Randproblem", das hier jedoch nicht besprochen werden kann. Stattdessen sei abschließend ein Beispiel betrachtet.

Dazu wird die Reihe "Umsatz von Roggen in Köln" (1531-1659) verwendet. Als "Trend" werden alle Bewegungen betrachtet, deren Periode länger als 60 Jahre ist. Zu seiner Extraktion wurde deshalb ein Tiefpaß mit $\lambda_1 = 2\pi/60$ benutzt. Zur Trendelimination diente ein dazu komplementärer Hochpaß. Zur Extraktion evtl. vorhandener "langer Wellen" wurde ein Bandpaß mit $\lambda_3 = 2\pi/60$ und $\lambda_4 = 2\pi/20$ verwendet, da alle Oszillationen mit einer Periodenlänge zwischen 20 und 60 Jahren als "lange Wellen" definiert wurden. Die erzielten Resultate sind in den folgenden Abbildungen wiedergegeben.

Es ist zu betonen, daß beim Filtern keinerlei a priori-Verlaufsformen spezifiziert werden müssen wie beim traditionellen Komponentenmodell. Es ist lediglich erforderlich, daß die interessierenden Komponenten im Frequenz-

Abb. 9 TURNOVER OF RYE IN COLOGNE 1531 - 1659



bereich nach substanzwissenschaftlichen Gesichtspunkten definiert werden. Wie sie im Zeitbereich verlaufen ist dabei unerheblich. Dieser Verlauf ist sozusagen erst a posteriori, d.h. als Resultat der Filterung, feststellbar.

4. Filter-Design und Spektralanalyse.

Immer wieder werden in Diskussionen Filter-Design und Spektralanalyse miteinander in Verbindung gebracht. Nicht selten wird beides - fälschlicherweise - gleichgesetzt. Da diese Vorstellung anscheinend nur schwer auszumerzen ist, sei hier speziell auf einen Vergleich dieser beiden Gebiete eingegangen.

Gemeinsam ist beiden, daß sie es mit dem Frequenzbereich zu tun haben. Das ist aber auch schon alles. Die Fragestellung der Spektralanalyse läßt sich kurz so charakterisieren:

Ausgangspunkt ist die Theorie der schwach stationären Prozesse. Im Rahmen dieser Theorie wird gezeigt, daß für jeden schwach stationären Prozeß unter bestimmten Bedingungen ein sog. "Spektrum" existiert. Falls es "normiert" wird, ist auch die Bezeichnung "Spektraldichte" gebräuchlich. Die Normierung hat zur Folge, daß die Fläche, welche diese (positive) Funktion mit der Frequenzachse einschließt, gleich Eins ist. Die Spektraldichte hat formal die gleichen Eigenschaften wie die Dichtefunktion einer Zufallsvariablen. Inhaltlich kann sie so interpretiert werden, daß sie den Frequenzgehalt eines Prozesses repräsentiert, genauer: seine Verteilung auf einzelne Frequenzbänder. Da das Integral über das Spektrum gleich der Varianz des stationären Prozesses ist, kann auch davon gesprochen werden, daß das Spektrum (bzw. die Spektraldichte) zeigt, welche Frequenzbänder den zeitlichen Verlauf eines stationären Prozesses hauptsächlich bestimmen. Dominieren z.B. die hohen Frequenzen, dann ist dieser Verlauf ziemlich unruhig, während bei Dominanz von niedrigen Frequenzen ein eher "glatter" Verlauf zu erwarten ist. Daneben kann es bei einzelnen Frequenzen "peaks" in einem Spektrum geben, die auf die Existenz einer zyklischen Komponente dieser Frequenz im Prozeß hinweist.

Im allgemeinen ist das Spektrum bzw. die Spektraldichte unbekannt. Liegt jedoch eine Realisation eines stationären Prozesses vor, d.h. eine (endliche) Zeitreihe, dann kann es bzw. sie geschätzt werden. Dazu gibt es mehrere Möglichkeiten, auf die hier nicht eingegangen werden kann.

Damit läßt sich der Unterschied zwischen Filter-Design und Spektralanalyse wie folgt zusammenfassen:

Bei der Spektralanalyse wird versucht, eine unbekannt Funktion auf der

Grundlage einer Stichprobe zu schätzen. Es liegt also ein Inferenzproblem vor, allerdings kein in der Statistik übliches, da kein(e) Parameter geschätzt wird (werden), sondern eine Funktion. Beim Filter-Design liegt dagegen ein Konstruktionsproblem vor. Werden z.B. IIR-Filter konstruiert, dann ergeben sich die Filter-Koeffizienten nicht via Schätzung aus einer vorliegenden Zeitreihe. Sie werden vielmehr ganz unabhängig von einer konkreten Reihe z.B. nach gewissen Optimalitätsprinzipien bestimmt. Dies hat mit statistischer Inferenz überhaupt nichts zu tun. Somit behandeln beide Gebiete nicht nur völlig verschiedene Fragestellungen, sie sind auch methodisch nicht vergleichbar.

5. Filter-und ARIMA-Modelle

Ein wesentlicher Schritt bei der ARIMA-Modellierung ist die Differenzbildung. Wie man leicht sieht, ist eine Bildung von Differenzen gleichbedeutend mit einer Filterung. Man kann zeigen, daß Differenzen nichts anderes sind als Hochpaßfilter. Leider handelt es sich dabei um ausgesprochen schlechte Hochpaßfilter. Verglichen mit der oben dargestellten Amplituden-Idealfom weisen sie vor allem zwei gravierende Nachteile auf:

a) Im Niederfrequenzbereich zeigt die Amplitudenfunktion eines Differenzenfilters einen nur langsam ansteigenden Verlauf, d.h. sie erreicht den Wert Eins erst bei relativ hohen Frequenzen.

b) Im Hochfrequenzbereich nimmt sie Werte an, die wesentlich größer als Eins sind (bei der höchsten Frequenz $\lambda = \pi$ ist $A(\pi) = 2$).

Hinzukommt, daß Differenzenfilter eine Phasenverschiebung zwischen der Originalreihe und der gefilterten Reihe bewirken.

Die Eigenschaft b) führt dazu, daß so gefilterte Reihen "aufgeraut" werden, also einen wesentlich unruhigeren Verlauf zeigen als die Ausgangsreihe. Mit steigender Differenzenordnung wird dieser Effekt noch verstärkt. Eigenschaft a) kann dazu führen, daß gerade die Informationen, für die man sich interessiert, verlorengehen. Diese Gefahr ist insbesondere bei der Suche nach "langen Wellen" gegeben, da solche, falls sie in einer Reihe vorhanden sind, bei sehr niedrigen Frequenzen zu suchen sind. Anders ausgedrückt: die "power", die solchen Wellen entspricht, residiert im Niederfrequenzbereich. Aber gerade dort weisen Differenzenfilter sehr kleine Werte auf, d.h. diese "power" wird durch die Filterung praktisch eliminiert. Ein Hochpaß, der für die vorliegende Fragestellung geeignet ist, müßte in diesem Bereich den Wert Eins annehmen. Nebenbei bemerkt, sind die gleichen Bedenken auch für eine

Suche nach "langen Wellen" auf der Basis von Wachstumsraten anzumelden, da diese approximativ als Differenzenfilter angesehen werden können.

Diese Defekte der Differenzenbildung schließen jedoch eine ARIMA-Modellierung von "langen Wellen" - Phänomenen keineswegs aus. Dazu ist es allerdings erforderlich, die Differenzenbildung durch eine der substanzwissenschaftlichen Fragestellung angemessene Hochpaß- bzw. Bandpaßfilterung (wenn man nur an der "Welle" interessiert ist) zu ersetzen. Deswegen sollte dann auch zutreffender nur von ARMA-Modellierung gesprochen werden. Ein Beispiel soll die Verknüpfung von Filtern und Modellierung veranschaulichen. Dazu wird die Jahresreihe "Gross Fixed Capital Formation in the United Kingdom 1830-1979" verwendet, die nachstehend zusammen mit ihrer "langen Welle" wiedergegeben ist. Diese ergibt sich aus der Originalreihe mit Hilfe eines Bandpasses, der alle zyklischen Bewegungen mit einer Periodenlänge von 20 bis 60 Jahren konserviert.

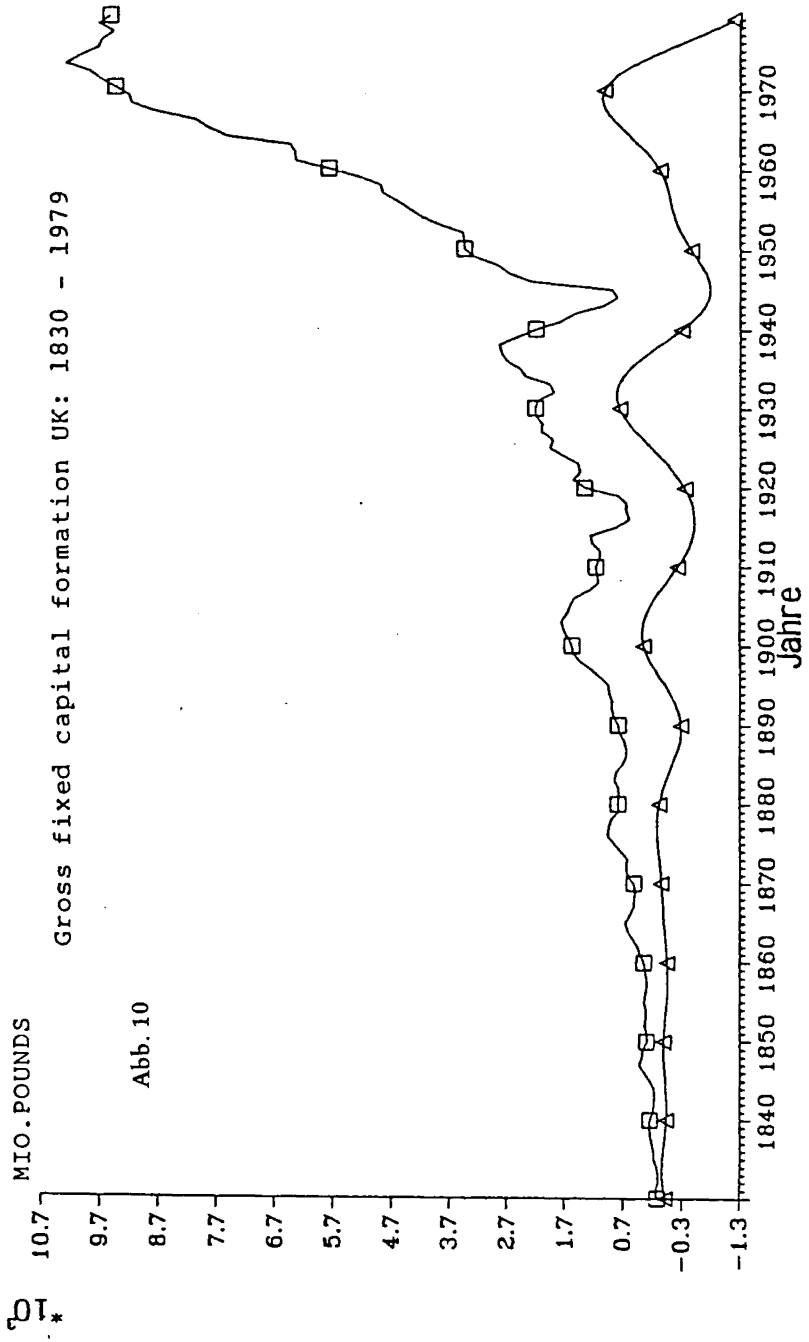
Versucht man die offensichtlich nicht-stationäre Originalreihe in der üblichen Weise zu modellieren, dann findet man als bestes Modell den ARIMA(0,1,4)-Prozeß:

$$X_t - X_{t-1} = u_t + 0.47u_{t-1} + 0.15u_{t-2} + 0.24u_{t-3} + 0.19u_{t-4}$$

Dieser Prozeß läßt keinerlei zyklische Struktur erkennen (dazu müßte ein AR-Teil mindestens der Ordnung zwei vorhanden sein). Die "Welle" läßt sich sehr gut durch den ARMA(2,1)- Prozeß

$$X_t - 1.889X_{t-1} + 0.943X_{t-2} = u_t - 0.45u_{t-1}$$

modellieren. Zu dieser Modellierung sei noch eine kurze Bemerkung angefügt: Da die "Welle" einen sehr glatten Verlauf zeigt (alle hochfrequenten Schwingungen wurden ja durch den exakten Bandpaß eliminiert), verbleiben nach der ARMA-Modellierung keine Residuen, die als weißes Rauschen interpretiert werden könnten - entgegen der "Box-Jenkins-Philosophie". Die betragsmäßig sehr kleinen Residuen zeichnen vielmehr den Verlauf der "Welle" nach. Diese Inkonsistenz mit dem Box-Jenkins-Ansatz läßt sich allerdings vermeiden, wenn man die Amplitudenfunktion des Bandpasses im Stoppband des Hochfrequenzbereiches nicht exakt auf Null setzt. Damit bleibt noch etwas "noise" in der "Welle" erhalten, und die erwähnte Inkonsistenz verschwindet.



10
*

Offensichtlich liegt es nahe, den hier vorgeführten Ansatz zu einem multivariaten zu verallgemeinern. Mit Hilfe von Transferfunktionen-Modellen könnten Abhängigkeiten von "langen Wellen" untersucht werden, was für historische Fragestellungen von Interesse sein dürfte. Das scheint aber noch Neuland zu sein. Bisher sind dazu keine Arbeiten bekannt geworden.

Literatur:

- Box, G.E.P., Jenkins, G.M.: Time Series Analysis: Forecast and Control, San Francisco 1976.
- Cryer, J.D.: Time Series Analysis, Duxbury Preß, Boston 1986.
- McCleary R., Hay R.A. jr.: Applied Time Series Analysis for the Social Science, Sage Publ. Beverly Hills and London, 1980.
- Metz, R.: Ansätze, Begriffe und Verfahren der Analyse ökonomischer Zeitreihen, in: Historical Social Research / Historische Sozialforschung, Vol. 13, No. 3 (1980), S.23-103.
- Metz, R.: Zur empirischen Evidenz "langer Wellen", in: Kyklos, Bd. 37, H. 2 (1984), S.266-290.
- Stier, W.: Konstruktion und Einsatz von Digitalfiltern zur Analyse und Prognose ökonomischer Zeitreihen, Westdeutscher Verlag, Opladen 1978.
- Stier, W.: Saisonbereinigungsverfahren - Ein Überblick, in: OR - SPEKTRUM, 1985, 7, S.63-79.
- Stier, W.: Methoden der Zeitreihenanalyse, Springer-Verlag, erscheint 1991.

Lange Wellen im Bildungssystem

von Volker Müller-Benedict

1. Einleitung

In der Geschichte der akademischen Karrieren gab es immer wieder generelle und fachspezifische "Überfüllungskrisen", in denen wesentlich mehr examinierte Studenten die Hochschulen verließen, als aktuell vom Beschäftigungssystem aufgenommen wurden. Zur systematischen Untersuchung dieses und damit zusammenhängender Phänomene wurden von der Göttinger Gruppe des DFG-Projekts "Qualifikationskrisen" die in historischen Quellen vorhandenen Daten über die in den letzten 170 Jahren (bis 1945) an deutschen Universitäten Studierenden aufgenommen und für die Datenverarbeitung verfügbar gemacht ¹.

Die in fast allen der erhobenen Zeitreihen deutlich sichtbaren Schwingungen ließen die alte Frage nach der Existenz von "langen Wellen", die früher in der Ökonomie diskutiert wurde ², für einen anderen Teilbereich der Gesellschaft, das Bildungssystem, neu entstehen. Eine mögliche Antwort auf diese Frage soll in diesem Beitrag gegeben werden.

Außer dieser Frage fiel im Rahmen sozialhistorischer Untersuchungen ³ die Unabhängigkeit der Variablen des universitären Bildungssystems von wichtigen gesellschaftlichen Einflußgrößen wie z.B. demographischer Entwicklung, Wirtschaftskonjunktur, dirigistischen Eingriffen auf. Dafür nur kurz zwei Beispiele: 1. Die Mangelphase in Preußen (1860 - 1875), in der ein allgemeines Defizit an Hochschulabsolventen spürbar wird, fällt in eine Zeit vollkommen linearer demografischer Entwicklung und allgemein guter Wirtschaftsaussichten. 2. Die sofort nach der Machtergreifung von den Nationalsozialisten erlassenen NC-Quoten für ein Hochschulstudium zur Eindämmung der Überfüllung, die an allen Universitäten des Deutschen Reiches seit 1930 bestand, wurden vom eigendynamischen Rückgang der Studentenzahlen so weit unterschritten, daß sie gegenstandslos waren und nach einem Jahr wieder abgeschafft wurden ⁴.

¹ Wichtige Teile dieser Daten wurden bisher im "Datenhandbuch zur deutschen Bildungsgeschichte. Teil 1" (TITZE 1987) veröffentlicht.

² S. z.B. SPREE 1978.

³ S. TITZE/NATH/MÜLLER-BENEDICT 1985.

⁴ S. NATH 1988.

Dadurch wird die zweite Frage aufgeworfen, wieviel Eigendynamik das universitäre Bildungssystem besitzt, ein wie großer Teil des Auf und Ab der Zeitreihen, d.h. von Überfüllung und Mangel an den Hochschulen, sich auf lange Wellen zurückführen läßt.

Weil lange Wellen nur mit Hilfe von Indikatoren analysiert werden können, die aus den historischen Quellen über lange Zeit als relativ unveränderte Kategorie erfaßt werden können, wurden für die Analyse die folgenden fünf Zeitreihen von Studierenden mit historisch seltener Länge ausgewählt:

- * Für die Universität Tübingen gibt es die Zahl der Studierenden der evangelischen Theologie von WS 1760 bis Wintertrimester 1941,
- * für die Universität Göttingen die Studierenden der evangelischen Theologie, von Jura und Allgemeiner Medizin und die Zahl der Studierenden insgesamt von SS 1767 bis Wintertrimester 1941 jeweils in ununterbrochener Reihenfolge ⁵.

Die Ergebnisse für diese speziellen Zeitreihen lassen sich jedoch auf die große Mehrheit der Studentenströme an allen Universitäten generalisieren. Das zeigt einerseits direkte Berechnung, andererseits wird es durch die Tatsache erklärt, daß für ein Studium an einer deutschen Hochschule im Großen und Ganzen überall ähnliche Bedingungen bestanden, was Immatrikulationsvoraussetzungen, Studiendauer, Studienfachwahlmöglichkeiten etc. anbelangt.

Wegen der Länge der Zeitreihen - über 174 Jahre deutscher Bildungsgeschichte - stellt sich als dritte Frage die Frage nach strukturellen Veränderungen, die in dieser Zeit stattgefunden haben und die Geschichte des Universitätsbesuchs in so unterschiedliche Abschnitte unterteilen, daß man von Strukturbrüchen sprechen kann.

Diesen drei Fragen ist schon in der Geschichte und auch heute in sozial-historischen Forschungen nachgegangen worden. Zyklische Entwicklungen, die sich quasi naturwüchsig "wie Ebbe und Flut" durchsetzen, wurden be-

⁵ Um die kurzfristigen Schwankungen von Sommer- auf Wintersemester auszugleichen, die u.a. von unterschiedlichen Abiturterminen herrühren, wurden aus Sommer- und Wintersemester (bzw. aus den 3 Trimestern der Jahre 1919, 1939 und 1940) ein Jahresdurchschnittswert errechnet. Innerhalb des langen Zeitraums liegen einige singuläre historische Ereignisse, wie z.B. die Freiheitskriege, die Revolution 1848, der 1. Weltkrieg. Es wurde darauf verzichtet, die Zeitreihen zu diesen Zeitpunkten irgendwie zu bereinigen, da sie erstens keine offensichtlichen Ausreißerwerte produzieren und zweitens durch bestimmte Zählmethoden, z.B. die Einbeziehung der Beurlaubten im 1. Weltkrieg, eine gewisse statistische Kontinuität gewährt bleibt.

schrieben. Die Eigendynamik der realen Entwicklung machte schon Anfang dieses Jahrhunderts den Versuch des Göttinger Volkswirtschaftlers LEXIS, bestimmte normale, gesellschaftlich wünschbare Studentenzahlen für einzelne Studienfächer zu ermitteln, letztlich zunichte ⁶.

Hier sollen die Zeitreihen mit Methoden untersucht werden, die dem naturwissenschaftlichen und ökonomischen Repertoire entstammen, Methoden der statistischen Zeitreihenanalyse. Dazu sehen wir zunächst von der Tatsache ab, daß diese Zeitreihen durch die Entscheidungen vieler historischer Subjekte zustandegekommen sind und daß wir selbst Subjekte in dem noch andauernden gesellschaftlichen Zusammenhang sind, dem die Zeitreihen entstammen, verlegen also die Zeitreihen quasi in einen naturwissenschaftlichen Objektbereich. Dort haben wir die Möglichkeit, sie mit Hilfe der statistischen Methoden dem harten Test zu unterwerfen, was sie ohne Rücksicht auf die historischen Zusammenhänge aus sich heraus aussagen und welche Gesetzmäßigkeiten in ihnen unabhängig von sozialwissenschaftlichen Betrachtungsweisen stecken.

Was könnte der Gewinn einer solchen Untersuchung sein? Die Methoden könnte man als eine über das Heuristische hinausgehende, den optischen Eindruck quasi mikroskopisch erweiternde Beschreibung der äußeren Gestalt der Zeitreihe ansehen. Ihre Ergebnisse haben deshalb in dem Sinn Objektivität, als sie in jeder sozialhistorischen Analyse in irgendeiner Weise berücksichtigt werden müssen, weil jede Analyse letztlich auch die Gestalt der Zeitreihen erklären muß.

Die Ergebnisse selbst sind an sich nur - wenn auch raffinierte - Beschreibung, Abbild der Geschichte. Würde man sie als Erklärung nehmen, triebe man Ideologie. Sie wirken erst aufklärerisch dadurch, daß sie auf ihren Sinn in sozialhistorischen Analysen geprüft werden können. Es muß nachweisbar sein, daß die statistischen Einzelergebnisse tatsächlich als Resultierende historischer Abläufe und Handlungsstränge wahrscheinlich sind. Dann jedoch erhält die sozialhistorische Analyse durch die statistischen Ergebnisse eine über die Ansichten der analysierenden Forscher hinausgehende Bedeutung, weil sie Strukturen erklärt, die genau zu den Parametern führen, die objektiv, d. h. vom forschenden Subjekt unabhängig, in den Zeitreihen festgestellt werden können.

⁶ LEXIS fertigte 1889 und 1891 für den preußischen Kultusminister zwei Denkschriften über die Normalzahl der Studierenden an. 1905 bemerkte er in einer neuerlichen Analyse über "Bedarf und Angebot in den gelehrten Berufszweigen", nachdem er wichtige "Normalzahlen" der Ärzteentwicklung korrigieren mußte: "Eine eigentliche Normalzahl der Studierenden läßt sich für diesen frei zugänglichen Beruf nicht feststellen".

Um diese Methoden anwenden zu können, müssen wesentliche Voraussetzungen über die Zeitreihen gemacht werden. Sie werden in Kap. 2 erläutert. In Kap. 3 werden die Frage des Strukturbruchs behandelt und weitere technische Voraussetzungen erörtert. Kap. 4 analysiert das Vorhandensein und die Länge von Zyklen, Kap. 5 die Frage des Anteils an Eigendynamik, der durch diese Zyklen erklärt werden kann. In Kap. 6 werden dann die vorangehenden Ergebnisse auf ihren Sinn in einer sozialhistorischen Analyse untersucht und ihre historische Erklärung versucht.

Die Methoden selbst sind zu elaboriert, um sie hier ausführlich darstellen zu können⁷. Andererseits können aber die Ergebnisse, die auf ihnen beruhen, ohne ihre Kenntnis nicht zuverlässig eingeschätzt werden. Deshalb mußte versucht werden, die grundsätzlichen Voraussetzungen, Ansätze und Probleme der Methoden deutlich werden zu lassen.

2. Die methodischen Voraussetzungen

Die bisher aus den sozialhistorischen Analysen gewonnenen Erkenntnisse über diese Fragen lassen sich in der Annahme zusammenfassen, daß im universitären Bildungssystem eigendynamische Prozesse wirkten und lange Wellen erzeugten, die sich trotz der dauernden Veränderung der aktuellen historischen Situation immer wieder durchsetzten und bestimmend für die Höhe seiner Variablen waren. Diese Annahme benötigen wir, um als wichtigste Voraussetzung für die statistischen Methoden davon ausgehen zu können, daß wir es hier mit einem sog. "stochastischen Prozeß" zu tun haben. Für die hier betrachteten Zeitreihen bedeutet das, daß ihre Zukunft mit berechenbarer Wahrscheinlichkeit von ihrer Vergangenheit abhängt.

Es scheint zunächst selbstverständlich, daß z.B. die Studentenzahl eines folgenden Jahres, wenn sie in den letzten fünf Jahren 800, 850, 900, 940, 1000 betrug, mit großer Wahrscheinlichkeit um die 1050 herum liegen wird und mit sehr viel geringerer Wahrscheinlichkeit etwa 200 oder 2000 betragen wird. Die Forderung der Berechenbarkeit dieser "Übergangswahrscheinlichkeiten" und ihrer Konstanz über abgrenzbare Zeiträume hinweg zwingt die Zeitreihe jedoch in einen festgelegten Rahmen (der mathematisch stochastischer Prozeß heißt), der, statistisch gesehen, d.h. mit mehr oder weniger großen, im Mittel feststehenden Abweichungen, ihr eine immer gleiche Gestalt verleiht. Diese statistische Definition von "gleichem" Aussehen der ersten und letzten Teile der Zeitreihe ist jedoch so weit, daß das naive menschliche Auge diese Art von Ähnlichkeit im Aussehen einer Zeitreihe nicht entdecken kann. Ein

⁷ S. KOOPMANS 1979, BOX/JENKINS 1976.

Versuch, diesen Rahmen, der die Gestalt der Zeitreihe bestimmt, real-historisch zu erfassen, wäre etwa, ihn als alle strukturellen Gegebenheiten anzusehen, die am Zustandekommen von Studentenzahlen beteiligt sind, also z.B. die Laufbahnstrukturen (nicht die aktuelle Lage auf dem Arbeitsmarkt), die Immatrikulationsvoraussetzungen, die Bildungsinstitutionen etc.

Für die Annahme, diese Zeitreihen als stochastische Prozesse aufzufassen und in so starkem Maße von ihrer Vergangenheit abhängig zu machen, spricht eine wichtige historische Einsicht: Daß es einen Zusammenhang von vorangehender Mangel- und darauffolgender Überfüllungsphase an den Universitäten gibt und umgekehrt, daß hier über Jahrzehnte hinweg ein historischer Prozeß sich vollzieht, der in der aktuellen Wahrnehmung der meisten Zeitgenossen als ein- und erstmaliger Mangel oder Überfüllung erscheint, aber tatsächlich einer historischen Entwicklung entstammt, die über Generationen währt.

Das ist keineswegs selbstverständlich und kann von Historikern genauso gut bezweifelt werden. In diesem Fall hätte jede der aufgetretenen Überfüllungskrisen ihre eigene, nur aus der jeweiligen historischen Konstellation stammende, von der vorhergehenden Welle unabhängige Genese. So könnte z.B. der Hauptgrund für die Überfüllungskrise im Vormärz (Ende der 1820er bis Mitte der 1840er Jahre) der Nachholbedarf an Bildung wegen der vorhergehenden Befreiungskriege sein (in denen u.a. zeitweise alle Universitäten geschlossen waren) oder die große akademische "Berufsnot" Anfang der 1930er Jahre auch als eine Folge des durch den 1. Weltkrieg verlorengegangenen "deutschen Lebensraums" angesehen werden, wie es tatsächlich behauptet wurde ⁸.

Die Methoden der Zeitreihenanalyse wurden zunächst für die Analyse physikalischer Schwingungsphänomene entwickelt. Sie werden z.B. in der medizinischen Diagnostik von Sprachstörungen (Prediktorenanalyse), in der Akustik, bei der Sternsystemberechnung, U-Boot-Sonarortung und seit längerem auch in der Ökonometrie angewandt. Daß die Voraussetzung der Methoden für physikalische oder biologische zeitabhängige Schwingungsprozesse erfüllt ist, liegt in ihrer Natur. Bei unseren historischen Zeitreihen ist das Vorhandensein von Schwingungen offensichtlich, aber ihre Herkunft aus einem stochastischen Prozeß kann nur - und das sei nochmals betont - vorausgesetzt, nicht bewiesen werden.

Welchen praktischen Nutzen bringen diese ausgefeilten Methoden für die Beschreibung historischer Zeitreihen, wenn man dafür diese einschränkenden

⁸ S. z.B.: SCHAIRER, R.: Die akademische Berufsnot. Jena, 1932.

Voraussetzungen machen muß? Sie erlauben es zunächst, innerhalb ihres statistischen Untersuchungsrahmens Begriffe wie Zyklus, Zykluslänge, Strukturbruch, Eigendynamik usw. als berechenbare Größen zu definieren und als Quantitäten in den Daten nachzuweisen. Damit ist ein Instrument vorhanden, die Fülle der optisch-subjektiven Eindrücke von "Zyklen" in den Zeitreihen und der durch Analyse der historischen Quellen gewonnenen Einsichten über wiederkehrende, eigendynamische Entwicklungen zusammenzufassen und zu objektivieren, allerdings unter einem rein quantitativen, ohne Interpretation inhaltsleeren Gesichtspunkt.

Wichtiger noch ist ihre Fähigkeit, Strukturen in den Zeitreihen zu entdecken, die einer einfachen Betrachtung entgehen. So können sie z.B. eine mit höchst unterschiedlicher Zyklusdauer schwingende Zeitreihe als ein Gemisch von nur zwei Zyklen mit jeweils fest definierter Länge erkennen oder für zwei Zeitreihen mit ähnlicher Gestalt ganz unterschiedliche Bildungsgesetze nachweisen.

Wenn wir also im folgenden die Zeitreihen für unsere Analyse als Pfade stochastischer Prozesse annehmen müssen, richten wir analytisch unser Augenmerk ausschließlich auf den langfristigen Zusammenhang der historischen Entwicklung und blenden die Möglichkeit, daß viele unabhängige kurzfristige Ereignisse die Gestalt der Zeitreihe entscheidend beeinflußt haben, aus. Sie können unter dieser Annahme nicht mehr auftreten, oder höchstens indem sie als Bedingungen der langfristigen Struktur angesehen werden, so z.B. so wichtige wie die Einführung des Abiturs als Immatrikulationsvoraussetzung 1788, die Revolution 1848 oder die Inflation 1924. Gleichzeitig jedoch müssen wir wegen der Unsicherheit bei der Grundvoraussetzung nachweisen, daß die statistisch festgestellten Zyklen und alle anderen auf diese Weise statistisch festgestellten Fakten überhaupt historisch einen Sinn haben. Denn sollte das nicht der Fall sein, müßten wir eher zum Schluß kommen, daß die Grundvoraussetzung eines stochastischen Prozesses falsch ist, als daß wir historisch bisher unbekannte Zyklen konstatieren, für deren Erklärung dann ganz neue historische Gründe gefunden werden müßten⁹.

⁹ Hier möchte ich weiter gehen als SPREE, S.39: "Die Erklärung der die Modelle und ihre Ergebnisse modifizierenden sozioökonomischen Strukturveränderungen, formal ausgedrückt: der Variationen der das Modell kennzeichnenden Parameterkonstellationen, die angesichts der Langfristigkeit der Prozesse mit Notwendigkeit stattfinden, kann nur im Sinne historisch-spezifischer Analysen erfolgen". Nicht nur die Variation der Parameter, sondern auch die Größe der Parameter muß durch historisch-spezifische Analysen erklärbar sein.

3. Stationarität und Strukturbruch

Die Methoden benötigen zweitens eine technische Voraussetzung, die Stationarität der Zeitreihen. Sie dürfen in Mittelwert und Varianz nicht wachsen (kein Trend), und die Art der Abhängigkeit der zukünftigen von den vorangehenden Werten muß über die Zeit gleich bleiben. Diese Abhängigkeit wird statistisch u.a. durch die "Autokorrelationsfunktion" beschrieben¹⁰. Die ersten beiden Forderungen lassen sich durch geeignete Transformationen der Zeitreihen, eine spezifische Art der Trendbereinigung, erfüllen. Die Analyse der Autokorrelationsfunktion läßt als erstes Ergebnis eine Aussage über Strukturveränderungen zu

Bei den hier vorliegenden Zeitreihen treten alle diese Arten von Nichtstationarität auf. Die Studierenden der evangelischen Theologie in Göttingen bilden die einzige Zeitreihe, die im betrachteten Zeitraum nicht wächst. Ihre Daten werden deshalb in den folgenden Berechnungen immer absolut verwendet¹¹.

Alle anderen Zeitreihen wachsen. Zugleich mit der Erhöhung des Niveaus wachsen die Schwankungen um das Niveau, die Ausschläge der langen Wellen. Um diese Form der Nichtstationarität zu eliminieren, werden zunächst alle Reihen logarithmiert. Dieses Vorgehen entspricht dem natürlichen Verständnis, daß eine Zunahme absoluter Zahlen von 40 auf 60 eine etwa gleiche Bedeutung hat wie eine Zunahme von 4000 auf 6000, nämlich jeweils um 50%, und ganz verschieden ist von einer Zunahme von 4000 auf 4020. Dieses natürliche Verständnis kann man z.B. in unserem Fall mit der Tatsache erklären, daß die Inanspruchnahme der personellen und räumlichen Kapazitäten der Universitäten immer im Verhältnis zur vorherigen Situation als Über- bzw. Unterbelastung empfunden wird, nicht in der absoluten Differenz.

Die wachsenden Zeitreihen müssen vom Trend bereinigt werden. Trend und lange Wellen sind jedoch in der Praxis nur schwer voneinander abzugrenzen, da sie sich nur quantitativ voneinander unterscheiden: Trends können als Zyklen mit einer Länge, die größer als die Zeitreihe selbst ist, aufgefaßt werden; lange Wellen sind Zyklen, deren Länge in der Größenordnung

¹⁰ Die Autokorrelationsfunktion gibt die Korrelation der Zeitreihenwerte mit sich selbst an, und zwar für alle Abstände von 1,2,3,...usw. Jahren. Sie ist also eine Funktion der zeitlichen Abstände, deren Wert zwischen 1 und -1 schwankt und - wie bei dem bekannten Korrelationskoeffizient - je nach Nähe an $+/-1$ oder 0 einen starken oder schwachen Zusammenhang der Zeitreihenwerte mit diesem zeitlichen Abstand anzeigt.

¹¹ Sie können damit auch als Vergleichsgrundlage für etwaige Nebenwirkungen der Transformationen dienen. S. die Tabellen.

der Zeitreihenlänge nahe kommt¹². Da wir aber für unsere Analyse notwendigerweise an der vollständigen Beibehaltung der langen Wellen interessiert sind, bedienen wir uns zur Trendbereinigung einer Methode, die genau auf die Zyklenlänge abgestimmt werden kann: des sog. Kerbenfilters¹³.

Eine Trendbereinigung mit einer deterministischen Trendfigur - gebräuchlich sind Polynome oder gleitende Durchschnitte - schlägt sich in unbekanntem oder verzerrenden Veränderungen der Autokorrelationsfunktion der Zeitreihe nieder. Filtert man die Zeitreihen mit dem Kerbenfilter, so enthalten sie keine zyklischen Komponenten von Länge größer als die Zeitreihenlänge mehr, alle kürzeren Zyklen sind jedoch unverfälscht erhalten¹⁴. Da Trends immer als Zyklen, die länger als die Zeitreihe selbst sind, aufgefaßt werden können, werden sie durch den Kerbenfilter eliminiert.

Diese beiden Operationen zur Trendbereinigung, die Logarithmierung und die Kerbenfilterung, sind insofern rein technischer Natur, als sie in bezug auf die zu untersuchenden Fragestellungen keine Bedeutung für die sozialhistorische Interpretation haben; das Wachstum dieser Zeitreihen soll hier nicht untersucht werden. Sie sind ebenfalls so gewählt, daß sie für die weitere statistische Analyse nicht besonders berücksichtigt werden müssen. Daß die beiden Transformationen der Logarithmierung und Kerbenfilterung auch vom optischen Eindruck her in etwa das aus der Gestalt der Zeitreihe herausnehmen, was man als Einfluß eines langfristigen Trends empfinden kann, zeigt Bild 1 am Beispiel aller Studierenden in Göttingen.

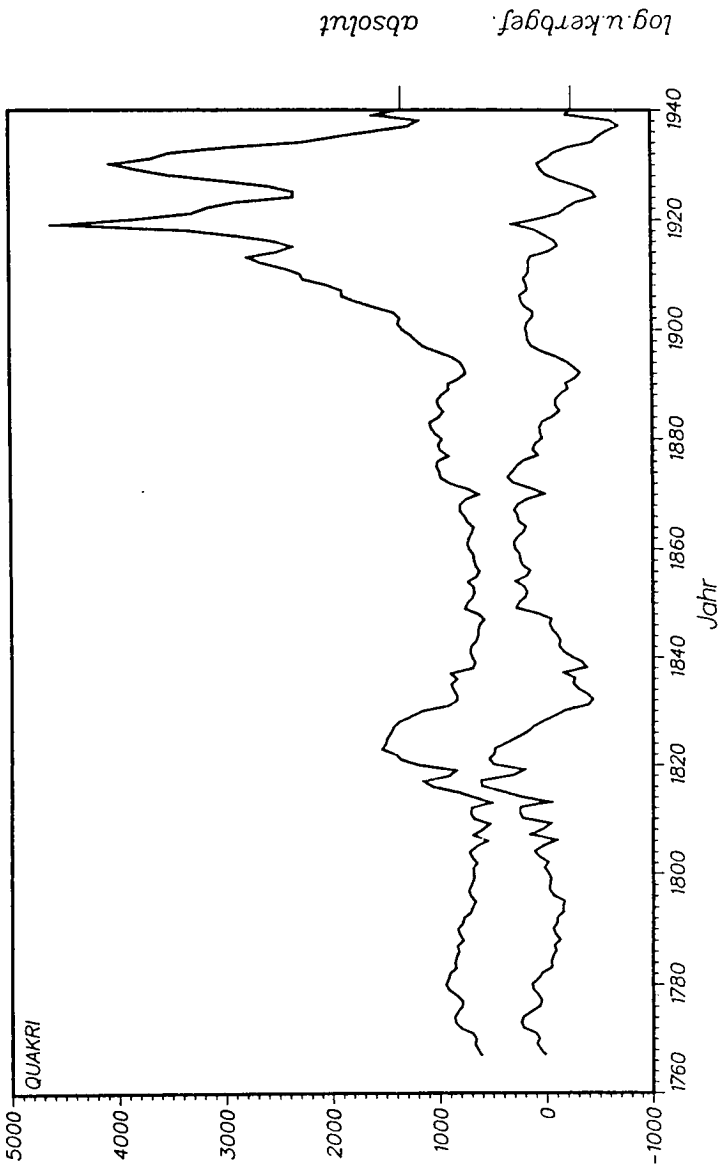
Die letzte Art von Nichtstationarität ist die Veränderung der Autokorrelationsstruktur. Sie rührt von denjenigen historischen Entwicklungen her, die die Struktur des Hochschulbildungssystems grundlegend verändern. Diese Entwicklungen müssen nämlich letztlich auch auf die Art der Abhängigkeit der zukünftigen von den vergangenen Zeitreihenwerten durchschlagen, also statistisch als anderer stochastischer Prozeß mit anderer Autokorrelationsfunktion identifizierbar sein. Wenn wir berücksichtigen, daß der Zugang zur Universität seit der Einführung des Abiturs in Preußen 1788 bis heute zumindest offiziell durch denselben Mechanismus allgemeingültiger gleicher Leistungskriterien bestimmt wird, und daß die Universitäten als Orte mit relativ starker Autonomie in der Verwaltung beharrlicher als vielfach anderswo ihre

¹² Zur Diskussion über die Abgrenzung von Trends und Konjunkturzyklen s.a. SCHULTE 1981, S.108ff.

¹³ S. SCHULTE 1981.

¹⁴ Zum Konzept der Zerlegung der Zeitreihe in zyklische Komponenten s. Kapitel 4. Die Durchlaßfunktion des Filters, die anzeigt, welche Frequenzen mit welchem Faktor zwischen 0 und 1 "durchgelassen" werden, zeigt Bild 4.

Bild 1: Die Studierenden an der Universität Göttingen von 1767- 1940. Die Zeitreihe in absoluten Zahlen und als transformierte Version.



überkommene Struktur beibehalten konnten, soll hier zunächst trotz des langen Zeitraums eine gleichbleibende Struktur der Prozesse angenommen werden.

Zur Prüfung dieser Annahme gleichbleibender Struktur müssen die Zeitreihen gesplittet und die Struktur der Teile miteinander verglichen werden. Wenn Strukturveränderungen stattgefunden haben, so sicher nicht plötzlich durch punktuelle Ereignisse wie Reichsgründung, Regierungswechsel o.ä. Denn das Bildungsverhalten der Bevölkerung kann sich als kürzeste Möglichkeit im Generationenabstand ändern, sich also immer nur ein allmählicher Strukturwandel ergeben. Deshalb kann eine Aufteilung der Zeitreihen nicht von vornherein nach bestimmten historischen Ereignissen gewählt werden. Die Zeitreihen wurden deshalb nach pragmatischen Gesichtspunkten¹⁵ halbiert (um 1850 geteilt) und gedrittelt, und als Prüfgrößen die jeweiligen Varianzen, Autokorrelationsfunktionen (ACF) und Spektren¹⁶ berechnet.

Zunächst ergab sich, daß bei Halbierung der Zeitreihen die Unterschiede in den Prüfgrößen der Teile größer waren als bei Vergleichen der Drittel. Deshalb wurde die Hypothese der Strukturkonstanz nur für die halbierten Zeitreihen weiter geprüft. Ein F-Test für die jeweiligen Varianzverhältnisse der beiden Teile ergibt signifikante Unterschiede für die Evang. Theologie Tübingen sowie Jura und Medizin Göttingen¹⁷.

Die zweite und dritte Prüfgröße sind die ACFs und Spektren der beiden Teile der Zeitreihen. Als Beispiel für das unterschiedliche Aussehen der ACF in beiden Teilen diene hier die Zeitreihe der Jurastudenten in Göttingen (Bild 2). In Bild 2 ist zu sehen, daß für die Zeit vor 1850 hauptsächlich Zyklen länger als 32 Jahre vermutet werden können: Das Minimum der ACF (stärkster negativer Zusammenhang der Zeitreihenwerte) liegt bei 16 Jahren, d.h. die Kombination Minimum und darauffolgendes Maximum tritt in der

¹⁵ Das größte praktische Hindernis ist die Kürze der Zeitreihen. Zur Entdeckung von Zyklen sollten Zeitreihen mindestens eine Länge vom Mehrfachen des längsten auftretenden Zyklus haben. In diesem Fall von 174 Jahresdaten ist deshalb Drittelung die unterste Grenze.

¹⁶ Zur Erklärung des Spektrums s. nächstes Kapitel.

¹⁷ F-Wert 1,47; 1,75 und 2,39; bei kritischer Schranke von 1,45 auf dem 5%-Niveau (F-Wert Evang. Theologie und alle Studierenden Göttingen: 1,16 und 1,13). Obwohl die Voraussetzung der Normalverteilung für den F-Test bei autokorrelierten Zeitreihen nicht gegeben ist, kann der Test erste Anhaltspunkte geben, da die beiden zu prüfenden Varianzen durch die in beiden Fällen vorhandene positive Autokorrelation in der gleichen Richtung unterschätzt werden. Zur Durchführung des F-Tests muß die Zeitreihe so betrachtet werden, als ob alle vorkommenden Werte in dem gleichen Jahr erhoben worden wären.

Bild 2: Autokorrelationsfunktion der Zeitreihe Jurastudenten in Göttingen

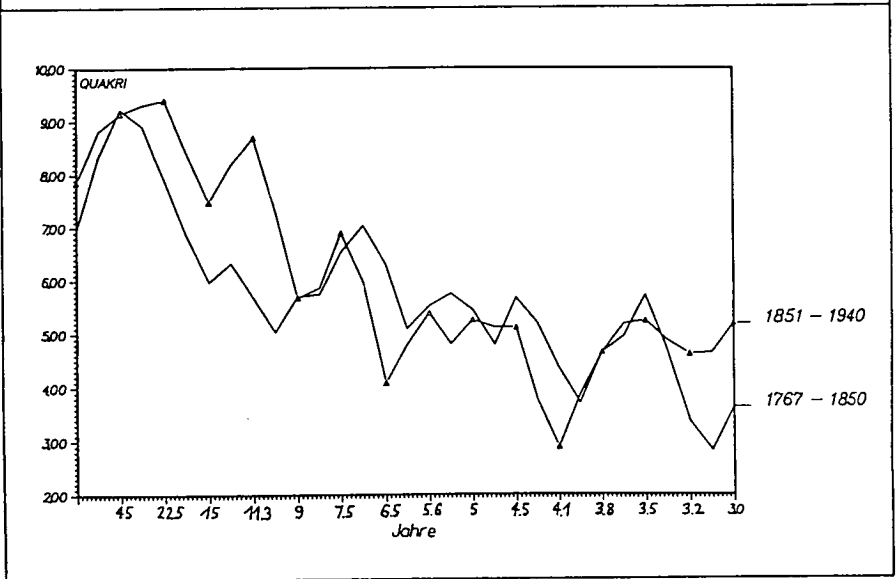
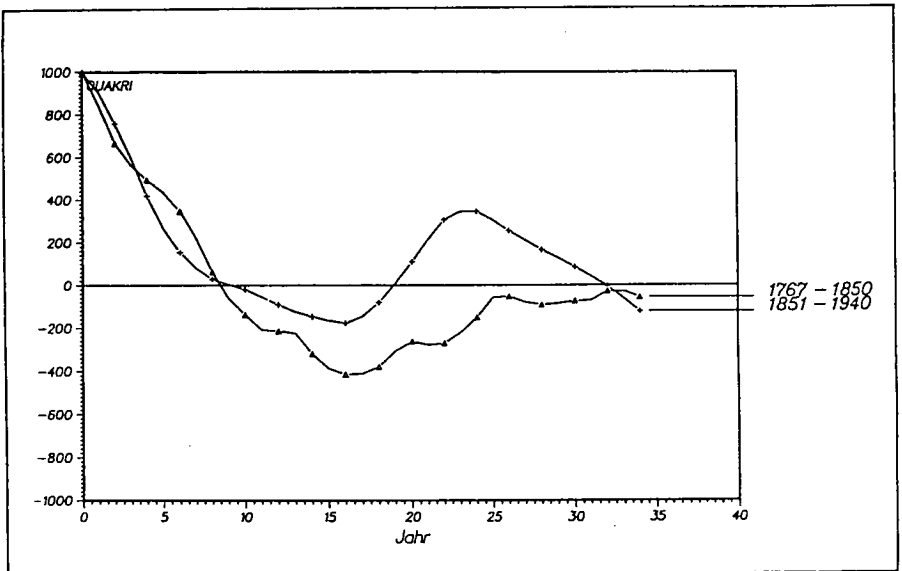


Bild 3: Logarithmiertes Spektrum der Zeitreihe Jurastudenten in Göttingen

Zeitreihe vor 1850 am häufigsten im Abstand von 16 Jahren auf. Nach 1850 gibt es kürzere Zyklen: Das Maximum der ACF (stärkster Zusammenhang der Zeitreihenwerte) liegt bei 21 Jahren, d.h. zwei aufeinanderfolgende Maxima oder zwei aufeinanderfolgende Minima treten in der Zeitreihe nach 1850 am häufigsten im Abstand von 21 Jahren auf. Dieser Abstand von 21 Jahren kann jedoch auch Folge zweier oder mehr Zyklen sein, die gleichzeitig wirken, aber verschiedene Längen haben. Die Unterscheidung dieser Zyklengemische gelingt mit der Spektralanalyse¹⁸. Für alle anderen Zeitreihen ergeben sich ebenfalls deutliche Unterschiede im Aussehen von ACF und Spektrum der zwei Teile.

Auf weitere Tests wurde angesichts der offensichtlichen Unterschiede (die im nächsten Kapitel statistisch signifikant werden) und der geringen Aussagefähigkeit der Tests¹⁹ verzichtet. Alle drei Prüfgrößen zusammen ergeben, daß die Annahme der Konstanz der Autokorrelationsstruktur aufgegeben werden muß. Die Zeitreihen werden im folgenden immer in beiden Teilen, bis 1850 und nach 1850, untersucht. Als erstes Ergebnis bleibt festzuhalten, daß nach Mitte des letzten Jahrhunderts im tertiären Bildungssystem eine entscheidende Strukturveränderung stattgefunden hat. Ihre Auswirkungen auf die Gestalt der Zeitreihen soll im folgenden untersucht werden.

4. Die zyklische Struktur

Die Frage, ob und welche langen Wellen statistisch in diesen Zeitreihen nachweisbar sind, wurde zunächst anhand der Spektren der Zeitreihen überprüft. Das Wort Spektrum ist bekannt aus der Farbenlehre. Ein Prisma zerlegt das weiße Licht in die Regenbogenfarben, das sogenannte Farbenspektrum. Das weiße Licht ergibt sich als Summe aller Regenbogenfarben, genauer: die Frequenz des weißen Lichts ist die Summe der einzelnen Frequenzen der Regenbogenfarben. Denkt man sich die fließenden Übergänge zwischen den Farben des Regenbogens als scharfe Grenzen, so wäre die Frequenz des weißen Lichts die Summe der sechs Frequenzen der Grundfarben. In genau der gleichen Weise zerlegt die Spektralanalyse eine Zeitreihe (das "weiße Licht") in die in ihr enthaltenen Zyklen (die Regenbogenfarben). Dadurch wird die Zeitreihe als Summe der in ihr enthaltenen zyklischen Komponenten darstellbar, also etwa als Summe von drei Zyklen von 40, 15 und 7 Jahre Länge. Der aktuelle Wert der Zeitreihe zu einem Zeitpunkt t ergibt

¹⁸ S. nächsten Abschnitt, ebenso die Interpretation von Bild 3.

¹⁹ S. MCCLEARY/HAY 1980, S.121: "Yet none of these tests is as compelling as the visual evidence".

sich dann aus der Summe aller vorkommenden Zykluswerte zu diesem Zeitpunkt. Die sind im einzelnen davon abhängig, in welcher Phase sich der jeweilige Zyklus gerade befindet (Aufschwung, Abschwung) und welches Gewicht er im Verhältnis zu den anderen Zyklen hat.

Der heuristische Wert einer Spektralanalyse im Vergleich zu einer Auszählung der Maxima und Minima oder anderen "analysis by eye"-Methoden liegt eben darin, daß die Schwingungen der Zeitreihe als Gemische von unabhängig voneinander entstandenen, aber in ihr zusammen wirkenden Zyklen erkannt werden, und daß die Länge dieser Zyklen angegeben werden kann. Das ist spätestens bei drei gleichzeitig wirkenden Zyklen durch andere Methoden nicht mehr möglich. Insbesondere kann durch diese Zerlegung festgestellt werden, welche Zyklen hauptsächlich die Gestalt der Zeitreihe bestimmen.

Das Spektrum einer Zeitreihe ist eine Funktion, die zu jeder Frequenz einen Wert angibt, der das Gewicht des Zyklus mit dieser Frequenz²⁰ darstellt, das dieser im Vergleich zu den anderen Frequenzen für die Zeitreihe hat. Besteht z.B. eine Zeitreihe nur aus zwei Zyklen, einem 20-jährigen und einem 3-jährigen, wobei der 20-jährige doppelt so viel zum Aussehen der Zeitreihe beitragen soll wie der 3-jährige, so hat das Spektrum der Zeitreihe nur zwei Werte, und zwar bei $1/20$ und bei $1/3$, und der Wert bei $1/20$ ist doppelt so hoch wie der bei $1/3$.

Die Übergänge zwischen den Regenbogenfarben sind jedoch fließend, so daß letztlich für das weiße Licht alle Frequenzen notwendig sind. Genauso ist es bei einer Zeitreihe, die Pfad eines stochastischen Prozesses ist. Ihr Spektrum enthält ebenfalls alle Frequenzen, jedoch mehr oder weniger stark. Die "peaks" (Maxima) dieser Spektren zeigen die Zyklen an, die das Hauptgewicht für das Zeitreihenaussehen haben. Es gibt zusätzlich ein statistisch definiertes Kriterium, das anzeigt, wann Peaks signifikant sind.

Als Beispiel sind die logarithmierten Spektren der Reihe Jurastudenten Göttingen in Bild 3 und Medizinstudenten Göttingen in Bild 5 zu sehen. Die Peaks, d.h. die wichtigsten Komponenten, lassen sich deutlich im linken Bereich erkennen. Ebenso deutlich ist das unterschiedliche Aussehen in den zwei verschiedenen Zeiträumen. Ab 1850 tritt ein zweiter Peak von Zyklen um die 11 Jahren Länge auf, der vor 1850 nicht vorhanden ist.

Welche Art von Zyklen zeigt nun die Spektralanalyse für die Zeitreihen auf? Es treten generell längere (um die 30 Jahre), mittlere (10 - 15 Jahre) und

²⁰ Zyklus und Frequenz hängen wie folgt zusammen: Ein Zyklus von 10 Jahren hat die Frequenz von $1/10$ (nämlich pro Jahr), ein Zyklus von 20 Jahren die Frequenz $1/20$ usw.

kürzere (ca. 7 Jahre) Wellen auf. Das statistische Kriterium, ob die festgestellten Wellen signifikant sind, gibt dabei einen Anhaltspunkt, welche Wellen den größten Beitrag zur Ausprägung einer Zeitreihe beisteuern. Als lange Wellen, die hier besonders interessieren, sollen alle Wellen mit einer Länge größer gleich ca. 10 Jahre verstanden werden, in Anlehnung an die ökonomische Betrachtungsweise, bei der darunter die eindeutigen Konjunkturzyklen beginnen. Alle langen Wellen mit den dazugehörigen statistischen Größen sind in Tabelle I im Anhang (Kap. 7) aufgelistet.

Auf der Grenze zu kurzen Wellen stehen die um die 7 Jahre langen Wellen, die bei allen Zeitreihen der Jahre bis 1850 für die Universität Göttingen mehr Gewicht als die mittleren Wellen haben ²¹. Sie sollen hier nicht weiter untersucht werden, zumal sie alle nicht signifikant sind. Interessant ist jedoch, daß sie vor 1850 die jeweils zweitwichtigste zyklische Komponente dieser frühen Zeitreihen sind.

Das erste Ergebnis der Spektralanalyse ist, daß während des ganzen Zeitraums lange Wellen von 22,5 bis 45 Jahren Länge signifikant vorhanden sind. Die Längenangaben sind dabei nur sehr vorsichtig zu interpretieren. Zyklen zwischen 45 und 22,5 Jahren Länge sind wegen der - im Vergleich zu diesen Zykluslängen - relativen Kürze der Zeitreihen statistisch nicht separierbar, so daß sie besser unter der Aussage "um 30-Jahre lange" Wellen zusammengefaßt werden.

Als zweites ergibt sich, daß in allen Fällen ab 1850 Wellen von 11,3 bis 15 Jahren Länge signifikant auftreten. Diese Wellen sind bis 1850, außer in der evangelischen Theologie an der Universität Tübingen, nicht statistisch signifikant. Das ist der Grund der vorher festgestellten Strukturveränderung in den Zeitreihen: das Auftreten dieser kürzeren Zyklen. Das Verhältnis der Spektralwerte der kürzeren zu den längeren Zyklen ist ein weiterer Beleg für einen Strukturbruch. In allen Fällen erhöht sich das Gewicht ²² der kürzeren Zyklen, von durchschnittlich 1/25 auf ca. 1/2 der langen Zyklen.

Die Veränderung, die die späteren von den früheren Teilen der Zeitreihen unterscheidet, läßt sich also dadurch charakterisieren, daß in die Bewegung der Zeitreihen ein neues Element kommt: Zyklen mittlerer Länge. War das Aussehen der Zeitreihen im früheren Zeitraum hauptsächlich von einer dominierenden langen Welle um die 30 Jahre Länge geprägt, so sind es im späteren Zeitraum zwei Wellenbewegungen, die die Gestalt wesentlich beeinflus-

²¹ Evang. Theologie Wert: 1550; Jura Wert: 695; Medizin Wert: 549; alle Studenten Wert: 44000.

²² Hier nur definiert als der Wert der Spektraldichtefunktion, nicht als Integral. Für den Vergleich reicht diese Definition.

Bild 4: Die Frequenz-Antwort-Funktion des Kerbenfilters

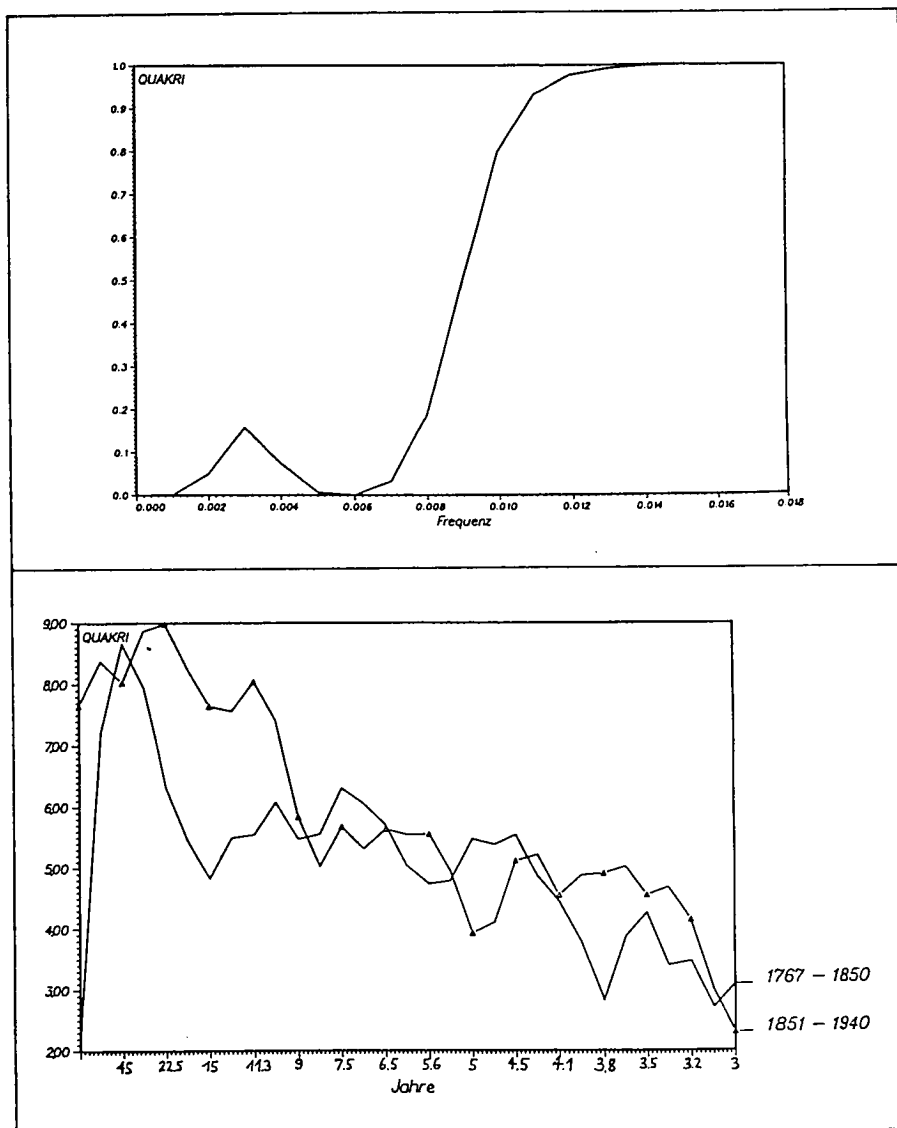


Bild 5: Das Logarithmierte Spektrum der Zeitreihe Medizinstudenten in Göttingen

sen. Die Einsicht, daß sich nach 1850 nicht die Dauer der Wellenbewegungen nach und nach verkürzt, sondern zusätzlich zur vor 1850 vorhandenen Struktur ein neuer Zyklus hinzukommt, ist hierbei wesentliches Ergebnis der Spektralanalysetechnik. Eine kontinuierliche Verkürzung der Zyklendauer hätte nämlich nur eine, entsprechend breite, Spitze im Spektrum zur Folge gehabt.

5. Eigendynamik

Die Frage einer Eigendynamik des Bildungssystems und hier speziell der Entwicklung der Zahl der Studierenden hat weitreichende Konsequenzen für die Beurteilung des historischen und zukünftigen Verlaufs. Insbesondere die Möglichkeiten der politischen Steuerung und Einflußnahme und die Prognosen zukünftiger Entwicklungen sind davon betroffen, ebenfalls die Frage der Abhängigkeit der Universitäten etwa von wirtschaftlichen oder anderen Vorgaben außerhalb des Bildungssystems. Besonders schwerwiegend ist in unserem Zusammenhang, daß hier lange Wellen wirksam sind und deshalb eigendynamische Entwicklungen kurzfristige Einflüsse sozusagen überdauern können.

Das Vorhandensein einer Eigendynamik im tertiären Bildungssystem wird in der Geschichte immer dann zum Thema, wenn die politische Steuerung nicht greift, obwohl sie nach eigenem Selbstverständnis mit genügend Machtfülle ausgestattet ist. Das Beispiel des Numerus clausus des nationalsozialistischen Ministers RUST wurde schon angeführt. Aus neuerer Zeit sei die Expansion der Studentenzahlen ab Mitte der 60er Jahre genannt, die weit über die optimistischsten Erwartungen der damaligen Bildungspolitiker hinausging²³.

Nach der Feststellung von langen Zyklen in den Zeitreihen erhebt sich also die Frage, wieviel der Bewegung der Zeitreihe aus eigendynamischen Entwicklungen von der Länge dieser Zyklen herrührt. Eine Entwicklung soll dabei dann als eigendynamisch verstanden werden, wenn es über ihre ganze zeitliche Ausdehnung ein einheitliches Erzeugungsgesetz gibt, das die jeweils nächsten zukünftigen Werte aus ihrer Vergangenheit erklären kann, ihre Dynamik also auf ihre eigene Gestalt zurückführt.

Für die statistische Analyse ergibt sich damit die Aufgabe, ein Modell,

²³ S. die Prognose des WISSENSCHAFTSRATS von 1964. Meine Untersuchung hier bezieht sich nur auf die Zeit vor dem 2. Weltkrieg. Es ist jedoch zu vermuten, daß in den 60er Jahren ein ähnlicher Strukturbruch stattgefunden hat wie in der Mitte des letzten Jahrhunderts, der u.a. an der sprunghaften Steigerung des Niveaus der - wie auch immer eingeschätzten - Bildungsbeteiligung abzulesen ist.

vereinfacht gesagt eine Gleichung, für jede Zeitreihe zu finden, die a) für jeden beliebigen Zeitpunkt mit möglichst großer Wahrscheinlichkeit den nächsten zukünftigen Wert mit Hilfe einiger zurückliegender Werte berechnen kann und b) selbst keine kurzfristigen Schwankungen aufweist (wie die Zeitreihe selbst), sondern nur an die langen zyklischen Komponenten der Zeitreihe angepaßt ist. Ist ein solches Modell gefunden, so stellt der Anteil der Varianz der Zeitreihe, der sich aus diesem Modell erklären läßt, genau den Beitrag dar, den die langen Wellen zur Bewegung der Zeitreihe liefern.

Unter der Voraussetzung eines stochastischen Prozesses ist das ein sog. Box-Jenkins-Modell²⁴. Ziel der folgenden Analyse ist es deshalb, aus den vorliegenden Zeitreihen Box-Jenkins-Modelle zu schätzen, die die im vorigen Abschnitt gefundenen Zyklen erzeugen. Es handelt sich hier also nicht um eine Box-Jenkins-Analyse mit dem üblichen Ziel, das bestangepaßte Modell herauszufinden, sondern um eine Anpassung von Modellen, die bestimmten Voraussetzungen genügen sollen, zum Zweck deskriptiver Analyse.

Dabei erfolgt die Suche nach einem Modell in zwei Schritten. Zunächst werden die Zeitreihen mit einem Glättungsverfahren von kurzfristigen Schwankungen befreit und der Anteil der nach der Glättung verbliebenen Varianz berechnet. Dann werden Modelle für die geglätteten Zeitreihen geschätzt und der Anteil der Varianz, den sie an den geglätteten Zeitreihen erklären, bestimmt. Aus beiden Anteilen zusammen ergibt sich dann, ein wie hoher Prozentsatz der Varianz der Originalzeitreihe durch die Eigendynamik der langen Wellen erklärt werden kann.

Auf Grund spezifischer Berechnungsschwierigkeiten erfordert die Darstellung langer Wellen in Box-Jenkins-Modellen nämlich eine Transformation der Zeitreihen. Sie müssen vorher von den Einflüssen kürzerer Zyklen bereinigt werden. Zu diesem Zweck werden nacheinander gleitende Durchschnitte der Ordnung 5 und der Ordnung 3 angewandt. Diese Art von Mittelung von jeweils 5 bzw. 3 aufeinanderfolgenden Zeitreihenwerten schaltet offensichtlich kurzfristige Schwankungen aus²⁵.

Der Anteil der Varianz, der den Zeitreihen durch die Glättung entzogen wird, läßt sich an der verminderten Varianz in Tabelle 2, Spalte 2 ablesen. Die Auswirkung dieser Glättung der Zeitreihen (mit einem Fachwort Tiefpaßfilterung) zeigt am Beispiel der Evangelischen Theologiestudenten in

²⁴ S. BOX/JENKINS 1976.

²⁵ Der doppelte gleitende Durchschnitt bildet einen symmetrischen Tiefpaßfilter mit ausreichender Hochfrequenzunterdrückung. S. KOOPMANS, 1979, Kap. 6.

Göttingen Bild 6. Man sieht anschaulich, daß die gefilterte Zeitreihe keine kurzfristigen Schwankungen mehr aufweist.

Im zweiten Schritt wurden für alle Zeitreihen jeweils drei verschiedene Modelle geschätzt, die sich dadurch unterscheiden, daß sie die Zeitreihe als Summe von höchstens ein, zwei oder drei Zyklen darstellen können. Diese Unterscheidung war nötig, weil die Schätzmethode für die optimale Anpassung von Modellen nicht automatisch zu Modellen mit zwei langen Wellen führt; je nach Vorgaben für die Modelle, die sich auf die Art und Anzahl der Koeffizienten beziehen, können sie zu einer bestimmten Zahl zyklischer Komponenten führen, die mehr oder weniger ausgeprägt sind. Für unsere Analyse kommt es jedoch darauf an, Modelle zu finden, die lange Wellen darstellen, d.h., daß sie statistisch signifikante Koeffizienten besitzen, die Zyklen von langer Dauer erzeugen. Durch die Entscheidung für diese drei Modelltypen wurde so auch ein Vergleich mit den im vorigen Kapitel festgestellten zyklischen Anteilen der Zeitreihen angestrebt: Führt die optimale Anpassung zu Modellen, deren zyklische Komponenten denen der Spektralanalyse entsprechen und die signifikant sind, ist dies eine weitere Bestätigung dafür, daß die Zeitreihen sich gerade aus diesen Zyklen erklären lassen.

Die berechneten ²⁶ Box-Jenkins-Modelle sind autoregressiv (AR-Modelle), d.h. sie erklären innerhalb gewisser wahrscheinlicher Grenzen den zukünftigen Wert der Zeitreihen ausschließlich als Summe gewichteter vergangener Werte, und zwar hier durch die bis zu 2, 4 oder 6 Jahre vorhergehenden Werte (AR(2)-, AR(4)-, AR(6) Modelle). Dabei kann das AR(2)-Modell höchstens einen, AR(4) zwei und AR(6) drei Zyklen erzeugen ²⁷. Die Modelle und weitere Parameter sind in Tabelle 3 im Anhang aufgeführt.

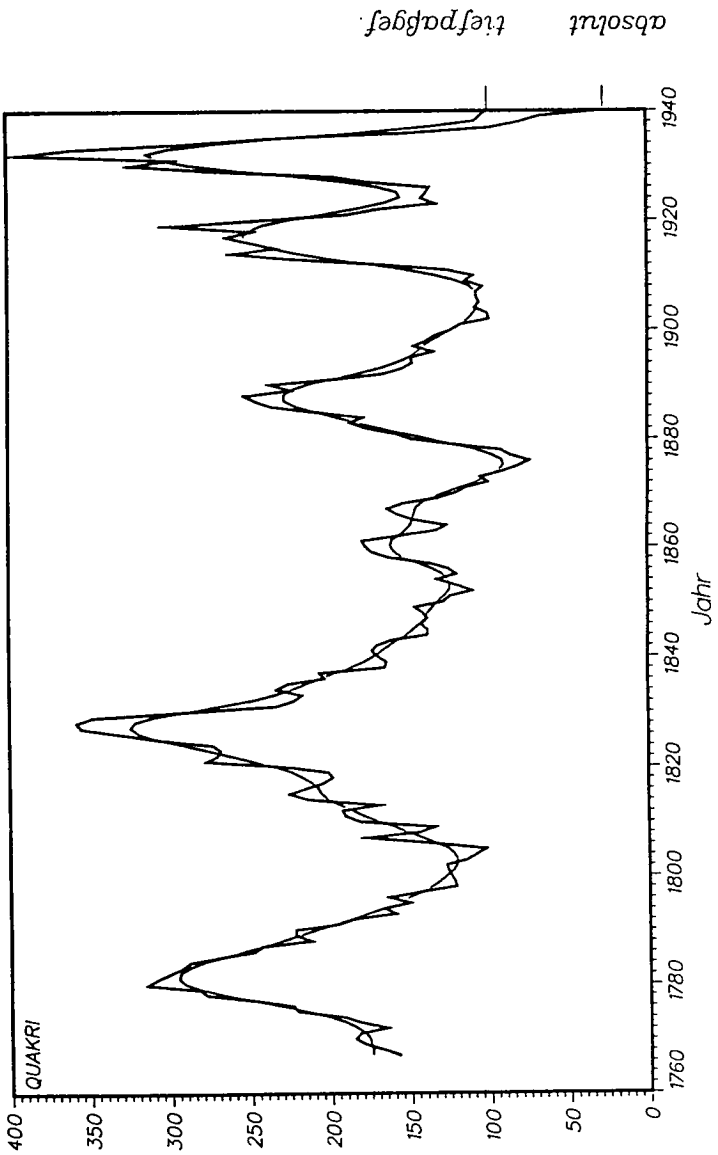
Einer allzu ausgeweiteten "Modellphilosophie", die die Beschäftigung mit den möglichen Modellen einer Untersuchung der historischen Verhältnisse vorzieht, soll mit dieser Art der Verwendung dieser Modelle vorgebeugt werden. Sie werden hier ausschließlich zur quantifizierenden Deskription der empirischen Zeitreihen benutzt. Die Bauart der Modelle, die bei der Erklä-

²⁶ Geschätzt wurde mit dem Programm P2T des Pakets BMDP.

²⁷ Die Berechnung der Autokorrelations- (ACF) und der partiellen Autokorrelationsfunktion (PACF) der Zeitreihen zeigt, daß für alle Zeitreihen ein rein autoregressives Modell in Frage kommt. Ein solches Modell weist sogenannte Pseudo- oder stochastische Zyklen auf, wenn Wurzeln seines charakteristischen Polynoms komplex sind.

Bild 6 (rechts): Die Studierenden der Evangelischen Theologie an der Universität Göttingen von 1767-1940: Die Zeitreihe in absoluten Zahlen und nach Bildung der gleitenden Durchschnitte von 5 und 3 Jahren.

Evang. Theologie Göttingen



rung der Zeitreihenwerte durch einen notwendigen Anteil an Zufall gekennzeichnet sind und deren Prozedur zum Herausfinden der Koeffizienten sich lediglich an statistischer Optimalität orientiert, läßt es nicht zu, sie als Verhaltensgleichungen für Aggregationen von Subjekten aufzufassen, wie dies beispielsweise bei den grundlegenden Modellen in der Makroökonomie oft praktiziert wird.

Die meisten der schon im Spektrum sichtbaren Zyklen kann man in den Modellen wiederfinden. Grundsätzlich wird in allen Modellen mit bis zu einem Zyklus der um die 30 Jahre lange Zyklus erkannt. Falls die 7-jährigen Zyklen, die vor 1850 die zweitwichtigste Komponente darstellen, in den Modellen mit 2 und 3 möglichen Zyklen vor 1850 auftreten, bestätigen die Modelle durch die Anzahl signifikanter Koeffizienten den Eindruck, den das Spektrum vom Verhältnis der um die 30- und 7-jährigen Zyklen vor 1850 vermittelte: für Jura und Medizin Göttingen waren sie schwächer, für Evang. Theologie und alle Studierenden Göttingen stärker vertreten. Für die Evang. Theologie Tübingen ist der 15-jährige Zyklus vor 1850 mit 5 signifikanten Koeffizienten im Modell mit bis zu 3 Zyklen gut repräsentiert.

Nach 1850 ergibt das größte Modell für die Evang. Theologie, Jura und Medizin Göttingen sowohl die langen als auch die neuen Zyklen von 11-15 Jahren Dauer mit signif. Koeffizienten. Sowohl bei der Evang. Theologie Tübingen als auch bei allen Studierenden Göttingen ergeben die Modelle mit 2 und 3 möglichen Zyklen immerhin den mittleren Zyklus. Die Modelle für die Zeit ab 1850 bestätigen somit das Auftreten von Zyklen mittlerer Länge. Insgesamt gesehen erzeugen diese Modelle fast alle stochastischen Zyklen, die schon in der Spektralanalyse als für diese Zeitreihen charakteristische Zyklen gefunden wurden.

Der sehr hohe Prozentsatz aufgeklärter Varianz (Tabelle 3, Spalten 3 und 4) zeigt die Güte der Modelle, sie lassen unter 1% ungeklärt. Bis auf diesen Rest stammt der nicht aufgeklärte Teil der Varianz der Originalreihen nur noch aus den kürzeren Schwingungen, die durch die Glättung im Schritt 1 herausgefiltert wurden, und bei den wachsenden Zeitreihen aus dem Trend, also Aufschwüngen von größerer Länge als die Zeitreihen selbst.

Die Gestalt der Zeitreihen - abgesehen vom Trend - läßt sich nach dieser Analyse zwischen knapp 57% (bei Evang. Theologie Tübingen 1851-1940) bis zu 87% (bei Evang. Theologie Göttingen 1767-1850) aus Eigendynamik erklären, die vor 1850 aus einem und nach 1850 aus 2 langen Zyklen stammt. Innerhalb dieses Modellrahmens können also die Zeitreihen zu einem beachtlichen Teil - im Durchschnitt etwa $2/3$ - aus einer ihnen selbst innewohnenden Struktur erklärt werden, ohne externe Einflüsse. Dabei hängt diese Struktur

zentral mit 2 Zyklen zusammen, von denen der zweite erst nach 1850 auftritt. Welche Bedeutung diese Zyklen haben könnten, soll im folgenden geklärt werden.

6. Die historisch-soziologische Bedeutung der statistischen Analyse

Von der Annahme ausgehend, daß Zeitreihen stochastische Prozesse repräsentieren, hatte die statistische Untersuchung zunächst das Ergebnis einer sich in der Zeit verändernden Struktur des Prozesses. Die Untersuchung legt nahe, daß ein entscheidender Strukturbruch nach der Mitte des letzten Jahrhunderts stattfand (Auftreten von kürzeren Zyklen). Gibt uns eine Analyse des Bildungsverhaltens, spez. des Universitätsbesuchs, Anhaltspunkte für eine Bestätigung eines solchen Bruches?

Wie eine neuere Untersuchung zeigt, war die Zeit von ca. 1860-1875 eine besondere Phase in der Universitäts- und Bildungsgeschichte²⁸. In dieser Zeit wurde das höhere Bildungssystem gleichsam auf ein höheres Entwicklungsniveau gehoben. Als beispielhafte Indikatoren weisen sowohl der Anteil der Lehrerstudenten an der Gesamtstudentenzahl als auch der Anteil der in der Philosophischen Fakultät tätigen Professoren (und damit der hauptsächlich Oberlehrer ausbildenden Professoren) die höchsten Wachstumsraten im gesamten betrachteten Zeitraum auf. Ein großer Teil der "Produktion" der Universitäten wurde also sofort wieder im Bildungssystem beschäftigt, es fand in dieser Zeit ein gewaltiger Eigenausbau statt. Vorausgegangen war in einer Phase der politischen Restauration eine ein Vierteljahrhundert währende Periode der Stagnation und Schrumpfung der Bildungsbeteiligung, sowohl an den Universitäten als auch an den Gymnasien. Das führte ab den 60er Jahren zu einem allgemeinen, d.h. alle Karrieren betreffenden, Nachwuchsdefizit, das auch so von den Zeitgenossen wahrgenommen wurde²⁹. In der anschließenden Phase des Eigenausbaus wurde die Fähigkeit des Bildungssystems insgesamt, höher qualifiziert Gebildete zu produzieren, gleichsam in einem Entwicklungssprung vergrößert. Danach setzte in fast allen Fakultäten ein Wachstum der Studentenzahlen ein (s. z.B. Bild 1 alle Studierenden, ebenso Jura, Medizin; Evang. Theologie ist aus spez. Gründen die Ausnahme). Ein Strukturbruch nach der Mitte des letzten Jahrhunderts ist demnach auch in der sozialhistorischen Analyse festzustellen.

Das zweite Ergebnis der statistischen Analyse ist die zyklische Struktur der

²⁸ S. TITZE et al., Kap. 2.

²⁹ S. z.B. HUCKERT 1895, S.621ff.

Zeitreihen. Sowohl als Ergebnis der Spektralanalyse, aber auch in den Modellen wiederzufinden, konnten die bedeutendsten zyklischen Anteile der Zeitreihen als lange Wellen von um die 30 Jahre und 11-15 Jahre Länge festgestellt werden, wobei die kürzeren Frequenzen jeweils erst ab Mitte des letzten Jahrhunderts auftraten. Alle noch kürzeren Frequenzen insgesamt haben durchschnittlich nur noch einen Anteil von einem Drittel an der Ausprägung der Zeitreihe.

Eine mögliche Hypothese zur Zyklizität dieser Zeitreihen, daß sie etwas mit den Konjunkturzyklen der Wirtschaft zu tun hätten, muß aufgrund dieser Ergebnisse sehr in Frage gestellt werden. Bei den für das 19. Jahrhundert von SPREE untersuchten ökonomischen Zeitreihen traten Konjunkturzyklen (5-10 Jahre) bei allen Reihen auf, die "Kuznetszyklen" (20-25 Jahre) nur bei der Hälfte aller Reihen, und die noch längeren "Kondratieffzyklen" (40-60 Jahre), die singulären Ereignissen wie Innovationsschüben zugeschrieben werden, gar nicht ³⁰. Wegen der unterschiedlichen Länge der Zyklen sind die Möglichkeiten einer Wechselwirkung gering. Eine Verbindung ergibt sich jedoch dadurch, daß der Zeitraum von 1850-1873 auch ökonomisch als eine Zeit überwiegend günstiger Wachstumsbedingungen zu charakterisieren ist ("Takeoff"-Phase).

Die beiden Frequenzen sind jedoch als strukturbestimmende Parameter des höheren Bildungssystems bekannt. Der längere Zyklus um die 30 Jahre ist die Länge einer durchschnittlichen Berufsdauer, d.h. die Zeit vom Eintritt ins Berufsleben, etwa auf eine Pastorenstelle oder Richterstelle, bis zum Tod oder zur Pensionierung bzw. Emeritierung. Die Altersstruktur der im Beruf stehenden Pastoren bzw. Juristen ist aber i.a. nicht gleichmäßig verteilt (auf Grund von unregelmäßigen Wachstumsschüben in der Karriere), so daß der Abgang durch Pensionierung und Tod immer eine zyklische Komponente von der Länge einer durchschnittlichen Berufsdauer aufweist. Diese langen Wellen wirken einmal als Hauptbestimmungsgröße der Ersatznachfrage direkt auf die Studentenzahlen ein und beeinflussen sie außerdem über die Tatsache der Berufsvererbung indirekt.

Weil die Altersstruktur der akademischen Berufe einerseits überhaupt nicht kurzfristig änderbar ist - auf Grund des Beamtenrechts, des Berechtigungswesens und der Tatsache, daß ein Berufswechsel um so unwahrscheinlicher wird, je länger die Ausbildung dauert -, sie aber andererseits zu jedem Zeitpunkt für etwa 30 Jahre - eine Berufsdauer - in die Zukunft den wichtigsten Teil der Nachfrage fortschreibt, ist einleuchtend, daß sie die beharrlich-

³⁰ S. SPREE 1978.

ste Komponente im Rekrutierungssystem der akademischen Karrieren darstellt. Als stärkste Komponente der Zeitreihen wiesen sie auch alle Spektren und Modelle aus. Gleichzeitig wird sie wegen ihrer Langfristigkeit bei der aktuellen politischen Diskussion oft unberücksichtigt bleiben. So ist z.B. die heutige Altersstruktur der Lehrer in der BRD so ungleichgewichtig wie niemals zuvor in der deutschen Geschichte, mit den entsprechenden zu erwartenden Konsequenzen (fast 4/5 aller Lehrer an Gymnasien waren 1978 jünger als 45 Jahre).

Der kürzere Zyklus um die 11 - 15 Jahre ist etwa die doppelte Länge einer Ausbildungsdauer, also vom ersten Semester bis zum abgeschlossenen Vikariat bzw. Referendariat. Aus einer einfachen Modellkonstruktion eines sog. Akzelerator-Multiplikatormodells mit den 2 Hypothesen a) konstante Nachfrage nach Absolventen und b) Orientierung der Studienanfänger an der aktuellen Situation der Absolventen (bei Überschuß an Absolventen über die Nachfrage weniger Studienanfänger, bei Mangel mehr als vorher) ergibt sich schon dieser Zyklus einer genau doppelten Ausbildungsdauer bei den Gesamtstudentenzahlen³¹. Für die Annahme b) gibt es eine Reihe von Indizien, die bestätigen, daß die Erstsemester sich, viel mehr als an allen Arten von Warnungen und Werbungen von staatlichen Stellen oder Berufsverbänden, an der aktuellen Arbeitsmarktsituation orientieren. Auch unter den erweiterten realen Bedingungen einer sich verändernden Arbeitsmarktlage kann also eine solche zyklische Komponente einer doppelten Ausbildungsdauer erwartet werden. Sie wird sich dann mit der Nachfragebewegung, die u.a. von der Altersstruktur bestimmt wird, vermischen³². Das erklärt jedoch nur genauer, daß sich die Studenten unter dem Zwang der Verhältnisse an den Verwertungsbedingungen ihrer Ausbildung orientieren. Dahinter steht ein starres Berechtigungssystem, das die sozialen Chancen an eine festgelegte Bildungsleiter knüpft, die sich über Gymnasium, Abitur, Fachstudium und evtl. Promotion immer stärker an eine bestimmte Karriere bindet³³.

Das Ergebnis, daß die kürzeren Zyklen erst ab der Mitte des letzten Jahrhunderts auftreten, bleibt noch mit anderen Theorien zu verbinden. Der schon oben erwähnte Strukturbruch im höheren Bildungssystem zu diesem

³¹ Diese Länge ergibt sich daraus, daß sich die Anbieter (Studenten) an der aktuellen Situation orientieren, ihre Produkte (die Absolventen) aber erst eine Ausbildungsdauer später auf dem Markt für die Nachfrager erscheinen. Diese Art von Modell ist auch als "Schweinezyklus" bekannt.

³² Dazu und zu den weiteren Komponenten einer Modellstruktur vgl. TITZE et al., Kap. 3.

³³ S. SCHELSKY 1957, HURRELMANN 1975, bes. Kap. 2.3.

Zeitpunkt hängt selbstverständlich auch mit der Industrialisierung und dem steigenden Bedarf des sich entwickelnden Deutschen Reichs an hochqualifizierten Arbeitskräften in Industrie, Verwaltung usw. zusammen. Die Industrialisierung ist jedoch nur auf dem Hintergrund der Umgestaltung des gesamten Arbeitsbereichs von mehr feudalistischen zu mehr kapitalistischen Zusammenhängen und von eher ständisch gebundenen zu wirklich offenen Arbeitsmärkten zu verstehen. Die sichere Erwartung, mit einem akademischen Abschluß auch eine Stelle zu bekommen, und sei es, daß man, z.B. als Absolvent der Evang. Theologie zunächst nur als Vikar beschäftigt, solange warten mußte, bis der bisherige Stelleninhaber des zukünftigen, aber schon festgelegten Kirchspiels das Zeitliche segnete, wurde auf den offener und freier werdenden Arbeitsmärkten mehr und mehr in Frage gestellt. So hat auch sicher auf die Frage: "Studiere ich überhaupt? Und wenn ja, was?" die Tatsache, ob man mit dem abgeschlossenen Studium überhaupt eine Arbeit bekommt, erst ab diesem Zeitpunkt immer mehr entscheidenden Einfluß gewinnen können. Und erst unter der Voraussetzung, daß die Verwertbarkeit des akademischen Abschlusses auf einem freien Arbeitsmarkt eine Entscheidungsgröße für die Wahl des Studienanfängers darstellt, kann man von der obigen Verhaltenshypothese ausgehen und mit Recht Zyklen einer doppelten Ausbildungsdauer erwarten.

Für die starke Eigendynamik der Studentenströme sind eingangs schon zwei Beispiele erwähnt worden. Auch Statistikexperten der Berufsstände haben sich in der Stärke dieser Bewegung einst verschätzt³⁴. Besonders Politik und Verwaltung überschätzten die Möglichkeiten der Steuerung der Studentenströme immer wieder und wirkten mit ihren Eingriffen z.T sogar zyklisch, weil sie die Länge der Zyklen nicht im Blick hatten. Das statistische Ergebnis des hohen Prozentsatzes, den die langen Zyklen zur Dynamik der Zeitreihen beitragen, bestätigt diese historischen Analysen. Dieses Ergebnis ist allerdings im vorausgesetzten Rahmen univariater Box-Jenkins-Modelle zu sehen. Weiterführende Modelle, die z.B. die Altersstruktur und die Nachfrage als Variablen einbeziehen, könnten Faktoren aufdecken, die für die Eigendynamik bestimmend sind. Letztlich ist jedoch ohne Frage, daß die Vergangenheit der Studentenströme zu einem großen Teil ihre Gestalt auch in mittel- und langfristiger Zukunft beeinflußt. Diese autokorrelierte Struktur müßte auch jedes weiterführende Modell hervorbringen, in dem die Studentenströme nicht als eigenständige, sondern als resultierende Größen erscheinen.

³⁴ S. z.B. die Schrift "Abmahnung vorm Studium der Medizin", die vom 31. Deutschen Ärztetag 1903 an viele Schulleiter verschickt wurde.

Mit der Verbindung der statistischen Ergebnisse mit entsprechenden Analysen historisch-soziologischer Herkunft ist zuletzt auch die Frage nach dem Sinn der Anwendung der Zeitreihenanalysemethoden ein Stück weit beantwortet. Die empirisch-statistische Grundlage für die Thesen der historischen Forschung konnte dadurch erweitert werden. Aber auch die beste Anpassung eines statistischen Modells an eine historische Zeitreihe kann im Sinn der Geschichtswissenschaften nie Beweis dafür sein, daß der die Zeitreihe hervorbringenden geschichtlichen Entwicklung ein einheitliches Bildungsgesetz zugrundeliegt. Welche Entwicklungen in der Geschichte, auch wenn sie mehrmals erschienen, zufällig und singulär waren und welche dagegen systematisch und immer wiederkehrend sind, kann nur die historische Forschung entscheiden. Daß diese Methoden aber in der historischen Forschung mit Erfolg angewendet werden können, könnte darauf hindeuten, daß im Bildungssystem Strukturen und funktionale Beziehungen bestehen, die unabhängig von dem, was die handelnden Subjekte erstreben, auch "Geschichte machen".

7. Tabellenanhang

Tabelle 1: Alle langen Wellen der Spektren der beiden Teile aller Zeitreihen stehen in Spalte 1, wobei die Jahreslänge der entsprechenden Zyklen angegeben wird. In Spalte 2 steht der dazugehörige geschätzte Spektralwert, in Spalte 3 dann der dazugehörige untere Rand des 90%-Vertrauensbereichs (VB) dieses Spektralwerts und in Spalte 4 die Varianz der Reihe. Die Varianz ist der Wert, den das Spektrum theoretisch hätte, also wenn alle Werte der Reihe zufällig wären. Liegt dieser Wert unterhalb des 90%-VB, so ist der Peak auf dem 5%-Niveau signifikant³⁵, was in Spalte 5 aufgeführt wird. Für Spalte 6 wurde das Verhältnis der Spektralwerte der kürzeren zu den längeren Zyklen berechnet.

Hinweis:

Dieser Beitrag wurde erstmals in der Zeitschrift "Hochschulausbildung", 2/88, veröffentlicht. Die Herausgeber danken dem Leuchtturm-Verlag für die freundliche Erlaubnis zum Nachdruck.

³⁵ S. GOTTMAN, S.223. Hier reicht die einseitige Abschätzung.

Tabelle 1: Signifikanz der Zyklen

Fach/Uni Zeit- raum	Spektrum Peaks bei. Jhr.	Power	untere Grenze 90%-VB	Vari- anz	sig. nifi- kant?	Verhältn. krz./lng. Zyklen
Spalte	1	2	3	4	5	6
Ev.Theol.Tüb.						
1760-1850	45	5100	2920	353	ja	1/6
	15	890	423		ja	
1851-1940	45	9500	9520	1223	ja	1/1 > 9
	15	9900	2330		ja	
Ev.Theol.Gö.						
1767-1850	42	89900	92600	3960	ja	1/195
	9,3	616			nein	
1851-1940	30	29700	19100	9620	ja	1/0,9
	15	33200	15800		ja	
Jura Gö.						
1767-1850	42	10400	9960	816	ja	1/19
	12	558		nein		
1851-1940	22,5	12300	5850	1930	ja	1/2
	11,3	6050	2880		ja	
Medizin Gö.						
1767-1850	42	5900	2180	380	ja	1/13
	9,3	439	209		nein	
1851-1940	22,5	8010	3810	910	ja	1/2,6
	11,3	3120	1480		ja	
Alle Fächer Gö.						
1767-1850	42	898000	92200	51000	ja	1/56
	12	15900			nein	
1891-1990	45	892000	900000	58000	ja	1/3
	11,3	233000	111000		ja	

Tabelle 2: In Spalte 1 sind die Arten der Transformation der Zeitreihen aufgeführt. Dabei bedeutet Trendeliminierung Logarithmierung und Kerbenfilterung, Glättung zweifache Anwendung eines gleitenden Durchschnitts von 3 bzw. 5 Jahren. In Spalte 2 steht jeweils die Varianz nach der Transformation. In Spalte 3 ist die Varianz nach der zusätzlichen Glättung in % der Varianz nach der Trendeliminierung (Logarithmierung und Kerbenfilterung) angegeben.

Tabelle 2: Varianzverminderung durch Glättung

Fach/Uni Zeitraum	Art der Transformation	abs.	Varianz in %
Spalte	1	2	3
Evang. Theologie Tübingen			
Würzburg			
1761-1850	Trendeliminierung	353	
	" + Glättung	264	74,8
1851-1940	Trendeliminierung	1223	
	" + Glättung	698	57,1
Evang. Theologie Göttingen			
1767-1850	absolut	3963	
	Glättung	3481	87,0
1851-1940	absolut	4624	
	Glättung	3364	72,0
Jura Göttingen			
1767-1850	Trendeliminierung	816	
	" + Glättung	576	73,4
1851-1940	Trendeliminierung	1444	
	" + Glättung	1156	80,1
Medizin Göttingen			
1767-1850	Trendeliminierung	380	
	" + Glättung	246	64,7
1851-1940	Trendeliminierung	910	
	" + Glättung	625	68,7

Alle Studenten Göttingen

1767-1850	Trendeliminierung	51076	
	" + Glättung	40750	79,8
1851-1940	Trendeliminierung	58081	
	" + Glättung	48534	83,6

Tabelle 3: Die durch die Modelle nicht erklärten Varianzen ("Varianz der Residuen") sind absolut in Spalte 2, der Prozentsatz an durch die Modelle aufgeklärter Varianz der geglätteten Zeitreihen in Spalte 3, die Anzahl signifikanter Koeffizienten ³⁶ in Spalte 4 und die sich auf Grund der Koeffizienten ergebenden stochastischen Zyklen in Spalte 5 aufgeführt.

Tabelle 3: Box-Jenkins-Modelle

Fach/Uni Zeitraum	Art des Modells	Varianz der Residuen		Anzahl signif. Koeff.	Länge stoch. Zyklen
		abs.	in %		
Spalte	1	2	3	4	5

Evang. Theologie Tübingen

1761-1850	AR(2)	2,7	99,0	2	30,3
	AR(4)	2,0	99,2	3	14,8
	AR(6)	1,6	99,3	5	40,0;15,2
1851-1940	AR(2)	6,0	99,1	2	21,7
	AR(4)	4,2	99,3	3	15,4
	AR(6)	3,8	99,4	4	16,9;16,1

Evang. Theologie Göttingen

1767-1850	AR(2)	9,6	99,7	2	35,7
	AR(4)	5,1	99,9	4	37,0;10,2
1851-1990	AR(2)	22	99,3	2	21,2
	AR(4)	6,3	99,8	3	15,6
	AR(6)	5,3	99,8	4	29,4;14,7

³⁶ Signifikanzgrenze für einen Koeffizienten war ein T-Wert größer als 1,96.

Jura Göttingen

1767-1850	AR(2)	3,5	99,3	2	30,3
	AR(4)	2,6	99,5	3	29,4;9,8
	AR(6)	1,8	99,7	3	32,3;7,6
1851-1990	AR(2)	7,6	99,3	2	28,6
	AR(4)	3,0	99,7	3	14,5
	AR(6)	2,3	99,8	4	41,7;12,0;10,4

Medizin Göttingen

1767-1850	AR(2)	2,9	98,8	2	28,6
	AR(4)	2,1	99,9	3	35,7;43,4
	AR(6)	1,8	99,3	2	33,3;11,1;7,6
1851-1990	AR(2)	6,2	99,0	2	21,7
	AR(4)	6,3	99,8	2	15,2
	AR(6)	5,3	99,8	3	24,3;11,0

Alle Studenten Göttingen

1767-1850	AR(2)	149	99,6	2	34,2
	AR(4)	85	99,8	4	35,2;6,9
	AR(6)	72	99,8	4	35,2;6,6
1851-1990	AR(2)	404	99,2	2	33,3
	AR(4)	75	99,8	4	10,4
	AR(6)	40	99,9	4	11,1

Summary

During the last two centuries phases of a so called oversupply have occurred several times at German universities. The recurrence of these phases is shown in the undulations of the time series of students of different subjects. Five particularly long ones of these series are examined by means of the statistical methods of time series analysis, the result being that the purely statistical results point out long-term significant conditions of the higher educational system: the influence of the age structure of the people in professions, the length of the time of education (at the university) and the self-dynamic persistence of the structures as well as a major structural change in the middle of the last century.

Literatur

- BOX, J.; JENKINS, G.: Time Series Analysis. Forecasting and Control. San Francisco, 1976.
- GOTTMAN, J.M.: Time Series Analysis. A comprehensive introduction for social scientists. Cambridge, 1981.
- HUCKERT, E.: Darf man jetzt schon zum Studium der Philologie aufordern? In: Central-Organ für die Interessen des Realschulwesens 23 (1895).
- HURRELMANN, K.: Erziehungssystem und Gesellschaft. Hamburg, 1975.
- KOOPMANS, W.: The Spectral Analysis of Time Series. New York, 1979.
- LEXIS, W.: Bedarf und Angebot in den gelehrten Berufszweigen. In: Hochschul-Nachrichten 15 (1905), Heft 169/170.
- MCCLEARY, R.; HAY, R.A.: Applied Time Series Analysis for the social sciences. London, 1980.
- MÜLLER-BENEDICT, V.: Akademikerprognosen und die Dynamik des Hochschulsystems. Frankfurt, 1991.
- NATH, A.: Die Studienratskarriere im Dritten Reich. Frankfurt, 1988.
- SHELTSKY, H.: Schule und Erziehung in der industriellen Gesellschaft. Würzburg, 1957.
- SCHULTE, H.: Statistisch-methodische Untersuchungen zum Problem langer Wellen. Hanstein, 1981.
- SPREE, R.: Wachstumstrends und Konjunkturzyklen in der deutschen Wirtschaft von 1820 bis 1913. Göttingen, 1978.
- TITZE, H.: Datenhandbuch zur Deutschen Bildungsgeschichte. Teil 1. Band 1: Die Hochschulen in Preußen und Deutschland. Göttingen, 1987.
- TITZE, H.; NATH, A.; MÜLLER-BENEDICT, V.: Der Lehrerzyklus. Zur Wiederkehr von Überfüllung und Mangel im höheren Lehramt in Preussen. In: Zeitschrift für Pädagogik (1985) 1, S.97-126.
- TITZE, H.; LÜHRS, W.; MÜLLER-BENEDICT, V.; NATH, A.: Prüfungsauslese und Berufszugang der Akademiker 1860-1940. In: Lösche, P. (Hrsg.): Göttinger Sozialwissenschaften heute. Fragestellungen, Methoden, Inhalte. Göttingen 1990, S.181-251.
- WISSENSCHAFTSRAT: Abiturienten und Studenten. Entwicklung und Vorschätzung der Zahlen 1950 bis 1980. Tübingen, 1964.

Verläuft der langfristige Wertewandel zyklisch? Eine zeitreihenanalytische Überprüfung der Thesen von Namenwirth und Weber

von Rainer Metz und Helmut Thome

Seit Ende der sechziger Jahre haben J. Zvi Namenwirth und Robert Ph. Weber in einer Serie von Aufsätzen und Konferenzbeiträgen Daten und theoretische Erklärungsversuche zum langfristigen kulturellen Wandel zur Diskussion gestellt. Ihre 1987 vorgelegte Aufsatzsammlung "Dynamics of Culture"¹ kann als die Summe ihrer bisherigen Überlegungen gelten. Im Zentrum ihrer Arbeiten steht die These, der kulturelle Wandel (primär erfaßt als Wertewandel oder politischer Themenwechsel) verlaufe in zyklischen Wellen, einem regelmäßigen Auf und Ab von Prioritäten, mit denen eine Gesellschaft bestimmte Problemkomplexe bearbeitet.

Eine kritische Würdigung ihrer These könnte an verschiedenen Punkten ansetzen: den theoretischen Konzepten, den Datenquellen, den Methoden der Datenerhebung und Datenanalyse sowie der konsistenten oder weniger konsistenten Verknüpfung dieser Komponenten. Wir konzentrieren uns hier auf die Methodik der Datenanalyse, die "statistischen" Verfahren zur Identifikation zyklischer Verlaufsformen in Zeitreihen. Im ersten Abschnitt skizzieren wir die von Namenwirth und Weber gewählte Vorgehensweise und erläutern einige Probleme, die mit ihr verbunden sind. Im zweiten Abschnitt stellen wir ein alternatives Verfahren zur Identifikation von Zyklen vor, das auf der Theorie linearer Filter beruht. Im dritten Abschnitt wenden wir dieses Verfahren auf Daten an, die von Weber ² auf andere Weise analysiert worden sind. Der letzte Abschnitt enthält ein kurzes Resümee.

1. Die Vorgehensweise von Namenwirth und Weber

Namenwirth und Weber stützen ihre Thesen über den zyklischen Charakter des Wertewandels bisher im wesentlichen auf die Analyse zweier Da-

¹ Siehe J. Zvi Namenwirth und Robert Philip Weber, *Dynamics of Culture* (Boston u. a. 1987).

² Neben dem in Fußnote 1 genannten Werk, insbesondere dessen Kap. 4, beziehen wir uns auf R. Ph. Weber, *The Dynamics of Value Change, Transformations and Cycles: British Speeches from the Throne, 1689-1972*. Dissertation University Connecticut (1978).

tensätze. Namenwirth untersuchte die Parteiprogramme (party platforms) der Republikaner und der Demokraten für die 31 US-amerikanischen Präsidentenwahlen von 1844 bis 1964³. Weber analysierte in ähnlicher Weise 283 britische Thronreden der Zeit von 1689 bis 1972. Beide Forscher gewinnen ihre Daten aus einer computergestützten Inhaltsanalyse mit dem Programmsystem "General Inquirer"⁴: Die Wörter der einzelnen Dokumente werden automatisch ca. 70 Kategorien zugeordnet, die durch den "Lasswell Value Dictionary" definiert sind. Er enthält eine Reihe zentraler Wertkategorien (wie wealth, power, respect, rectitude), die so aufgeschlüsselt sind, daß neben den substantiellen Inhalten auch Aspekte der sozialen Allokation erfaßt werden können⁵. Für jede Kategorie wird die prozentuale Häufigkeit ermittelt, mit der die ihr zugeordneten Wörter bzw. Wortbedeutungen in jedem der zeitlich geordneten Dokumente vorkommen. Für jede Kategorie entsteht somit eine Zeitreihe, die das relative Gewicht eines bestimmten Wertindikators zum jeweiligen Zeitpunkt widerspiegelt. In der Interpretation werden "kovariierende" Kategorien zu theoretisch gedeuteten "Themen" zusammengefaßt (siehe unten).

In etwa der Hälfte bis zwei Drittel der Zeitreihen (je nach Land und Epoche) wollen die Autoren die bereits erwähnten zyklischen Verlaufsformen festgestellt haben. (Die anderen Reihen enthalten zu wenig Varianz oder führen zu keinen konsistenten Ergebnissen in der Datenanalyse, siehe unten.) In den amerikanischen und den britischen Reihen der "kapitalistischen Periode" (ab 1796) wurden lange Zyklen ("L-cycle") mit einem Durchschnitt von etwa 150 Jahren identifiziert, die von kürzeren Zyklen ("K-cycle") mit einer mittleren Periode von ca. 50 Jahren überlagert werden. Für die "merkantilistische" Epoche Großbritanniens (1689-1795) umfassen die entsprechenden Perioden 72 und 32 Jahre. Das langfristige Auf und Ab der Themen wird in Anlehnung an Parsons' (und Bales') berühmtes AGIL-Schema interpretiert: Soziale Systeme müssen vier funktionale Problembereiche "bearbeiten", denen sie sich bei knappen Ressourcen mit zyklisch (!) wechselnder Dringlichkeit zuwenden. Sie werden, teilweise abweichend von Parsons' Terminologie, als "adaptive", "instrumental" (goal attainment), "integretive" und "expressive" (Latency) bezeichnet⁶. So erläutert Weber z. B. den Phasenverlauf in den britischen Reihen ab 1795 wie folgt:

³ Siehe Kapitel 3 und 6 in Namenwirth u. Weber (s. Fußn. 1).

⁴ Siehe in diesem Band den Beitrag von Züll und Mohler.

⁵ Zur näheren Beschreibung siehe Namenwirth u. Weber (Fußnote 1), Kap. 2.

⁶ Siehe Namenwirth u. Weber (s. Fußn. 1), S. 73 - 76, 89 f.

"The war with France in the 1790s and the Second World War correspond to the Adaptive problem. Concern with the economic well-being of the nation is the hallmark of the Instrumental problem. The Integrative problem addresses the relationship among social classes and includes the extension of the franchise to the urban working class. The Expressive problem corresponds to the Irish Home Rule debate."⁷

Die Problemkomplexe entsprechen den "Themen", zu denen die Kategorien zusammengefaßt sind. Die Zuordnung der einzelnen Wertkategorien zu den Themen, also die Zuordnung der Indikatoren zu den theoretischen Konstrukten, ist nicht stabil, sie variiert über die einzelnen Epochen und zwischen den Nationen⁸. Unseres Erachtens sind die wechselnden Zuordnungen theoretisch nicht ausreichend kontrolliert, sondern ziemlich ad hoc vorgenommen worden. Auch die innerhalb einer Epoche getroffenen Zuordnungen werfen Fragen auf. Warum z. B. ist für die kapitalistische Periode "Rectitude Religious" dem "Instrumental Theme", "Rectitude Ethics" dem "Expressive Theme" und "Rectitude Total" dem "Integrative Theme" zugeschlagen?⁹

Die vermeintlich zyklischen Schwankungen in der Priorität der Themen werden vor allem auf ausgeprägte Verschiebungen in der Wählerschaft ("electoral realignment") zurückgeführt, hinter denen als kausal treibende Kraft die Abfolge der Generationen bzw. "population replacement" und Mitgliederaustausch in entscheidenden gesellschaftlichen Institutionen vermutet werden¹⁰. Die kürzeren, etwa fünfzigjährigen Zyklen werden u. a. im Sinne von Kondratieff-Wellen als Folge ökonomischer Entwicklungen interpretiert¹¹.

Die Zyklen sind in mehreren Schritten identifiziert worden. Die britischen Zeitreihen (auf die wir uns in unserer Re-Analyse beschränken) weisen keine gleichbleibenden Intervalle auf; die Thronreden wurden vor 1958 nicht regelmäßig, sondern in Abständen von drei Monaten bis drei Jahren gehalten. Weber hat deshalb jede dieser Reihen über Vier-Jahres-Intervalle aggregiert (A-Reihen): die Worthäufigkeiten jeder einzelnen Kategorie wurden von Thronrede zu Thronrede über vier Jahre addiert, sodann ihr Prozentanteil, gemessen an der aggregierten Nennhäufigkeit aller Kategorien, errechnet. In einem zweiten Ansatz wurden die einzelnen Reihen mit Hilfe eines asymme-

⁷ Ebd., S. 105.

⁸ Siehe ebd., S. 98, Fußn. 15.

⁹ Ebd., S. 106.

¹⁰ Ebd., S. 114 ff.

¹¹ Ebd., S. 93 ff.; Weber 1978 (s. Fußn. 2), S. 98 ff.

trischen Gleitmittelwertes mit einem Stützbereich von 9 Jahren gefiltert (M-Reihe) gemäß

$$Y_t = 1/m(Y_t + Y_{t+1} + \dots + Y_{t+m}),$$

"where Y_t is the new value and m is the length of the filter" ¹². Da der Gleitmittelwert zu regelmäßigen Intervallen führen soll (was normalerweise die Voraussetzung, nicht das Ergebnis seiner Anwendung ist), kann mit $m=9$ nur die Zahl der Jahre, nicht die Menge der Beobachtungen gemeint sein.

Aggregation und gleitende Durchschnitte können bekanntermaßen zyklische Komponenten in den Zeitreihen hervorrufen, die in den ursprünglichen Daten nicht vorhanden sind ¹³. Weber glaubt, sich durch die Wahl zweier unterschiedlicher Glättungsverfahren hinlänglich gegen das Auftreten methodischer Artefakte geschützt zu haben: Werden in der A-Version einer Reihe die gleichen L- oder K-Zyklen identifiziert wie in der M-Version, scheinen diese vom Glättungsverfahren unbeeinflusst zu sein ¹⁴.

Aus allen Zeitreihen wurde zunächst die lineare Trendkomponente als Funktion der Zeit mit einfacher Regressionsanalyse (OLS-Regression) eliminiert. An die trendbereinigten Reihen wurden sodann in einem iterativen Fitting-Verfahren Sinuskurven für die langen Zyklen (L-Zyklen) angepaßt. Die Iteration stoppte, wenn eine weitere Änderung von Periodenlänge, Phase oder Amplitude die Anpassung nicht mehr verbessern konnte. Als Kriterium für die Anpassungsgüte diente das Bestimmtheitsmaß (coefficient of determination) r^2 ¹⁵.

Soweit L-Zyklen auf diese Weise angepaßt werden konnten, "erklärten" sie im Durchschnitt 24 % der Varianz in den trendfreien britischen Zeitreihen (A-Reihen) der kapitalistischen Periode von 1795 - 1972 ¹⁶.

¹² Namenwirth u. Weber (s. Fußn. 1), S. 93, Fußn. 12.

¹³ Siehe unten die Ausführungen zur Filtertheorie.

¹⁴ Ebd., S. 122.

¹⁵ Zur Beschreibung des Verfahrens siehe Weber 1978 (siehe oben Fußnote 2), S. 87 ff.

¹⁶ In Namenwirth u. Weber (s. Fußn. 1) wird lediglich über die mit M-Reihen erzielten Ergebnisse berichtet, die mit den A-Reihen replizierbar sein sollen. Über die A-Reihen berichtet Weber aber in seiner Dissertation (s. Fußn. 2). Dabei treten einige bemerkenswerte Diskrepanzen auf. So wird z. B. für die aggregierte Reihe POWER PARTICIPANTS ein Hochpunkt für 1970 ausgewiesen, dem bei einer Periodenlänge von $\lambda = 156$ auch ein Hochpunkt in 1814 entspricht. Für die Gleitmittelreihe wird dagegen bei einer Periodenlänge von $\lambda = 183$ ein Hochpunkt für 1874 angegeben.

Die Wellenlängen schwanken in der ersten Untersuchungsperiode (1689 - 1795; 107 Jahre) zwischen 49 und 110 Jahren bei einer mittleren Wellenlänge (Median) von 72 Jahren, im zweiten Untersuchungszeitraum (1795 - 1972; 178 Jahre) zwischen 80 und 200 Jahren bei einem Median von 148 Jahren. Das bedeutet, daß in den weitaus überwiegenden Fällen keine einzige Wiederholung des L-Zyklus in den Daten beobachtet werden konnte. (Die K-Zyklen, die wir hier nicht näher betrachten, wurden nach Subtraktion der längeren Zyklen in gleicher Weise durch Anpassen einer Sinuskurve "identifiziert").

Für die inhaltliche Interpretation "standardisierte" Weber die Wellenlängen der einzelnen Zyklen auf den Median der jeweiligen Epoche:

"The most parsimonious and hence desirable assumption is that change in value concerns proceeds at one wavelength rather than at a different wavelength for different value or content categories. Therefore, the estimated wavelengths and peaks, or points of maximal concern, were standardized to the median wavelength and median peak year."¹⁷

Die entsprechende Formel lautet:

$$\text{standard peak} = (\text{original peak} * \text{median wavelength}) / \text{observed wavelength}$$

Diese Vorgehensweise halten wir für ziemlich problematisch, da sie ein zirkuläres Element in die Argumentation einführt. Die bereits erwähnte theoretische Interpretation der Zyklen muß sich doch auf die beobachteten Phasenkonstellationen zwischen den Reihen beziehen. Zumindest hätte die Standardisierung für die einzelnen Themengruppen getrennt erfolgen sollen. Die Frage, ob sich die identifizierten Zyklen der verschiedenen Kategorien tatsächlich gemäß den theoretisch interpretierten Themenwechseln gruppieren lassen, wollen wir hier aber nicht weiter verfolgen. Statt dessen konzentrieren wir uns auf die Methode, mit der die Zyklen allererst identifiziert worden sind.

Weber interpretiert und rechtfertigt die iterative Anpassung deterministischer Sinuskurven u. a. mit folgenden Hinweisen:

"Fitting sine curves to the data presumes that the underlying process is deterministic rather than stochastic. In fact, these data can be analyzed with either method (...) but different answers result. In the case of sine curves, the

¹⁷ Namenwirth und Weber (s. Fußn. 1), S. 95.

results indicate long-term dynamics that are not stochastic. One can also fit second- or third-order autoregressive models to the data. These suggest that the observed dynamics over time result from previous values of the variable plus random shocks. However, textual evidence strongly suggests, and the theoretical arguments presented support the contention, that long-term deterministic cycles in changing thematic concerns exist in these texts. Thus we prefer the deterministic to the stochastic model."¹⁸

Diese Bemerkungen scheinen uns auf einer Reihe von Mißverständnissen zu beruhen:

1. Die Vorstellung eines deterministischen Zyklus impliziert die Annahme, daß Periodenlänge, Amplitude und Phase der Wellenbewegung über Zeit konstant sind. Eine solche Annahme ist aus keiner uns bekannten soziologischen Theorie kultureller Entwicklung ableitbar und dürfte auch kaum mit bisher vorliegenden empirischen Beobachtungen vereinbar sein. Das von Namenwirth und Weber herangezogene AGIL-Schema beispielsweise läßt allenfalls Prognosen zu, wonach bestimmte Themenkomplexe in sozialen Systemen sukzessive und in einer wiederkehrenden Folge unterschiedlicher Intensitäten bearbeitet werden. Daß den unterschiedlichen Themenkomplexen in den Phasen ihrer jeweiligen Hochkonjunktur die gleiche Zeitspanne erhöhter Aufmerksamkeit gewidmet wird oder daß themenspezifische Periodenlängen über lange Zeiträume konstant sind, läßt sich aus den Thesen von Parsons (und Bales) nicht ableiten. Die Annahme deterministischer Wertezyklen kollidiert sogar mit Namenwirths und Webers Erklärungsversuchen: Weder verlaufen die als Einflußfaktoren genannten Wählerschichtungen noch die ökonomischen Auf- und Ab-Bewegungen deterministisch. Auch die Prozesse der Generationenbildung (bzw. "population replacement"), die die Wertezyklen kausal antreiben sollen, sind in ihrem Ablauf sicher nicht deterministisch.

2. Wenn eine zyklische Bewegung nicht deterministisch verläuft, kann es dennoch u. U. pragmatisch sinnvoll sein, diese Komponente mit einem deterministischen Modell grob zu beschreiben und anschließend aus der Zeitreihe zu eliminieren, um, beispielsweise, irreguläre Schwankungen visuell deutlicher hervortreten zu lassen¹⁹. Im vorliegenden Falle gilt jedoch gerade

¹⁸ Ebd., S. 94.

¹⁹ Dabei kann man sich bestimmter Hilfsmittel bedienen, zum Beispiel: Je stärker sich ein zyklischer Prozeß einer deterministischen Verlaufsform nähert, um so stärker geht die entsprechende Spektraldichte von einem Band- in ein Linienspektrum über,

den Zyklen das besondere Interesse der Forscher. Da sie von Seiten der "inhaltlichen" Theorie nicht als deterministisch ausgewiesen werden können, ist es zumindest fraglich, ob sie sich durch deterministische Modelle hinreichend genau (nur für die Beobachtungsperiode) beschreiben lassen. Schon geringfügige Phasenverschiebungen können sich fatal auswirken, wenn die zur Debatte stehende Theorie explorativ aus beobachtbaren Phasenkonstellationen und Zyklenkohärenzen abgeleitet oder gar anhand solcher Konstellationen getestet werden soll. Auf keinen Fall kann die Frage der Adäquanz eines deterministischen Modells mit Hilfe einer reinen Fittingprozedur beantwortet werden. Wenn Weber eine spektralanalytische Zerlegung mit dem Argument zurückweist, sie erfordere Zeitreihen, die siebenmal so lang seien wie die Zyklusperiode,²⁰ so trifft dies seine eigene Methode erst recht. Um einen deterministischen Zyklus nachweisen oder auch nur entsprechende Annahmen empirisch stützen zu können, muß man den Zyklus in mehrfacher Wiederholung beobachtet haben. Eine sich nicht wiederholende Sinuskurve kann man an viele Reihen auch dann anpassen, wenn kein deterministisch-zyklischer, ja nicht einmal ein pseudoperiodischer Prozeß gegeben ist. So z. B. können Random-Walk-Prozesse wechselnde stochastische Trendverläufe ("drift") aufweisen, an die sich innerhalb einer begrenzten Untersuchungsperiode leicht eine Sinuskurve iterativ anpassen ließe. Eine sich nicht wiederholende Sinuskurve anpassen zu können, beweist hinsichtlich der stochastischen oder deterministischen Eigenschaften der zu analysierenden Reihe nichts.

3. Die Sinusmodelle, die Namenwirth und Weber angepaßt haben, sind außerdem mit dem großen Nachteil behaftet, daß ihre Parameter nicht statistisch testbar sind; jedenfalls wird für die iterative "Schätz"prozedur keine wahrscheinlichkeitstheoretische Verankerung angegeben. Webers Bemerkung: "when the population of speeches is analyzed and classical inferential statistics are inappropriate ..." ²¹ ist irreführend. Man will ja in der statistischen Zeitreihenanalyse stets von den Beobachtungsdaten (die allemal von Meßfehlern und sonstigen Zufallseinflüssen mit bestimmt sind) auf den "dahinter liegenden" Erzeugungsprozeß schließen, ihn "modellieren". Das

während die Autokorrelationsfunktion weniger stark gedämpft wird. Außerdem nimmt bei deterministischen Zyklen die im Periodogramm für die entsprechende Frequenz ausgewiesene Dichte proportional zur Länge der beobachteten Zeitreihe zu - ein Tatbestand, den man ebenfalls für diagnostische Zwecke nutzen kann. Siehe hierzu J. M. Gottman, *Time-Series Analysis. A Comprehensive Introduction for Social Scientists* (Cambridge u. a. 1981).

²⁰ Siehe Namenwirth u. Weber (s. Fußn. 1), S. 94.

²¹ Siehe Namenwirth u. Weber, *Dynamics* (Fußn. 1), S. 122.

setzt voraus, daß man die beobachteten Werte (die zudem aus einer diskreten Stichprobe stammen, die aus einem zeitkontinuierlichen Prozeß gezogen wurde) als Realisationen eines mehrdimensionalen Zufallsprozesses betrachtet.

Fassen wir unsere bisherige Argumentation zusammen: Ein sauberes methodisches Vorgehen muß zwei Aspekte trennen: (a) die Identifikation der Zyklen, die in einer Beobachtungsreihe vielleicht vorhanden sind; (b) ihre stochastische oder deterministische Modellierung, falls sie identifiziert werden können. Die Identifikationsinstrumente dürfen in ihrer Anwendung nicht schon selbst das Modell, nicht einmal den Modelltyp voraussetzen. Ein solches Instrument zur Identifikation von Zyklen liefert die Theorie linearer Filter, deren Grundgedanken wir im nächsten Abschnitt skizzieren²².

2. Zur Theorie linearer Filter

Bei der Identifizierung von Zyklen hat sich die Forschung bislang meist auf das "klassische" Komponentenmodell gestützt. Grundlegend für dieses Modell ist die Annahme, daß eine Zeitreihe die Summe (das Produkt) verschiedener Komponenten darstellt. Meist geht man dabei von den 4 Komponenten Trend, Konjunktur, Saison und Zufall aus. Für den Nachweis von Zyklen verfährt man dabei i.d.R. so, daß zuerst der Trend der Reihe bestimmt und dieser dann von den Originalwerten subtrahiert wird. Als Trend kommen entweder deterministische Funktionen der Zeit oder gleitende Mittelwerte zur Anwendung. Das Ergebnis einer solchen Trendbereinigung ist die trendfreie Reihe, die dann gemäß den Annahmen des Komponentenmodells die Komponenten Konjunktur, Saison und Zufall enthält. Unterstellt man Jahreswerte, was wir im folgenden tun wollen, verbleiben Konjunktur und Zufall. Um den Zufall auszuschalten, wird die trendfreie Reihe - meist mit Hilfe gleitender Mittelwerte - geglättet. Das Ergebnis - die glatte Komponente - wird als Konjunktur interpretiert. Obwohl der hier skizzierte Ablauf als sehr vereinfachtes Schema bezeichnet werden kann, entspricht er grundsätzlich, auch wenn ihm noch einige statistische Raffinessen hinzugefügt werden, dem traditionellen Vorgehen. Gegen dieses auch heute noch am meisten angewandte Verfahren sind einige schwerwiegende methodologische Einwände vorgebracht worden:

- 1) Die Festlegung einer bestimmten Trendfunktion läßt sich substanzwissenschaftlich meist nicht begründen. Die Entscheidung des Forschers z.B. für

²² Vgl. den Beitrag von Winfried Stier in diesem Band.

eine logistische oder Exponentialfunktion als Trendkurve ist immer mehr oder weniger willkürlich.

2) Auch das Ausmaß der Glättung, also die Länge des Stützbereichs der gleitenden Mittelwerte, läßt sich substanzwissenschaftlich kaum begründen; ob z.B. ein gleitender Mittelwert der Länge 9 oder 13 genommen wird, kann ohne zusätzliche Informationen kaum entschieden werden.

3) Viele Verfahren der Trendbereinigung und auch der Glättung stehen in dem Verdacht, künstliche Zyklen zu erzeugen (Slutsky-Effekt).

Diese Überlegungen zeigen, daß beim "klassischen" Komponentenmodell die Ergebnisse nicht frei sind von der Willkür des Forschers und zudem immer die Gefahr besteht, daß die gefundenen Zyklen durch die benutzten statistischen Verfahren künstlich erzeugt wurden. Aus diesen Gründen hat sich gegenüber diesem Vorgehen eine starke Methodenskepsis etabliert.

Abhilfe bei diesen Schwächen schien die in den 70er Jahren vor allem in der ökonomischen Forschung sich durchsetzende Spektralanalyse zu bringen. Die Spektralanalyse erlaubt bekanntermaßen die Identifikation der zyklischen Struktur einer Reihe, ohne daß diese vorab transformiert, d.h. geglättet werden muß. Das Ergebnis der Spektralanalyse ist das geschätzte Spektrum, das nicht nur Auskunft über die vorhandenen Zyklen, sondern auch über deren Bedeutung für die Gesamtvarianz der Reihe gibt. Mit Hilfe der Spektralanalyse sind einige umfangreiche Untersuchungen vor allem über das Konjunkturmuster durchgeführt worden. Erst die empirische Analyse langfristiger Zyklen hat auf die meist nicht erfüllten Anwendungsvoraussetzungen und Schwächen der Spektralanalyse deutlich hingewiesen.

Die Spektralanalyse basiert auf der Theorie der (schwach) stationären stochastischen Prozesse, d.h., die Ergebnisse können nur dann eine gewisse Signifikanz beanspruchen, wenn die zu analysierenden Zeitreihen schwach stationär sind, vor allem also keinen Trend aufweisen. Da die meisten Zeitreihen jedoch einen Trend zeigen, können diese Reihen - jedenfalls nicht unmittelbar - mit Hilfe der Spektralanalyse untersucht werden. Da die Spektralanalyse nur auf trendfreie Reihen sinnvoll angewandt werden kann, müssen die Zeitreihen vorab trendbereinigt werden. Die mit der Trendbereinigung verbundene Problematik bezüglich der spektralanalytischen Ergebnisse wurde in den meisten Untersuchungen nicht in der notwendigen Schärfe diskutiert. Erst allmählich hat sich die Erkenntnis durchgesetzt, daß die Art der Trendbereinigung die Ergebnisse wesentlich determiniert. Dies umso stärker, je kleiner die Frequenz der zu analysierenden Zyklen ist. M.a.W., gerade bei der Suche nach Langfristzyklen hängen die spektralanalytischen Ergebnisse

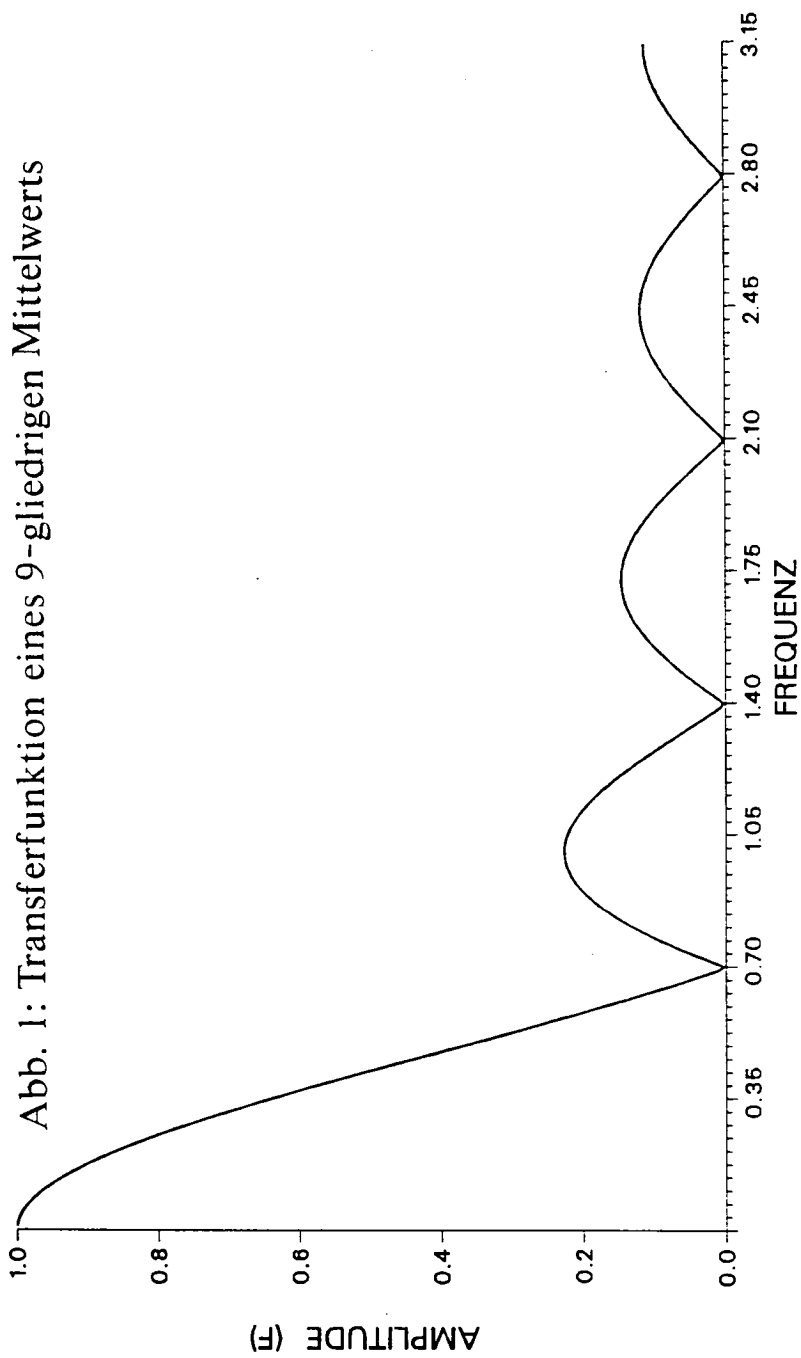
wesentlich von der durchgeführten Trendbereinigung ab. So kann man zeigen, daß der spektralanalytische Nachweis der sog. Kondratieffzyklen primär von der Art der Trendbereinigung abhängt.

Sowohl die Kritik an dem Komponentenmodell als auch die Einsicht in die starken Beschränkungen der Spektralanalyse haben zu völlig neuen Forschungen auf diesem Gebiet Anlaß gegeben. Vor allem die Erfahrungen bei der Suche nach "langen Wellen", aber auch die Suche nach einem "idealen" Saisonbereinigungsverfahren haben zur Entwicklung eines neuen methodischen Instrumentariums der Zeitreihenanalyse geführt, den sog. Filterverfahren. Als Filter bezeichnet man, vereinfacht ausgedrückt, ein Rechenverfahren (Algorithmus) mit dem man eine Input-Zeitreihe in eine Output-Zeitreihe transformiert. Eine Zeitreihe kann man als Summe sehr vieler sich überlagernder Schwingungen unterschiedlicher Schwingungsdauer auffassen. Deshalb kann man auch sagen, daß bei der Filterung bestimmte Schwingungen in die Outputreihe übertragen werden und andere nicht. Benutzt man statt des Begriffs "Schwingung" den der "Frequenz", kann man auch sagen, daß ein Filter den Frequenzgehalt einer Zeitreihe verändert. In der Praxis lassen sich verschiedene Filtertypen unterscheiden, je nach dem, welche Schwingungen übertragen werden. Ein Filterverfahren, mit dem man den Trend eliminiert, überträgt nur Schwingungen kurzer Dauer (hochfrequente Schwingungen), hier spricht man von einem Hochpass. Ein Filter, der nur Schwingungen langer Dauer (niederfrequente Schwingungen) überträgt, wird als Tiefpass bezeichnet.

Wie ein bestimmtes Verfahren - ein bestimmter Filter - den Frequenzgehalt der Inputreihe verändert, läßt sich nun - und das ist der entscheidende Vorteil dieses Ansatzes - mit Hilfe der Filtertheorie genau berechnen. Für alle linearen zeitinvarianten Filter läßt sich die sog. Transferfunktion berechnen, mit deren Hilfe die Übertragungseigenschaften des Filters angegeben werden können. Ein Wert der Transferfunktion von 1 zeigt, daß die zugehörige Frequenz unverändert übertragen wird. Ein Wert von Null bedeutet, daß die entsprechenden Frequenzen bei der Filterung vollständig eliminiert werden. Der Frequenzbereich, in dem die Transferfunktion den Wert 1 (0) hat wird als Paßband (Stoppband) bezeichnet.

Nach diesen Überlegungen wird klar, daß auch ein gleitender Mittelwert nichts anderes ist als ein Filter. Da er eine lineare zeitinvariante Transformation darstellt, läßt sich auch seine Transferfunktion berechnen. Für die von Namenwirth/Weber benutzten 9-gliedrigen Mittelwerte hat die Transferfunktion die in Abb. 1 dargestellte Form.

Abb. 1: Transferfunktion eines 9-gliedrigen Mittelwerts



Es zeigt sich, daß alle Frequenzen zwischen $1/9$ und 0 abgeschwächt übertragen werden. Die Funktion hat bei $1/9$ ihre erste Nullstelle, so daß alle Schwingungen mit dieser Frequenz vollständig eliminiert werden. Da die Transferfunktion in keinem Bereich Werte größer als 1 hat, entstehen keine Slutzky-Effekte.

Da zahlreiche substanzwissenschaftliche Fragestellungen auf Filter mit ganz unterschiedlichen Transferfunktionen hinauslaufen, war es von Anfang an das Ziel, ein Verfahren zu finden, mit dem sich beliebige Transferfunktionen realisieren lassen. Realisiert wurde dieses Vorhaben mit einem neu entwickelten Filterverfahren, das gegenüber traditionellen Verfahren einige bedeutsame Vorteile aufweist. Bei diesem Verfahren wird der gesamte Filterprozeß (also Spezifikation der Transferfunktion und die eigentliche Filterung) direkt im Frequenzbereich vollzogen. Die Vorgehensweise ist - stark vereinfacht - etwa die folgende.

Gemäß dem substanzwissenschaftlichen Ansatz werden die Frequenzen spezifiziert, die übertragen (eliminiert) werden sollen. Daraus resultiert eine bestimmte Transferfunktion, deren Paß- und Stoppband entsprechend diesen Definitionen festgelegt wird. Für die Filterung wird nun die Zeitreihe in den Frequenzbereich transformiert (praktisch mit Hilfe der Fast Fourier Transformation). Das Ergebnis ist die sog. Fourier-Transformierte. Diese wird nun gemäß der vorgegebenen Transferfunktion verändert, d. h. im Paßbandbereich (Stoppbandbereich) mit 1 (0) multipliziert. Das Ergebnis wird in den Zeitbereich zurücktransformiert (mathematisch mit Hilfe der Inversen Fourier Transformation), womit man als Ergebnis die gefilterte Zeitreihe erhält.

Dieses Verfahren erfüllt nachweislich wichtige Voraussetzungen, die man an Filter stellen muß:

- 1) Die Transferfunktion ist exakt 1 (0) im Paßband (Stoppband)
- 2) Paß- und Stoppband sind völlig variabel, damit läßt sich jeder gewünschte Filtertyp spezifizieren.
- 3) Der Filter erzeugt keine Verzögerungen zwischen Input und Output; hier sagt man auch, daß die Phase identisch Null im gesamten Frequenzbereich ist.

Die letztgenannte Forderung ist nicht selbstverständlich und für die empirische Forschung von großer Bedeutung. Hat ein Filter nämlich eine Phase, die nicht gleich Null ist, bedeutet dies, daß bei der Zyklusanalyse unerwünschte Verschiebungen der Hoch- und Tiefpunkte eintreten.

Der entscheidende Vorteil des Filter-Design Ansatzes besteht darin, daß sich die Amplitudenfunktionen und damit die Filter für beliebig definierbare

Schwingungskomponenten exakt realisieren lassen. Ergebnis der Filterung sind die jeweiligen in der Zeitreihe enthaltenen Schwingungen, deren Schwingungsdauer dem Paßband der vorgegebenen Amplitudenfunktion zuzuordnen ist. Damit ist gewährleistet, daß evtl. vorhandene Schwingungen so als Zeitreihe dargestellt werden, wie sie als erzeugende Komponenten dem Gesamtprozeß zugrunde liegen. Stellt man sich z.B. eine Zeitreihe vor, die aus mehreren sich überlagernden periodischen Schwingungen mit unterschiedlicher Amplitude und Periodendauer zusammengesetzt ist, so erlaubt der Filter - bei entsprechender Amplitudenvorgabe - die exakte Reproduktion jeder einzelnen dieser Schwingungen. Darüber hinaus ist gewährleistet, daß der Filter keine Schwingungen erzeugt, die in der zu filternden Reihe gar nicht vorhanden sind. Wird also das Paßband des Filters so gewählt, daß ihm entsprechende Schwingungen in der Reihe nicht vorhanden sind, ist das Ergebnis des Filters praktisch eine Null-Linie.

Diese Eigenschaften haben für die Interpretation der Ergebnisse weitreichende Konsequenzen, und zwar in zweierlei Hinsicht. Der Filter reproduziert - wie bereits erwähnt - genau und damit ausschließlich die Komponenten, deren Periodendauer dem Paßband des Filters entspricht. D.h., der Filteroutput ist genau in dieser Form Bestandteil des Filterinputs. Darin unterscheidet sich dieser Ansatz von allen Verfahren, bei denen die gesuchten Komponenten geschätzt werden. Die Spezifikation der Amplitudenfunktion und das eigentliche Filtern erfolgen gänzlich unabhängig von irgendwelchen schätztheoretischen Überlegungen oder Kriterien. Der Filter-Design Ansatz kann in diesem Sinne als parameterfreies Verfahren betrachtet werden, worin sicherlich ein entscheidender Vorzug gegenüber den traditionellen Ansätzen zu sehen ist. Bei diesem Verfahren werden also keine Modelle vorausgesetzt, deren Parameter zu schätzen wären.

Der zweite hier zu berücksichtigende Punkt betrifft die definitorische Festlegung der gesuchten Komponenten und die Möglichkeit, den Realitätsgehalt und die Bedeutung der gefilterten Komponenten zu beurteilen. Da, wie bereits erwähnt, dieses Verfahren die Operationalisierung beliebiger Definitionen erlaubt, wird z. B. der Trend nicht als ein spezifischer Funktionstyp festgelegt, sondern es werden die Frequenzkomponenten definiert, die den Trend konstituieren sollen. Welche aber sind das: Sind es alle Schwingungen mit einer Dauer von Unendlich bis 100 Jahren oder von Unendlich bis zu 200 Jahren: Hier zeigt sich sehr rasch, daß viele substanzwissenschaftliche Begriffe, wie z.B. Trend oder Zyklus, in aller Regel viel zu ungenau sind, um ohne weitere Annahmen operationalisiert werden zu können. Da selbstver-

ständig das resultierende Ergebnis von der jeweiligen Definition (im Frequenzbereich) abhängt, ist offensichtlich, daß der zugrundeliegenden Definition entscheidende Bedeutung zukommt. Im Falle der Trenddarstellung hängt der resultierende Trendverlauf eben von der verwendeten Trenddefinition ab.

Nicht anders ist es beim Zyklus. Die Form des gefundenen Zyklus wird sich in Abhängigkeit von der vorgegebenen Zyklendefinition gestalten. Die Beurteilung von zyklischen Prozessen im Hinblick auf ihren Realitätsgehalt wird in der Praxis noch dadurch erschwert, daß konkrete Zeitreihen Schwingungskomponenten aller Frequenzbereiche enthalten.

Als Problem bleibt: Wie - anhand welcher Kriterien - läßt sich der Realitäts- oder Bedeutungsgehalt der extrahierten Komponenten beurteilen? Anders gefragt, wie läßt sich die Definition der zu extrahierenden Komponenten so festlegen, daß diesen ein forschungsspezifischer Bedeutungsgehalt zukommt? Für eine Beantwortung dieses Problems muß man sich zu allererst fragen, was es bedeutet, den Realitätsgehalt bestimmter Zeitreihenkomponenten beurteilen zu wollen. Mit dem Realitätsgehalt ist offensichtlich die Vorstellung verbunden, daß die gefilterte Komponente für die zugrunde liegende Zeitreihe von herausragender Bedeutung ist. Und zwar herausragend im Vergleich zu anderen Schwingungskomponenten (mit ähnlichem Erklärungswert). Für den Gesamtprozeß, der durch die trendfreie Reihe repräsentiert wird, muß die herausgefilterte (Zyklen-)Komponente einen "wesentlichen" Beitrag liefern.

Zur Beurteilung dieses Problems zieht man in der Forschung häufig die durch die jeweilige Komponente erklärte Varianz heran. Dieser Maßstab (im allgemeinen das Quadrat des Pearson Produkt-Moment Korrelationskoeffizienten r) ist aber sowohl für die Festlegung der Definition als auch für die Beurteilung des Realitätsgehalts wenig geeignet. Erstens nimmt die erklärte Varianz bei einer Vergrößerung des Paßbandes prinzipiell zu und zweitens bleibt das Amplitudenverhältnis der beiden Komponenten in dieser Maßzahl unberücksichtigt²³.

U. E. kann im Augenblick nur die Spektralanalyse Hinweise auf die in trendfreien Reihen enthaltenen dominierenden zyklischen Prozesse liefern. Um hier Mißverständnissen vorzubeugen: Der Filter-Design-Ansatz hat nichts mit der Spektralanalyse zu tun. Die Spektralanalyse ist ein Verfahren, mit dem für trendfreie Reihen - also stationäre Prozesse - die Struktur der zugrunde liegenden Zyklen geschätzt wird, während es sich beim Filtern um eine Transformation von Prozessen gemäß vorgegebener Kriterien handelt.

²³ Vgl. unten die Ausführungen zu unserer Tabelle 1.

Wenn hier nach dem Bedeutungsgehalt spezifischer Komponenten für den Gesamtprozeß gefragt wird, so lassen sich erste Hinweise auf solcherart bedeutende Komponenten am besten mit Hilfe der Spektralanalyse ableiten. Gemäß den spektralanalytischen Ergebnissen können dann entsprechende Filter bzw. Amplitudenfunktionen gewählt werden. Die Ergebnisse der Filter stellen die real-historische Ausformung der Komponenten im Zeitbereich dar und sind damit die Grundlage für die substanzwissenschaftliche Beschreibung und Erklärung des gesuchten Prozesses.

Das hier skizzierte Vorgehen trägt der Tatsache Rechnung, daß sich die Existenz von Zyklen streng genommen nur mittels einer explorativen Datenanalyse nachweisen läßt. Insofern erscheinen auch die von Namenwirth und Weber durchgeführten Fitting Prozeduren wenig geeignet, da bekannt ist, daß sich an konkrete Zeitreihen eine Vielzahl unterschiedlicher Kurven anpassen läßt und daß es für die Beurteilung der Güte der Anpassung keine wirklich "harten" Kriterien gibt. Der Nachweis von zyklischen Komponenten birgt immer die Gefahr, daß genau das bestätigt wird, was der Forscher erwartet.

Da wir uns bei den folgenden Berechnungen an den von Namenwirth und Weber aufgestellten Hypothesen über die Existenz von Zyklen orientieren, spielt die normalerweise substanzwissenschaftlich zu begründende Definition der zu extrahierenden Komponenten hier vorerst nur eine untergeordnete Rolle. Daß trotzdem gegenüber diesen Autoren abweichende Ergebnisse erwartet werden können, liegt in der Tatsache begründet, daß beim Filtern die herausgefilterte Reihe in der jeweiligen Form immer schon Bestandteil der zu filternden Zeitreihe ist, was beim Kurven-Fitting ja nicht der Fall ist.

Darüber hinaus führen wir alternative Berechnungen durch, die auf unterschiedlichen Komponentendefinitionen beruhen, da wir die Annahmen von Namenwirth und Weber für zu restriktiv halten. Diese auf unterschiedlichen Definitionen beruhenden Berechnungen sollen vor allem Problematik und jeweilige Aussagefähigkeit der Ergebnisse demonstrieren.

3. Analyseergebnisse

Aus Webers Liste erfolgreich identifizierter L-Zyklen für den Zeitraum von 1795 bis 1972 haben wir, mehr oder weniger willkürlich, zehn Zeitreihen ausgewählt²⁴, bei denen die jeweils angepaßte Sinuskurve laut Weber zwischen 9 % und 54 % der Gesamtvarianz der trendbereinigten Reihe aus-

²⁴ Wir danken R. Ph. Weber, der uns seine Daten zur Verfügung stellte. Geringfügig abweichend von der in seiner Arbeit gegebenen Periodisierung der kapitalistischen Epoche, sind die uns überlassenen Aggregatdaten von 1798 bis 1974 datiert.

macht. Wir beschränken uns hier auf die vierjährlich aggregierten Daten, die weniger stark geglättet sind als die Gleitmittelreihen. In dem Band von Nannenwirth und Weber²⁵ finden sich nur Angaben zu den Gleitmittelreihen; entsprechende Informationen über die aggregierten Reihen sind jedoch Webers Dissertation²⁶ zu entnehmen.

Wir haben die im vorigen Abschnitt skizzierte Filtermethode in verschiedener Weise eingesetzt und die dabei erzielten Ergebnisse mit Webers Sinuskurven verglichen. In einem ersten Ansatz (im folgenden: "Methode 1") wurden aus den Zeitreihen alle Schwingungskomponenten mit einer Periodenlänge von über 200 Jahren eliminiert. Dies entspricht dem üblichen Vorgehen, dem "Trend" alle Schwingungen zuzuweisen, deren Periodenlängen über die Länge des Beobachtungszeitraumes hinausgehen. Mit einem weiteren Filter wurden sodann aus den Restreihen alle Schwingungskomponenten extrahiert, die eine Länge zwischen 90 und 200 Jahren aufweisen. Das heißt, wir isolieren und bündeln in jeder der zehn Zeitreihen die Schwingungen desjenigen Frequenzbereichs, den Weber für seine "long cycles" in Anspruch nimmt. (Einer seiner Zyklen, RECTITUDE ETHICS, ist mit einer Periode von $\lambda = 208$ etwas länger, eine Reihe, POWER-ARENA, die hier nicht betrachtet wird, ist mit $\lambda = 80$ etwas kürzer.) Danach konnte geprüft werden, ob diese Schwingungen überhaupt als periodisch anzusehen und wie im Vergleich zu Webers Sinuskurven die Hoch- bzw. Tiefpunkte plaziert sind.

In einem zweiten methodischen Zugang ("Methode 2") haben wir Webers eigene Trenddefinition übernommen, also lediglich die lineare Trendkomponente aus den Zeitreihen eliminiert und aus der bereinigten Reihe die Schwingungskomponenten mit einer Periodenlänge von $\lambda > 90$ herausgefiltert. Die so extrahierten Komponenten wurden wiederum nach ihrer Periodizität und der Lage der Hoch- und Tiefpunkte beurteilt.

In einem dritten Schritt ("Methode 3") haben wir nach der linearen Trendbereinigung aus der Restreihe nur diejenigen Schwingungskomponenten herausgefiltert, deren Frequenzen derjenigen Sinuskurve entsprechen, die Weber spezifisch für diese bestimmte Reihe angepaßt hat. Das heißt, es wurden alle Schwingungskomponenten eliminiert, deren Periodenlängen um mehr als (ungefähr) ± 5 Jahre von der Periodenlänge der "gefitteten" (deterministischen) Sinuskurve abweichen. Webers These ist ja, daß die kulturelle Entwicklung von einer bestimmten, deterministisch ablaufenden Problemlösungsdynamik angetrieben wird. Wenn diese These stimmt, muß sich

²⁵ Siehe oben Fußnote 1.

²⁶ Siehe oben Fußnote 2.

der entsprechende Zyklus in der beobachteten Reihe mit einem relevanten Varianzanteil auffinden lassen, nachdem das Gemenge von Schwingungskomponenten anderer Frequenzen aus der Reihe eliminiert worden ist.

Wir wollen zunächst die einzelnen Verfahrensschritte anhand der Reihe RECTITUDE-ETHICS veranschaulichen:

Abb. 2 zeigt die aggregierte Reihe sowie den Trend, der mit Hilfe des Tiefpass-Filters bestimmt wurde, der Schwingungslängen von ∞ bis zu 200 Jahren reproduziert. Abb. 3 bringt die um diesen Trend bereinigte Reihe (gezackte Linie), die alle Schwingungen (einschließlich "white noise") mit der Periodenlänge $2 < \lambda < 200$ Jahren enthält (Hochfrequenzreihe). Außerdem zeigt sie mit der glatten Linie die zusammengefaßten Schwingungen, die eine Periodenlänge von $200 > \lambda > 90$ aufweisen. Sie verlaufen nicht streng periodisch; das linke Wellental ist länger als das rechte und weniger tief. Ein Hochpunkt läßt sich für das Jahr 1902 ermitteln, während Weber einen Hochpunkt für das Jahr 1880 (standardisiert: 1875) festgestellt hat. Tiefpunkte in den Jahren 1950 und 1848/49 legen eine Periodenlänge (wenn überhaupt) von etwa 100 Jahren nahe. Weber hingegen hat hier eine Periodenlänge von 176 Jahren ermittelt.

Im nächsten Analyseschritt (Methode 2) wird die Reihe als erstes um den linearen Trend (eingezeichnet in Abb. 4) bereinigt. Die Restreihe ist in der gezackten Linie der Abb. 5 sichtbar. Die glatte Linie in Abb. 5 stellt alle (zusammengefaßten) Schwingungskomponenten der trendfreien Reihe mit $\lambda > 90$ dar. Die Periodizität verschwindet nun noch mehr; der linke, sehr flache Tiefpunkt liegt bei 1806, also fast doppelt so weit vom Hochpunkt (in 1901) entfernt wie der rechte Tiefpunkt (1953) - nicht gerade ein Beleg für einen deterministischen Zyklus. Selbst wenn man die in der rechten Hälfte sichtbare Welle als periodisch auffassen wollte, wären ihr Hochpunkt und ihre Länge (wiederum etwa 100 Jahre) mit Webers Sinuskurve nicht vereinbar.

Bei der dritten Analysevariante werden nach der linearen Trendbereinigung aus der Restreihe in einem engen Frequenzband nur diejenigen Schwingungskomponenten extrahiert, die der Sinuskurve entsprechen, die Weber an diese Reihe angepaßt hat (siehe glatte Linie in Abb. 6). Diese "Schwingungen" weichen nur noch geringfügig, optisch nicht mehr wahrnehmbar, von einer Abszissenparallele im Nullpunkt ab. (Auch Weber hat für seine Sinuskurve in diesem Fall nur 9 % Varianzanteil reklamiert.) Numerisch läßt sich ein Hochpunkt für das Jahr 1810 ermitteln - 70 Jahre vor dem Peak, den Weber identifiziert hat. Akzeptiert man den Tiefpunkt im Jahre 1892 als Pe-

Abb. 2: Rectitude-Ethics

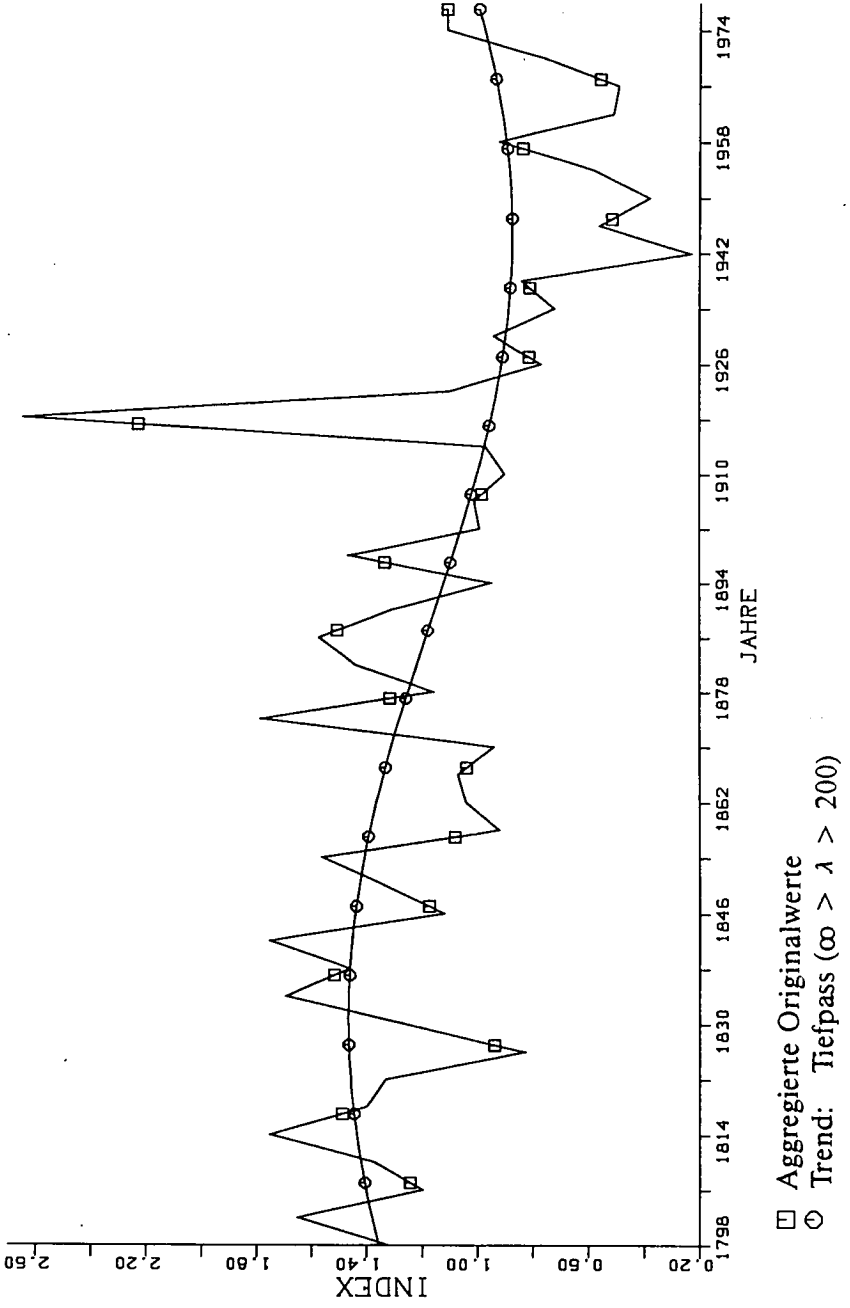
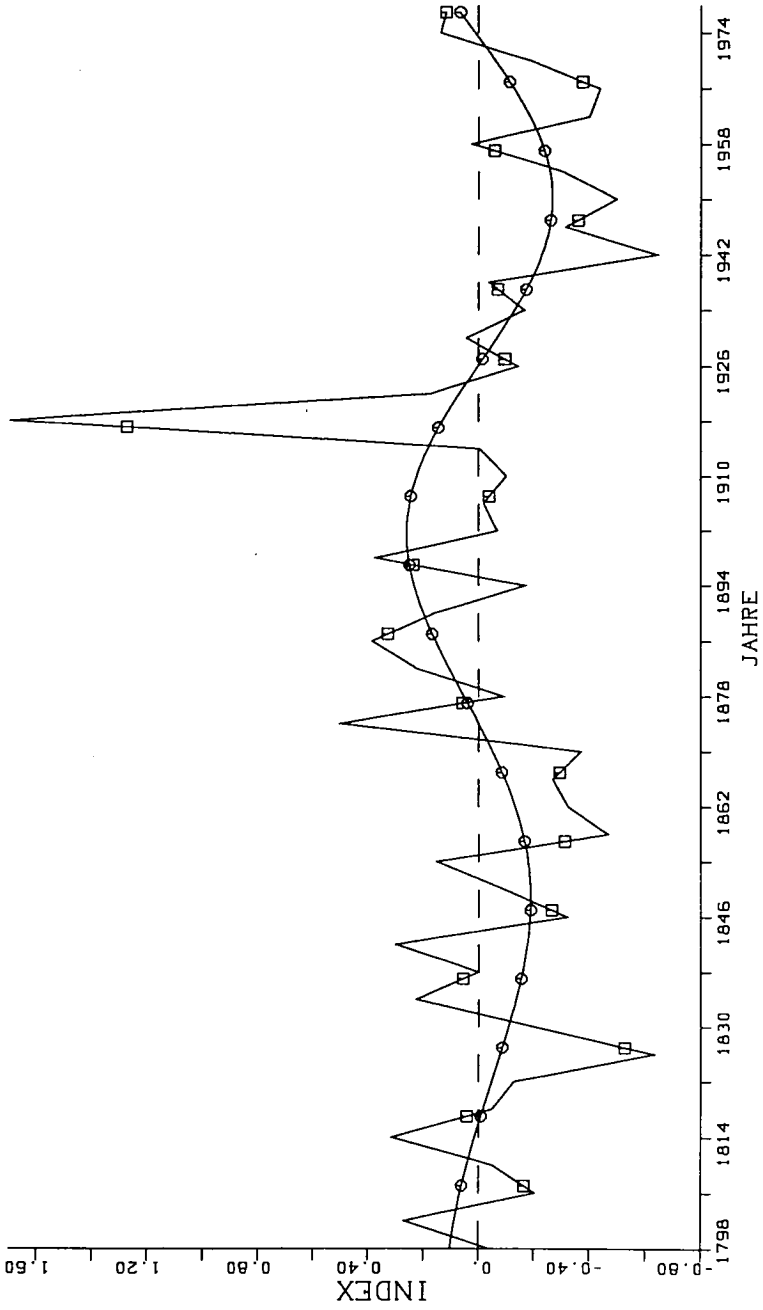


Abb. 3: Rectitude-Ethics



□ Trendfreie Werte: Hochpass ($200 > \lambda > 2$)
 ○ Glatte Komponente: Bandpass ($200 > \lambda > 90$)

Abb. 4: Rectitude-Ethics

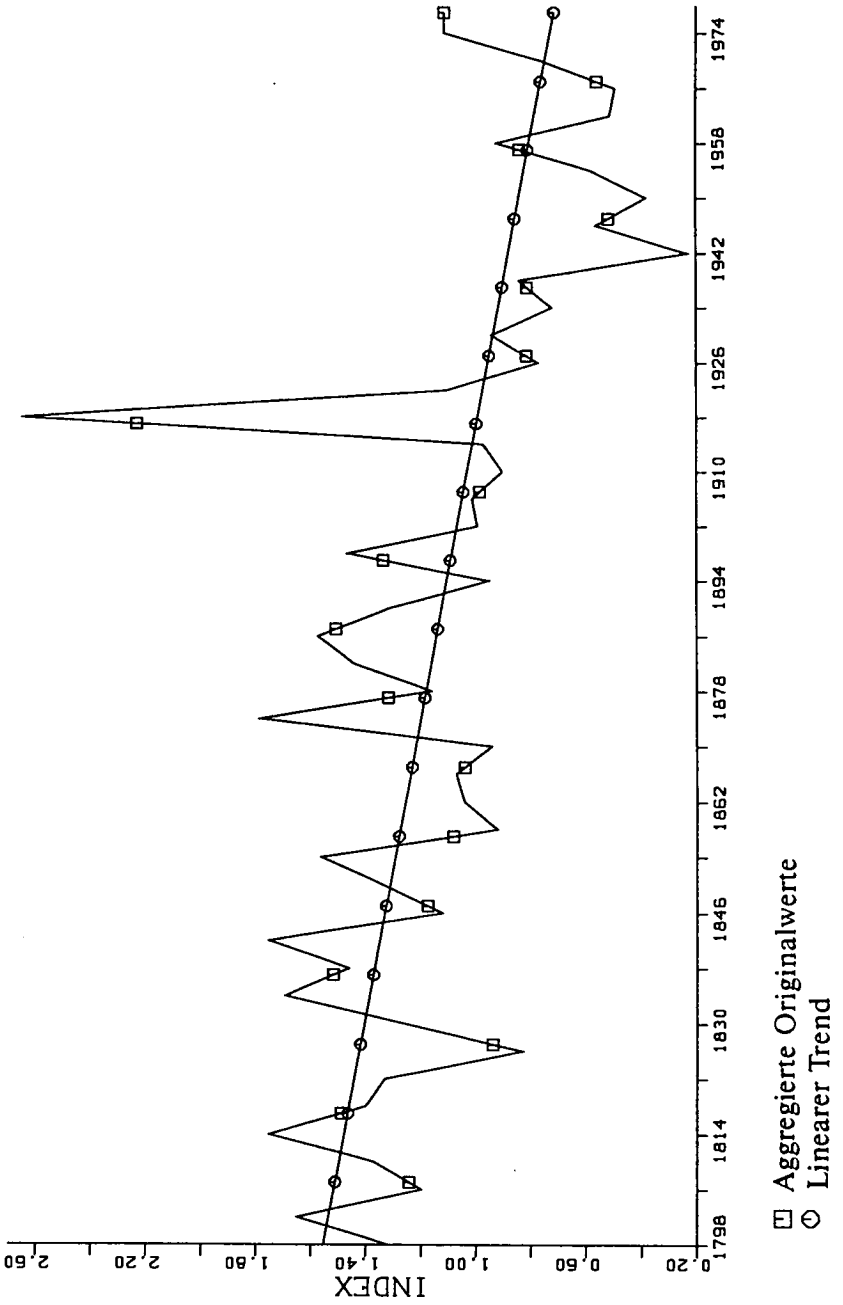


Abb. 5: Rectitude-Ethics

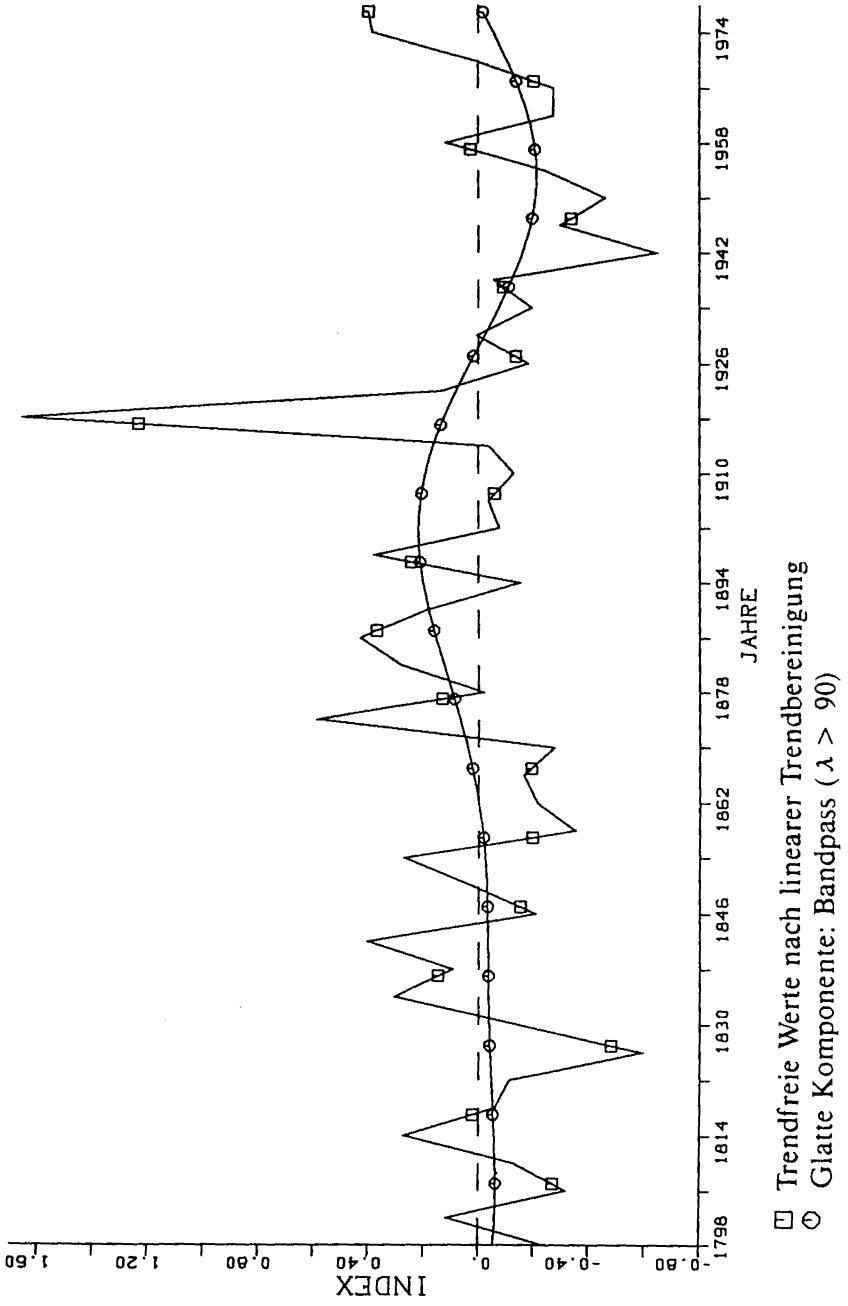
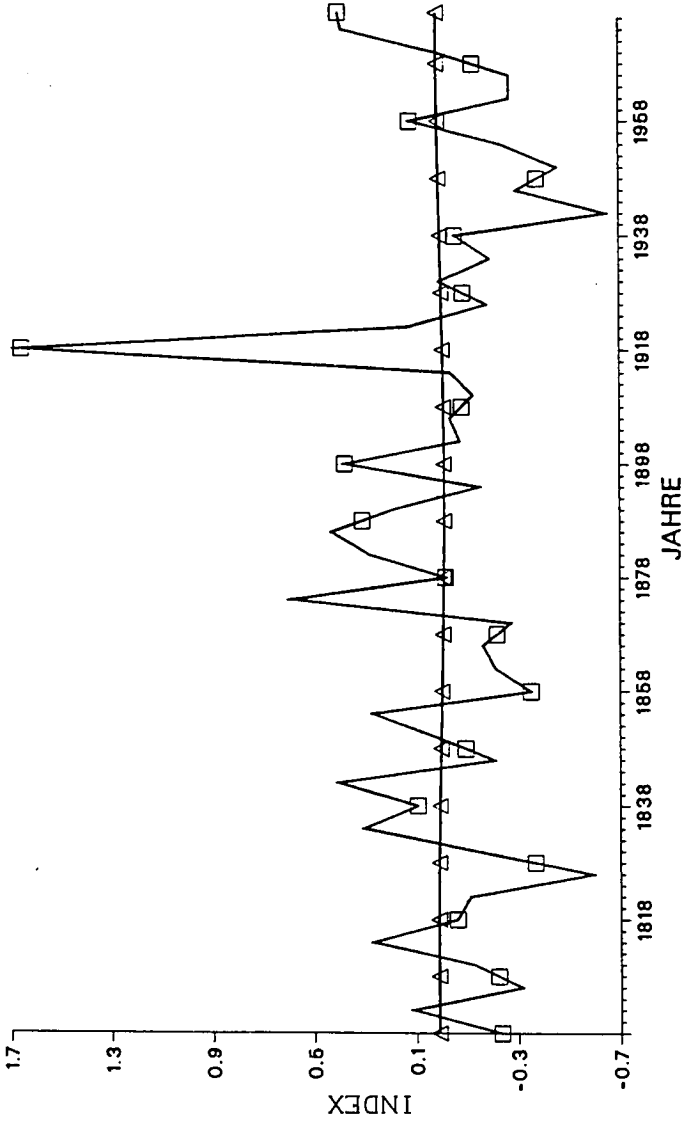
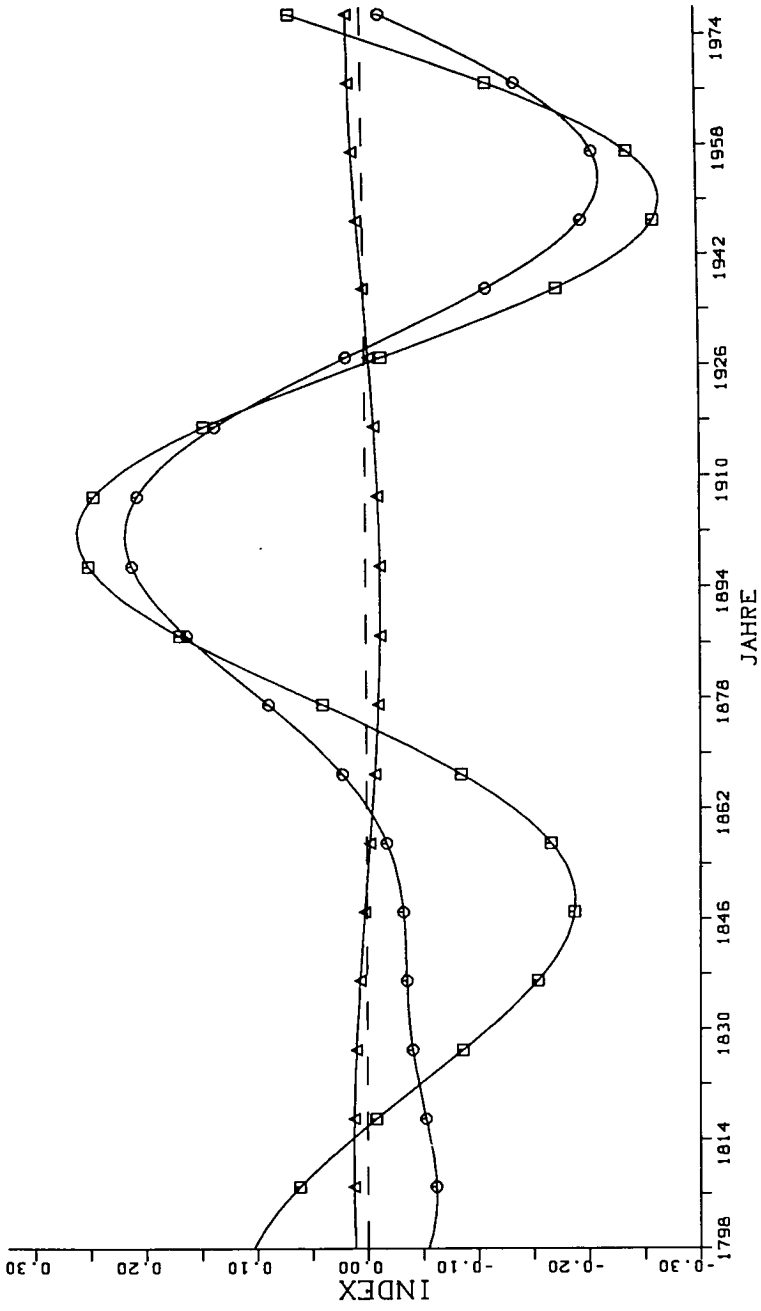


Abb. 6: Rectitude-Ethics



- Trendfreie Werte nach linearer Trendbereinigung
- △ Glatte Komponente: Bandpass ($190 > \lambda > 170$)

Abb. 7: Rectitude-Ethics



□ Glatte Komponente (zu Tiefpass): Bandpass ($200 > \lambda > 90$)
○ Glatte Komponente (zu linearem Trend): Bandpass ($\lambda > 90$)
△ Glatte Komponente (zu Tiefpass): Bandpass ($190 > \lambda > 170$)

Abb. 8: Power Gains

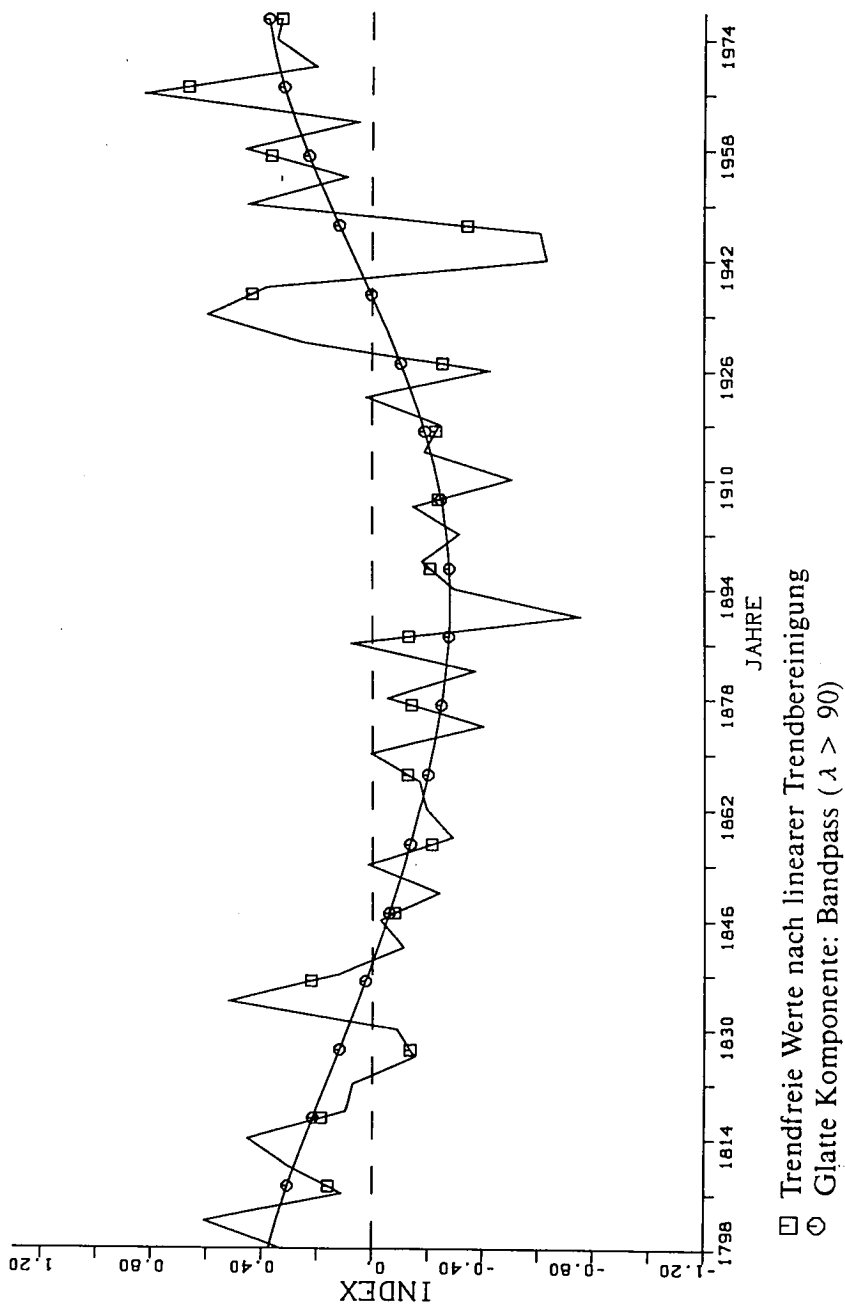
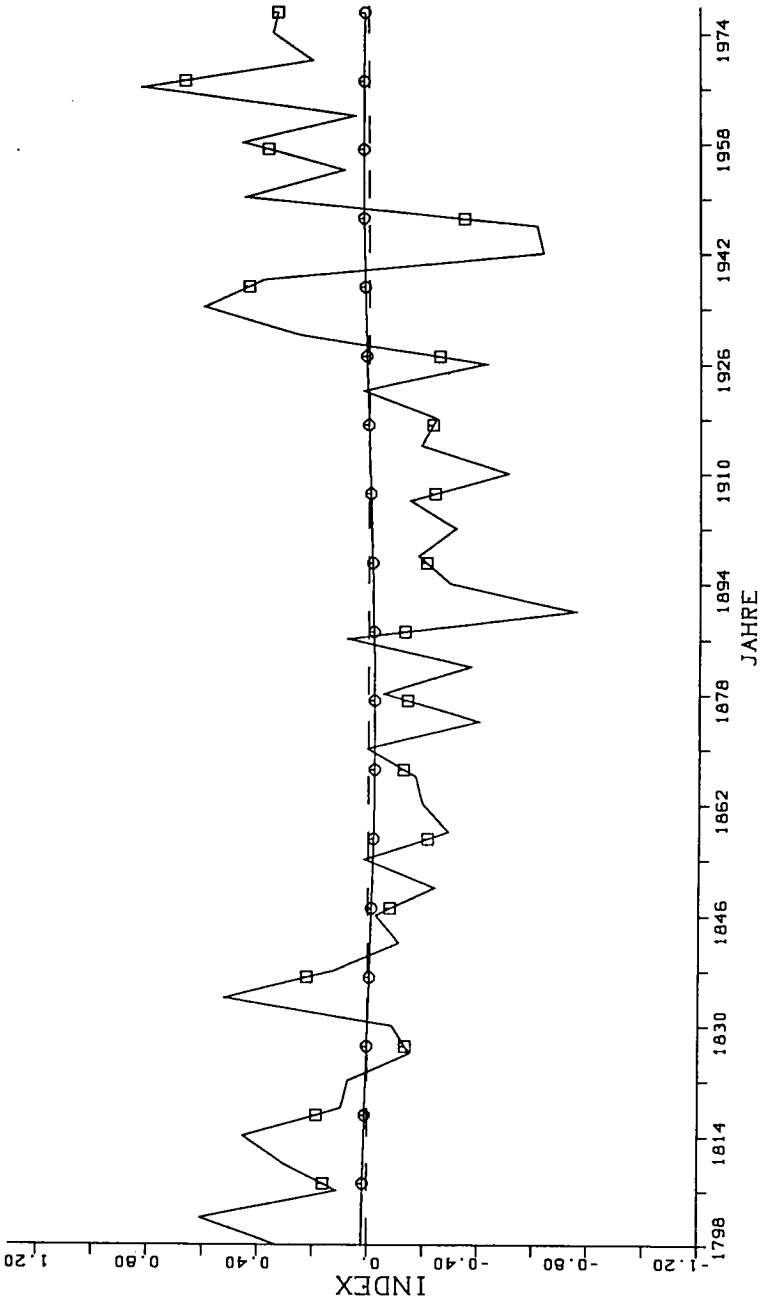


Abb. 9: Power Gains



□ Trendfreie Werte nach linearer Trendbereinigung
 ○ Glatte Komponente: Bandpass ($180 > \lambda > 160$)

riodenindikator, ergibt sich daraus eine Periodenlänge von $(1892-1810) \cdot 2 = 164$ Jahren, die nahe an Webers $\lambda = 176$ liegen (was nicht anders zu erwarten ist, da ja für den Bandpass-Filter der Frequenzbereich vorgegeben wurde, der Webers Sinuskurve entspricht). Abb. 7 faßt die nach den drei filtertheoretischen Varianten ermittelten Schwingungen in vergrößertem Maßstab noch einmal zusammen.

Als Ergebnis läßt sich also festhalten: Der von Weber für RECTITUDE ETHICS behauptete Zyklus läßt sich filtertheoretisch nicht bestätigen. Nach den ersten beiden Methoden lassen sich lediglich Schwingungen feststellen, die sowohl eine andere Länge als auch andere Hoch- und Tiefpunkte aufweisen. Nach der dritten Methode, die Webers Frequenzvorgabe folgt, ergibt sich gegenüber Webers Sinuskurve eine Phasenverschiebung, die Hoch- und Tiefpunkte nahezu vertauscht.

Welches Bild ergibt sich für die anderen neun Zeitreihen: Tab. 1 faßt die wichtigsten Ergebnisse zusammen, soweit sie aus Webers Methode sowie aus unseren Methoden 2 und 3, dem "weichesten" und dem "härtesten" Test, resultieren. Beide akzeptieren die lineare Trendbereinigung, Methode 3 läßt jedoch nur Wellenlängen zu, die Webers Frequenzvorgabe entsprechen. Die mit den filtertheoretischen Methoden identifizierten Schwingungen sind nicht deterministisch. Deshalb entfällt hierbei die Angabe des Parameters λ für die Periodenlänge. Man kann aber aus der zeitlichen Differenz (multipliziert mit 2) benachbarter Hoch- und Tiefpunkte auch für diese Schwingungen eine Längenangabe erhalten, die mit dem von Weber ermittelten Lambda-Parameter vergleichbar ist. Bei POWER AUTHORITY z. B. liegen Hoch- und Tiefpunkt nach Methode 2 100 Jahre auseinander, was einer Schwingungslänge von 200 Jahren entspräche (falls überhaupt eine Periodizität vorliegt, was bei lediglich einem Hoch- und Tiefpunkt weder bestätigt noch definitiv ausgeschlossen werden kann). Demgegenüber nennt Weber für diese Reihe eine Schwingsdauer von $\lambda = 144$. Wie Weber geben auch wir den quadrierten Pearson Produkt-Moment-Korrelationskoeffizienten r^2 als Maß der "Ähnlichkeit" zwischen ausgefilterten Komponenten und trendbereinigter Originalreihe an. Als Indikator für die Güte der Anpassung ist diese Maßzahl aber, wie bereits erwähnt, nur bedingt aussagefähig. Das ergibt sich daraus, daß harmonische Wellen mit gleicher Phase und Periode auch dann perfekt miteinander korrelieren, wenn ihre Amplituden völlig unterschiedlich sind. Ein niedriger Wert für r^2 enthüllt zwar eine geringe Ähnlichkeit zwischen (trendbereinigter) Originalreihe und angepaßter Reihe (vorausgesetzt die beiden Komponenten sind nicht phasenverzögert) ein hoher Koeffizient besagt aber nicht, daß ein

Tabelle 1: Vergleich der Analyseergebnisse

Zeitreihen	Weber			Filter 2					Filter 3				
	HW	λW	$r^2 W$	$H_1 M_2$	$H_2 M_2$	$T_1 M_2$	$T_2 M_2$	$r^2 M_2$	$H_1 M_3$	$H_2 M_3$	$T_1 M_3$	$T_2 M_3$	$r^2 M_3$
ENLIGHTENMENT TOTAL	1884	176	42	1881	--	1806	--	53	1888	--	--	--	53
POWER AUTHORITY	1972 1828	144	38	1874	--	1974	--	45	1920	--	1852	--	6
POWER PARTICIPATION	1970 1814	156	54	1877	--	1945	--	63	1874	--	1954	--	59
WEALTH PARTICIPATION	1848	128	32	1847	--	1919	--	44	1846	--	1913	--	43
WEALTH TOTAL	1956 1816	140	13	1824	1950	1886	--	16	1903	--	1829	--	15
WELL BEING TOTAL (*)	1829 1954	125	42	1834	1951	1892	--	19	1838	1949	1892	--	18
RECTITUDE ETHICS	1880	176	9	1901	--	1806	1953	20	1810	--	1892	--	9
RECTITUDE TOTAL	1880	208	19	1895	--	1951	--	16	--	--	--	--	3
POWER GAINS	1968 1804	164	36	--	--	1893	--	50	1955	--	1875	--	33
POWER TOTAL	1980 1788	192	20	--	--	--	--	35	--	--	1886	--	24
Erläuterung der Symbole													
"H" steht für "Hochpunkt", "T" für Tiefpunkt; die Indizes geben an, ob es sich um einen ersten oder zweiten Hoch- bzw. Tiefpunkt handelt.													
"W" steht für "nach Webers Methode" (1. Spalte); "M ₂ " und "M ₃ " (2. und 3. Spalte) bezeichnen unsere filtertheoretischen Methoden 2 und 3.													
r ² bezeichnet den Anteil der jeweils erklärten Varianz.													
"λ" bezeichnet die Periodenlänge.													
"(*)" markiert eine Gleitmittelreihe													

hoher Anteil an Varianz "erklärt" wird. Auf dieses Problem werden wir weiter unten zurückkommen.

In der folgenden Zusammenfassung berücksichtigen wir auch Ergebnisse, die wir mit unserer ersten Methodenvariante erzielt haben. (Sie unterscheidet sich, wie oben dargelegt, von der zweiten Variante nur durch eine andere Form der Trendbereinigung.) Außerdem stützen wir unsere Interpretation auf "Plots", wie wir das oben bei der Diskussion der Ergebnisse für RECTITUDE ETHICS schon gezeigt haben. Aus Platzgründen können hier jedoch nicht alle Plots reproduziert werden. Wir beginnen mit 4 Indikatoren, die Weber²⁷ dem Problemkomplex "Integration" zugeordnet hat:

²⁷ Siehe Namenwirth u. Weber (s. Fußn. 1), S. 106.

ENLIGHTENMENT-TOTAL: Die Hochpunkte stimmen bei allen vier methodischen Varianten in etwa überein. Filtertheoretisch ergeben sich jedoch keine oder nur schwache Hinweise auf eine mögliche Periodizität. Nach Methoden 2 und 3 erhält man eine Kurve, die wie die erste Hälfte einer periodischen Schwingung aussieht, dabei aber fast den gesamten Beobachtungszeitraum ausfüllt, also nicht einmal in einfacher Wellenlänge sichtbar wird. Nach Methode 2 erhält man neben dem Hochpunkt für 1881 einen Tiefpunkt für das Jahr 1806, was rein rechnerisch einer Periodenlänge von 150 Jahren entsprechen würde, die mit Webers $\lambda = 176$ in etwa kompatibel wäre.

POWER-AUTHORITY: Die nach den filtertheoretischen Methoden 1, 2 und 3 ermittelten Hochpunkte (1866, 1874, 1920) liegen weit von den Hochpunkten entfernt, die Weber mit 1972 und, implizit über $\lambda = 144$, mit 1828 angibt. (Für die Gleitmittelreihe hat er allerdings einen Hochpunkt von 1873 notiert.) Greift man nur die Schwingungen heraus, die seiner Sinuskurve entsprechen (Methode 3), verbleibt ein Varianzanteil von lediglich 6 %. Hinsichtlich der Periodizität der Schwingungen nach Methoden 1 und 2 gilt das gleiche wie für ENLIGHTENMENT-TOTAL.

POWER-PARTICIPANTS: Hier liegen die nach den Methoden 1, 2 und 3 bestimmten Hochpunkte (1875, 1877, 1874) nahe beieinander. Weber nennt mit 1970 bzw. (implizit durch $\lambda = 156$) 1814 Hochpunkte, die von diesen weit entfernt liegen. Allerdings notiert er für die Gleitmittelreihe mit 1874 auch hier wieder einen völlig anderen Hochpunkt. Mit keiner der drei Filtermethoden konnten entweder zwei Hoch- oder zwei Tiefpunkte festgestellt werden, so daß die Periodizität erneut äußerst unsicher ist.

WEALTH-PARTICIPANTS: Hier spielt die Trenddefinition bei der Bestimmung der Hochpunkte eine gewisse Rolle. Nach Methode 1 ist er für das Jahr 1829 anzusetzen, bei linearer Trendbereinigung (Methoden 2 und 3) für 1847 bzw. 1846, was Webers Hochpunktbestimmung (1848) entspricht. Sieht man von der Wiederholungsproblematik ab, wird, bei linearer Trendbestimmung, die Webersche Sinuskurve relativ gut bestätigt: Anzumerken ist jedoch, daß dieser Zyklus nach Phase und nach Periodenlänge nicht parallel zu den drei zuvor diskutierten Reihen verläuft, obwohl doch alle gemeinsam Indikatoren des "Integrative Theme" sein sollen. Auch die Phase von ENLIGHTENMENT TOTAL weicht erheblich (etwa 60 Jahre) von den Phasen der Reihen POWER AUTHORITY und POWER PARTICIPANTS ab.

Als nächstes betrachten wir zwei Reihen, die dem "Instrumental Theme" zugeordnet sind.

WEALTH-TOTAL: Wird der Trend nach Methode 1 eliminiert, ergeben sich innerhalb der Untersuchungsperiode zwei Hoch- und zwei Tiefpunkte, die aber nicht zyklisch miteinander verbunden sind. Bei linearer Trendbereinigung (Methode 2) nähert sich der Kurvenverlauf mit zwei Hochpunkten einer periodischen Form stärker an, ohne sie zu erreichen. Immerhin ist der Abstand der beiden Hochpunkte (1950 und 1824) mit der von Weber identifizierten Periodenlänge von 140 Jahren halbwegs vereinbar. Wenn man jedoch nur die Schwingungskomponenten herausgreift, die der Weberschen Sinuskurve für WEALTH-TOTAL entsprechen (Methode 3), ergibt sich für sie ein Hochpunkt im Jahre 1903, 50 Jahre von Webers Datierung entfernt. Sowohl nach Webers, als auch nach unserer Methode wird mit 13% bzw. 15% nur ein geringer Varianzanteil erklärt.

WELL-BEING TOTAL: Hier werden bei allen drei filtertheoretischen Methoden Schwingungen sichtbar, die, bei zwei Hochpunkten, einem periodischen Verlauf nahekommen. Die beiden Hochpunkte nach Methode 1 (1838, 1937) liegen 99 Jahre, die beiden Hochpunkte nach Methode 2 (1834, 1951) liegen 117 Jahre auseinander, während Weber (für die Gleitmittelreihe) eine Periodenlänge von 125 Jahren bei etwa gleichen Hochpunkten wie denen nach Methode 2 angibt. Sie stimmen auch relativ gut mit den Hochpunkten nach Methode 3 überein. Allerdings ist der hierfür ausgewiesene Varianzanteil von 18 % gegenüber Webers 42 % doch ziemlich gering.

Nachdem die Ergebnisse für RECTITUDE-ETHICS schon ausführlich diskutiert worden sind, kommen wir nun zum zweiten Indikator des "Expressive Theme".

RECTITUDE TOTAL: Die Hochpunkte nach Methoden 1 und 2 (1898, 1895) liegen in der Nähe des von Weber ermittelten Hochpunkts (1880). Eine zyklische Verlaufsform kann (nach diesen beiden Methoden) weder ausgeschlossen noch bestätigt werden. Doch indizieren die nach Methode 2 ermittelten Hoch- und Tiefpunkte (1895 und 1951) eine Schwingungslänge von 110 Jahren, die erheblich kürzer ist als die von Weber angegebene Periodenlänge von 208 Jahren. Schwingungskomponenten, die dieser Periodenlänge entsprechen würden (Methode 3), sind in der Reihe überhaupt nicht vorhanden; der entsprechende Bandpaßfilter reproduziert eine Gerade, die sich mit der Abszisse deckt.

Betrachten wir zum Schluß noch zwei Indikatoren, die Weber dem "Adaptive Theme" zurechnet.

POWER GAINS: Wenn wir die Schwingungskomponenten nach Methode 1 extrahieren, ergibt sich für sie ein Varianzanteil nahe Null. Lediglich die nach Methode 2 extrahierten Schwingungskomponenten erreichen einen nennenswerten Varianzanteil. Sie zeigen, mit einem Tiefpunkt in 1893, eine langgezogene Kurve, einem Wellental ähnlich, das nicht symmetrisch ist und offensichtlich einer langfristigen Bewegungsdynamik folgt, die innerhalb des Beobachtungszeitraums nicht zu identifizieren ist. Weber paßt an diese Reihe eine Sinuskurve mit $\lambda = 164$ Jahren an, deren Tiefpunkt (1886) nur geringfügig von 1893 abweicht - ein typisches Beispiel dafür, wie verführerisch die Ergebnisse des "Curve fitting" sein können: Ein in Wirklichkeit sehr langfristiger Trend (siehe Abb. 8) erscheint als Zyklus mit (relativ) kurzer Periodenlänge, der (laut Weber) innerhalb des Beobachtungszeitraums immerhin 36% der Varianz erklärt. Die nach Methode 3 herausgefilterte Schwingungskomponente, die der Frequenz der Weberschen Sinuskurve entspricht, ist zusammen mit der trendfreien Reihe in Abb. 9 dargestellt. Man sieht, daß sie sich kaum von der Nulllinie unterscheidet. Dennoch wird ein relativ hohes $r^2 = 0.33$ berechnet, weil die (kaum wahrnehmbaren) Auf- und Abschwünge in etwa mit denen der Originalreihe übereinstimmen. Wie oben bereits erläutert, berücksichtigt r^2 jedoch nicht das Ausmaß der Amplituden, die hier völlig unterschiedlich sind. Deshalb wäre es in diesem Falle irreführend, wegen $r^2 = 0.33$ von einer 33%igen Varianzaufklärung zu sprechen. Dies wird deutlich, wenn man Abb. 9 und Abb. 8 vergleicht. Offensichtlich erklärt die nach Methode 3 extrahierte Schwingungskomponente (s. Abb. 9) einen erheblich geringeren Anteil der Varianz der trendfreien Reihe als die nach Methode 2 extrahierten Komponenten (Abb. 8).

POWER-TOTAL: Bei linearer Trendbereinigung gemäß Methode 2 ist Periodizität auszuschließen. Bei der Trendbereinigung nach Methode 1 wird ein Schwingungsablauf erkennbar, der sich einer Periodenform nähert, doch die beiden Tiefpunkte (1830, 1927) sind mit der von Weber ermittelten Periodenlänge von 192 Jahren nicht vereinbar. Webers Angaben implizieren zudem einen weit davon entfernten Tiefpunkt im Jahre 1884. Die nach Methode 3 extrahierten Schwingungen zeigen keinen Hochpunkt, nur einen einzigen Tiefpunkt. Dessen Datierung (1866) stimmt mit der von Weber überein; die Periodizität ist aber bestenfalls unsicher.

4. Resümee

Fassen wir zusammen: Zyklen, deterministische zumal, die sich innerhalb der Beobachtungsperiode nicht wenigstens einmal wiederholen, können zwar theoretisch behauptet, empirisch aber mit keiner Methode belegt werden. Sinuskurven, die sich innerhalb des Beobachtungszeitraums nicht wiederholen, gar mit einer einzigen Welle über ihn hinausgehen, lassen sich an viele Zeitreihen anpassen, die keine zyklischen Schwingungen, sondern nur langfristige, nicht-periodische Niveauverschiebungen aufweisen. Evidenz gegen eine Zyklientheorie kann aber auch dann beigebracht werden, wenn die Beobachtungszeiträume für positive Belege zu kurz sind: Wenn der behauptete Zyklus tatsächlich vorliegt, müssen sich Schwingungen entsprechender Frequenz und mit nennenswertem Varianzanteil (nicht nur im Sinne von r^2) mittels filteranalytischer Methoden auch dann nachweisen lassen, wenn die Beobachtungsperiode relativ kurz ist.

Zur Überprüfung von Webers Zyklientheorie haben wir bei zehn seiner Zeitreihen drei verschiedene Testvarianten mit Hilfe der Filtertheorie durchgeführt. Der härteste Test ist mit der dritten Methodenvariante gegeben, die Schwingungen innerhalb eines relativ engen Frequenzbandes herausfiltert, das der von Weber jeweils identifizierten Sinuskurve entspricht. Die Übereinstimmung dieser Schwingungen mit Webers Sinuskurve ist nach vier Kriterien überprüfbar: (a) nach der Periodizität: wenn sie aus den genannten Gründen schon nicht beweisbar ist, läßt sie sich wenigstens nicht ausschließen? (b) nach der Datierung der Hoch- bzw. Tiefpunkte), (c) nach dem Varianzanteil an der trendbereinigten Reihe (unter Berücksichtigung der oben erwähnten Problematik). Außerdem dürfen sich (d) bei Anwendung der beiden anderen filtertheoretischen Varianten (Methoden 1 und 2) keine bezüglich der Hoch- und Tiefpunkte gegenläufigen Schwingungen zeigen.

Der "weicheste" Test ist mit der zweiten Methodenvariante gegeben, bei der zunächst der lineare Trend (als Funktion der Zeit) und anschließend alle Schwingungen mit einer Periode von $\lambda \leq 90$ Jahren aus der Reihe eliminiert wurden. Bei der ersten Variante ist diese Testsituation nur wenig verschärft, da immerhin der gesamte Frequenzbereich zwischen 90 und 200 Jahren zugelassen wird. Mit diesen beiden Varianten sind neben der Lage der Hoch- und Tiefpunkte auch die Periodenlängen der Weberschen Sinuskurven überprüfbar.

Die Kriterien von Test 3 werden am ehesten von zwei Zeitreihen, WEALTH PARTICIPANTS und WELL-BEING TOTAL, erfüllt. Einer dritten Reihe, POWER PARTICIPANTS, könnte man das Prädikat "Test be-

standen" zuerkennen, wenn man Webers Sinuskurve für die Gleitmittelwerte, nicht diejenige für die aggregierten Daten zugrunde legte. Allerdings ist die Periodizität in diesem Falle sehr unsicher. Eine zyklenähnliche Bewegung im Sinne Webers läßt sich auch für ENLIGHTENMENT TOTAL nicht unbedingt ausschließen. Da es aber innerhalb der Untersuchungsperiode nach Methode 3 nur einen Hochpunkt und keinen Tiefpunkt gibt, ist die Periodizität noch unsicherer als im Falle von POWER PARTICIPANTS.

Wenn man nur die Schwingungen betrachtet, die mit den beiden ersten Methodenvarianten extrahiert wurden, so zeigt noch WEALTH TOTAL (nach linearer Trendbereinigung) einen zyklenähnlichen Verlauf, der mit Webers Sinuskurve vergleichbar ist, ohne allerdings einer strengen Periode zu folgen. Außerdem verlaufen diese Schwingungen gegenläufig zu der (allerdings schwachen) Schwingung, die mit der dritten Methodenvariante herausgefiltert wurde.

Bei den übrigen fünf Reihen ist die Evidenz bezüglich der Weberschen Sinuskurven nach allen Testvarianten in mindestens einem der og. Kriterien negativ.

Die Bilanz unserer Ergebnisse spricht also gegen Webers Theorie. Fünf eindeutigen Widerlegungen stehen drei nicht entscheidbare Fälle und zwei "Bestätigungen" gegenüber. Dabei sind die Bestätigungen noch insofern großzügig als auch in diesen Reihen keine Wiederholung des Zyklus innerhalb der Untersuchungsperiode beobachtet werden kann. Zyklen in der Form und mit der theoretischen Bedeutung, die Weber ihnen zumißt, sind in dessen eigenen Daten mit einer angemessenen Methode nicht nachweisbar.

Sicherlich wäre es wünschenswert, auch noch die anderen der von Weber konstruierten Zeitreihen in gleicher Weise zu überprüfen. Die Evidenzen dürften sich jedoch kaum zugunsten der Weberschen Theorie verschieben. Für diese Vermutung spricht, daß die zehn von uns untersuchten Zeitreihen ohne Kenntnis der Ergebnisse ausgewählt wurden. Außerdem wird für die Sinuskurven, die Weber für diese zehn Reihen identifiziert hat, ein durchschnittlicher Varianzanteil von 30.5 % an der Gesamtvarianz der trendfreien Reihen angegeben; der Durchschnitt für alle Reihen beträgt 24 %, so daß unsere Auswahl in dieser Hinsicht nicht unfair ist.

Natürlich haben wir mit diesen Ergebnissen nicht die allgemeine These widerlegt, die kulturelle Entwicklung weise zyklische oder zyklenähnliche Komponenten auf. Als nicht haltbar betrachten wir aber Webers These, die langfristige kulturelle Entwicklung vollziehe sich generell in Form deterministischer Zyklen oder lasse sich mit solchen Zyklen vereinfachend, aber ange-

messen repräsentieren. Uns scheint die Annahme deterministischer Zyklen theoretisch nicht sinnvoll und methodisch unfruchtbar. Zeitreihen sollten als stochastische Prozesse aufgefaßt werden, die nicht nur eine innere Dynamik aufweisen, sondern auch von externen Ereignissen (und deren kollektiver Verarbeitung) vorangetrieben werden.

TEIL. III:

Computereinsatz in der Textanalyse

Der General Inquirer III Ein Dinosaurier mit Zukunft

von Cornelia Züll und Peter Ph. Mohler

1.0 Einleitung

Der General Inquirer wurde zu Anfang der sechziger Jahre von Philip J. Stone als inhaltsanalytisches Pendant zu Herbert Simons "General Problem Solver" entwickelt. Er entstand damals in engem Zusammenhang mit Entwicklungen zur maschinellen Übersetzung. Dementsprechend wurde zusammen mit den Programmen zugleich an der Entwicklung von Wörterbüchern (Diktionären) gearbeitet, mit deren Hilfe die gesellschaftliche Wirklichkeit vollautomatisch in sozialwissenschaftliche Begrifflichkeit übersetzt werden sollte. Auch wenn sich in der Folgezeit dieser hohe Anspruch nicht ganz hat einlösen lassen, so blieb der General Inquirer das einzige Programm für die computerunterstützte Inhaltsanalyse (cui), zu dem es große, auch in historischen Studien bewährte, Klassifikationssysteme gibt. In diesem Beitrag werden vor allem diese Klassifikationssysteme und die Programme erläutert. Daneben wird auch das Ziel verfolgt, dem Leser einen ersten Überblick über diesen komplexen textanalytischen Ansatz zu verschaffen.

2.0 Abriß der Entwicklung des General Inquirers

Eine erste größere Publikation zum General Inquirer erschien bereits 1966¹. In dieser ersten, grundlegenden Publikation wurde der theoretische Hintergrund ausführlich dargestellt, insbesondere wurde der Begriff der Inhaltsanalyse schärfer als eine Methode der Inferenz von Texten auf soziale Wirklichkeiten definiert. Einen großen Teil des Buches nahmen aber schon Berichte zu Anwendungen des General Inquirers in der praktischen Forschung ein. Diese Anwendungen streuten über alle Gebiete der Humanwissenschaften, von der Psychologie über die Soziologie bis zur Geschichtswissenschaft.

Obwohl nach dieser ersten Publikation weiter am General Inquirer entwickelt wurde - es gab neben der Harvard Version noch eine Edinburgh Version, die Diktionäre wurden weiter validiert und ausgebaut - stagnierte die

¹ Stone, P. J. , D. D. Dunphy, M. S. Smith, and D. M. Ogilvie (eds.), 1966. The General Inquirer: A Computer Approach to Content Analysis. Cambridge: MIT Press.

Entwicklung, weil es nicht gelang, eine stabile, gut dokumentierte Standardversion der wissenschaftlichen Öffentlichkeit zur Verfügung zu stellen. Die Anwendung des General Inquirers erforderte zudem auch gute programmtechnische Kenntnisse, was seiner Verbreitung in den Geisteswissenschaften nicht förderlich war.

Um 1976 schien es dann einen neuen Aufschwung zu geben, als Kelly und Stone ihr Verfahren zur Disambiguierung von Homographen publizierten und damit eines der schwierigsten Probleme der cui wenigstens für die häufigsten amerikanisch-englischen Wörter lösten². Aber auch danach gelang es aus verschiedensten Gründen nicht, eine Standardversion zu entwickeln.

Trotz dieser äußerst ungünstigen Lage, flaute die Nachfrage nach dem General Inquirer bei Philip Stone nicht ab, sie stieg eher im Laufe der achtziger Jahre an. ZUMA, das schon zuvor mit TEXTPACK das erste voll portable cui-Programm entwickelt hatte, erklärte sich auf Grund der bestehenden Nachfrage und des großen analytischen Potentials dann 1986 bereit, zusammen mit Robert Philip Weber (Harvard) eine Standardversion für IBM Computer zu entwickeln. Weber hatte zuvor den General Inquirer in einer umfangreichen historischen Untersuchung³ eingesetzt und für solche Zwecke adaptiert. Die Entwicklung der Standardversion ist jetzt faktisch abgeschlossen, eine ausführliche Dokumentation liegt vor⁴ und das Programm wird seit Anfang 1989 an interessierte wissenschaftliche Institute ausgeliefert.

2.1 Der theoretische Ansatz des General Inquirers

Wie schon erwähnt stand am Anfang der Entwicklung die Idee, Texte in eine Metasprache der Sozialwissenschaften zu übersetzen. Heute würde man eher sagen, daß mit Hilfe des General Inquirers semantische Strukturen, Bedeutungen, eindeutig abgebildet werden sollen. Die Abbildung ist nicht als spezifische, womöglich nur auf eine Dimension eingeschränkte Abbildung, gedacht, sondern als eine allgemeine und mehrdimensionale. Mit "allgemein" ist gemeint, daß unterschiedlichste Forschungsansätze und Fragestellungen ein und dasselbe Klassifikationsverfahren benutzen. Damit soll eine Basis für

² Kelly, E. F., and P. J. Stone, 1975. *Computer Recognition of English Word Senses*. Amsterdam: North Holland.

³ Namenwirth, Z. J., and R. P. Weber, 1987. *The Dynamics of Culture*. Boston: Allen & Unwin.

⁴ Züll, C., R. P. Weber, and P. Ph. Mohler, 1989. *Computer-aided Text Classification for the Social Science: The General Inquirer III*. Mannheim: ZUMA.

interdisziplinäre, übergreifende und kumulative Forschung gelegt werden. Der General Inquirer ist von daher ganz bewußt auf vergleichende Forschung angelegt. Seine Klassifikationsschemata, die als Wörterbücher (Diktionäre) realisiert wurden, sind deshalb unter eher allgemeinen, denn spezifischen theoretischen Annahmen entwickelt worden (siehe 4.0). "Deduktiv entwickelt" heißt, daß die Einträge im Diktionär (Wörter) unabhängig von einer spezifischen Fragestellung allein aus theoretische Erwägungen jeweils einer oder mehreren Klassen (Kategorien) zugewiesen wurden. Dieser sogenannte a priori Ansatz der cui ist in der Literatur nicht unumstritten⁵. Trotz aller gegen den a priori Ansatz vorgebrachten Einwände muß man zugestehen, daß er, richtig eingesetzt, das Instrument der Wahl für groß angelegte vergleichende Studien ist.

2.2 Der General Inquirer in der historischen Sozialforschung

Von Anfang an wurde der General Inquirer zur Analyse historischer Texte und zur Beantwortung historischer Fragestellungen eingesetzt (vgl. diverse Beiträge in Stone et al. 1966). Die bislang bekannteste Untersuchung auf dem Gebiet der historischen Sozialforschung geht auf Zvi Namenwirth zurück, der mit einer systematischen Analyse der amerikanischen "Party Platforms" über 120 Jahre hinweg (1844-1964) zyklische Bewegungen der gesellschaftlichen Entwicklung nachweisen konnte. Robert P. Weber konnte die von Namenwirth postulierten Zyklen auch in Großbritannien an Hand von Thronreden (1689- 1974) nachweisen. In "Dynamics of Culture" (Namenwirth & Weber, 1987) haben beide ihre Ergebnisse nach theoretischen Grundlagen stringend und überzeugend zusammengefaßt.

Beide Studien weisen auf die Stärke des General Inquirers, der vorzüglich für die Analyse bekannter Klassen historischer Texte geeignet ist. Wenn erst einmal durch geeignete Vorstudien die Struktur einer Textklasse in bestimmten historischen Kontexten herausgearbeitet ist, kann mit dem a priori Ansatz des General Inquirers eine große Textmenge valide und zuverlässig klassifiziert werden. Zusätzlich erlaubt die allgemeine, mehrdimensionale Abbildung des General Inquirers einen Vergleich zwischen unterschiedlichen Textklassen. So könnte man sich einen Vergleich von Webers Analysen der Thronreden mit Parlamentsdebatten oder einen Vergleich von Testa-

⁵ Mohler, Peter Ph. 1989. "Computergestützte Inhaltsanalyse: Überblick über die linguistischen Leistungen", in: Lenders, W. et al. (Hg.). Handbook of Computational Linguistics. Berlin: de Gruyter.

menten mit Gesetzestexten vorstellen. Der General Inquirer Ansatz könnte also wesentlich zur Systematisierung in der historischen Sozialforschung beitragen.⁶

3.0 Wörterbücher im General Inquirer

Die Wörterbücher des General Inquirers enthalten neben den Wötereinträgen und den dazugehörigen Kategorien zusätzlich Regeln zur Disambiguierung (d.h. Erkennung von Homographen) der Wörter und zum Bestimmen von Redewendungen (z.B. "ship of state" und "Secretary of State"). Im Moment stehen drei große Wörterbücher, die zusammen mit dem General Inquirer weitergegeben werden, zur Verfügung. Diese drei Diktionäre - Lasswell Value, Harvard IV-3 und Harvard IV-4 - werden im Abschnitt 4 beschrieben.

Das Erstellen und Validieren von inhaltsanalytischen Diktionären ist im allgemeinen ein sehr aufwendiges Unterfangen. Trotzdem wird es häufig erforderlich sein, die sehr allgemein gehaltenen Diktionäre zu modifizieren und vor allem, der eigenen Fragestellung angepaßt, zu erweitern. Um Diktionäre zu erstellen, zu modifizieren und zu erweitern steht innerhalb des General Inquirers eine eigene Sprache zur Verfügung, in der alle Einträge und Regeln abgefaßt sein müssen.

Die Disambiguierung und Vercodung erfolgt auf Satzebene in drei Stufen: zunächst wird das Wort selbst überprüft, danach werden im selben Satz andere, spezifische Wörter gesucht und als letztes werden Kategorien überprüft, die anderen Wörtern im Satz bereits zugewiesen wurden. Danach kann die Bedeutung des Worts in den meisten Fällen bestimmt werden. Im Folgenden soll kurz auf den Aufbau der Wörterbücher und der Regeln im Wörterbuch eingegangen und die wesentlichen Merkmale aufgezeigt werden. Tabelle 3.0.1 zeigt am Beispiel des Wortes "LAST" den Aufbau eines Wörterbucheintrags.

⁶ Die General Inquirer Diktionäre liegen als vollentwickeltes Klassifikationssystem zur Zeit nur für englische Texte vor. Manuel Eisner versucht im Moment eine deutschsprachige Version des Lasswell Value Diktionärs (LVD) zu entwickeln. Der a priori Ansatz ist zudem nicht an den General Inquirer gebunden, er kann auch mit anderen Systemen, z. B. der LVD-Lib oder TEXTPACK V realisiert werden.

Tabelle 3.0.1: Wörterbucheintrag zu "LAST"

- LAST=TAGS: (1) DET.NUMB.ORD.MODIF.TIMESP.'ADJ "PREVIOUS, PAST, PRIOR, MOST RECENT '
 (2) DET.NUMB.ORD.POS.MODIF.TIMESP.'ADJ-ADV "FINAL, FINALLY '
 (3) LY.TIMESP.'IDIOM-ADV ""AT . . .," "AT LONG LAST"-- FINALLY '
 (4) SUPV.ENDS.'VERB "ENDURE, REMAIN '
 (5) MODIF.UNDEF*.'ADJ ""LASTING"--ENDURING, REMAINING'

- RULES: (7) TOR(K+0,K+0,,27,ROOT.)
 (8) WOR(K+1,K+1,,12,TIME.FALL.)
 (9) WOR(K-I,K-I,,APLY(I),THE.)
 (10) TOR(K-2,K-2,APLY(2),APLY(1),BE.)
 (12) TOR(K-1,K-1,,22,DET.)
 (13) TOR(K+1,K+3,,17,TIME.)
 (14) WOR(K+1,K+1,APLY(2),,DAY.MINUTE.MOMENT. SECOND.)
 (15) TOR(K-1,K-1,APLY(2),APLY(1),GEN.)
 (17) TORK(K-5,K+5,APLY(1),,POLIT.)
 (18) TORK(K-6,K+6,,APLY(2),CARD.)
 (19) WOR(K+1,K+1,APLY(2),APLY(1),ONE.)
 (22) WOR(K-1,K-1,DELID(3),,AT.LONG.)
 (23) TORK(K-I,K+I,APLY(I),,TIME.)
 (24) TOR(K-1,K-1,APLY(4),,TO.MOD.DO.LY.INDEF.)
 (25) TOR(K+1,K+1,APLY(4),APLY(2),DET.PREP.)
 (27) TOR(K+0,K+0,APLY(2),,LY.)
 (28) TOR(K+0,K+0,,APLY(4),ING.)
 (29) TOR(K+1,K+1,APLY(4),APLY(5),DET.PREP.)

END;

Jeder Eintrag im Wörterbuch besteht aus zwei Teilen: einer Liste aller möglichen Bedeutungen des Worts (TAGS, Bedeutungscode), gefolgt von Regeln (RULES, formale Disambiguierungsregeln), mit denen das Wort der richtigen Kategorie zugeordnet werden kann.

Im ersten Teil werden jeweils alle möglichen Bedeutungen eines Worts aufgelistet und mit Bedeutungs-codes versehen. Im obigen Beispiel hat das Wort LAST als Verb z.B. den Code 4. Nach dem jeweiligen Code werden alle Kategorien (syntaktische und semantische) aufgeführt, die dem Wort zugewiesen werden sollen, wenn es in dieser Bedeutung vorkommt.

Sind alle Bedeutungen definiert, folgen die Regeln für die Zuordnung der richtigen Bedeutung zu einem Wort. Ist das Wort immer eindeutig, entfällt dieser zweite Teil des Wörterbucheintrags. Die Regeln werden, wie vorher schon die verschiedenen Bedeutungen, in Form einer Liste angegeben.

Es gibt drei verschiedene Hilfen zum Testen: die GOTO Funktion, die SUPV-Funktion, die auf syntaktischer Ebene testet, und den normalen Kontext-Test, wie er unten beschrieben wird. Mit GOTO können Regeln und Tests übersprungen, bzw. Regeln gezielt angesteuert werden, mit SUPV kann festgestellt werden, ob es sich um eine Verb-Form handelt oder nicht und in Abhängigkeit davon zur nächsten Regel übergegangen werden. Die Kontext-Tests sind immer in der Form

test name(r-start,r-end,b-success,b-fail,items)

aufgebaut. Dabei definiert das erste Feld (test name) die Art der Suche. Es gibt die verschiedensten spezifischen Funktionen zu suchen, auf die hier nicht alle eingegangen werden kann. Die wichtigsten sind WOR, TOR, TSAME, WORK und TORK. WOR und WORK testen, ob die unter "items" als letztem Parameter angegebenen Wörter im definierten Bereich vorkommen oder nicht. Bei WORK ist das Schlüsselwort selbst vom Test ausgenommen, bei WOR dagegen wird es in den Test einbezogen. Bei TOR und TORK werden anstelle des Worts, die in der Itemliste angegebenen Codes mit den für das Wort (oder die Wörter) im angegebenen Bereich vergebenen Codes verglichen. TSAME prüft, ob dem Wort selbst schon ein bestimmter Code zugewiesen wurde.

Der Bereich, der überprüft wird, wird durch r-start und r-end in der oben beschriebenen Regel festgelegt. Die Angaben beziehen sich auf die Wortpositionen, die geprüft werden sollen. Ist r-start,r-end z.B. mit K+1,K+1 angegeben, so wird das Wort, das auf das gesuchte Wort folgt, getestet; ist K+5,K+5 angegeben, so werden die 5 Wörter vor und 5 Wörter nach dem gesuchten Wort in die Überprüfung miteinbezogen.

b-success legt fest, was geschehen soll, wenn die Bedingung erfüllt ist, b-fail entsprechend, was bei nicht-erfüllten Bedingungen geschehen soll. Dabei können als Ergebnis mit der Funktion APLY die gefundenen Codes zugewiesen werden. Es kann aber auch zu einer weiteren Regel gesprungen werden oder, wenn nichts angegeben ist, mit der nächsten Regel weitergetestet werden. Eine letzte Möglichkeit - die Funktion DELID - erlaubt die Behandlung von Idiomen oder idiomatischen-Redewendungen, d.h. es wird geprüft, ob ein Wort Teil einer Redewendung ist. Wenn ja, werden alle bereits zugewiesenen Codes wieder gelöscht und stattdessen der besondere Code für das Idiom zugewiesen. Im obigen Beispiel wird dies in der Regel 22 gemacht. Geprüft wird, ob es sich um AT LONG LAST handelt. In diesem Fall erhält das Wort den Code 3. Ist das nicht zutreffend, wird mit der nächsten Regel weitergetestet.

Im einzelnen läuft die Prüfung der Regeln für das Wort LAST folgendermaßen ab: zunächst wird geprüft, ob es sich bei dem Wort um den Wortstamm oder um eine andere Form (z.B. LASTING) handelt. Wenn nein, wird zur Regel 27 gesprungen, wenn ja, wird mit Hilfe der nächsten Regel getestet. Dort wird nun geprüft, ob auf LAST die Wörter TIME oder FALL unmittelbar folgen ($K+1, K+1$; "last time" oder "last fall"). Wenn ja, geht es weiter mit Regel 12, wenn nein, wird die nächste Regel geprüft. Nun wird geprüft, ob vor LAST das Wort THE steht ($K-1, K-1$; "the last"). Wenn nein, wird der Code 1 zugewiesen, das Wort als Adjektiv eingestuft, und es erhält folgende syntaktische und semantische Codes: DET, NUMB, ORD, MODIF, TIMESP. Wenn diese Bedingung nicht erfüllt ist, wird die nächste Regel angewandt. So werden alle Regeln nacheinander abgearbeitet, bis die Bedeutung des Worts feststeht.

4.0 Die verfügbaren Diktionäre zum General Inquirer: Harvard IV-3 und Harvard IV-4 Diktionär, Lasswell Value Diktionär

Der General Inquirer wird zusammen mit drei Diktionären, dem Harvard (Version IV-3 und IV-4) sowie dem Lasswell Value Diktionär ausgeliefert. Im folgenden soll kurz auf die theoretischen Ansätze und die allgemeine Struktur dieser Diktionäre eingegangen werden. Für eine ausführlichere Erörterung wird auf Züll et al. verwiesen.

4.1 Harvard IV-3 und IV-4

Die Harvard Diktionäre gehen auf eine Idee von Philip J. Stone zurück, der in einem umfassenden Klassifikationssystem psychologische und soziologische Ansätze zusammenfügte. Dies spiegelt sich auch im vollen Namen dieser Diktionäre wieder, der "Harvard Psycho-Sociological Dictionary" lautet. Dabei fanden aus der Psychologie insbesondere Vorstellungen von Freud und aus der Soziologie solche von Parsons Eingang in die Kategorienkonstruktion. Version IV-3 ist Stones Weiterentwicklung der im General Inquirer von 1966 beschriebenen Version III (Stone et al. 1966). Insbesondere hat er in der neuen Version versucht, vermehrt syntaktisch "reine" Kategorien zu bilden, indem er systematisch zwischen Nomen, Verben und Modalen unterscheidet. Dadurch erhält man zusätzlich zur Bedeutung auch noch Hinweise auf die Statik (Nomen) oder Prozesshaftigkeit (Verben) in den untersuchten Texten. Weiterhin wurde versucht, die Kategorien als Maßindikatoren mittels verschiedener Maßnahmen robuster als früher zu machen. So konnte es früher der Fall sein, daß eine Kategorie nur dann valide maß, wenn mehrere ihrer Wörter in einem Text auftraten. Trat dagegen von einer Kategorie in einem Text immer nur ein Wort auf, konnten erhebliche Verzerrungen entstehen (z.B. in der Kategorie ACTION NORM, wenn in einem Text nur über BREAKFAST gesprochen wurde, dann war dies eher ein Indikator für "ORAL" denn für ACTION NORM).

Entsprechend seiner allgemeinen Zielsetzung spannen die Kategorien ein weites Feld zwischen "Animal Noun", über "Hostile" bis zu "Work" auf. Zusätzlich zu der syntaktischen Trennung in Nomen, Verben und Modale, den semantischen Hauptkategorien enthält der Harvard IV-3 eine Reihe von Hilfskategorien (Markerkategorien), wie z.B. "CONJ" für Konjunktionen, "QUOTE" für Anführungszeichen usw., die u.a. auch für die Disambiguierung benötigt werden. Die Version IV-4 ist eine vollständig revidierte Fassung, die von Dexter C. Dunphy, Cederik G. Bullard und Elinor R.M. Crossing (University of New South Wales/Australien) entwickelt wurde. Dunphy et al.⁷ überprüften zuerst die hierarchische Ordnung des Klassifikationssystems und validierten dann in einem aufwendigen Verfahren die einzelnen Kategorien. Ihre Version des Harvard Diktionärs enthält 65 Markerkategorien als Hilfskategorien für die Disambiguierung, 79 "First Order" Kategorien, die den semantischen Kern ausmachen, und 25 weitere "Second Order" Kategorien, die zusätzliche Dimensionierungen der Analyse erlauben.

⁷ Dunphy, D. C., C. G. Bullard, and E. E. M. Crossing. 1974. Validation of the General Inquirer Harvard IV Dictionary. Paper presented at 1974 Pisa Conference on Content Analysis.

4.2 Lasswell Value Diktionär

Der Lasswell Value Diktionär (LVD) geht auf eine Idee von Zvi Namenwirth zurück, der damit eine Umsetzung der Theorien von Lasswell und Kaplan⁸ in ein inhaltsanalytisches Klassifikationsschema anstrebte. In neueren Veröffentlichungen wird dieser Anspruch eher zugunsten einer Definition als allgemeines Klassifikationsschema denn als direkte Operationalisierung bestimmter theoretischer Überlegungen zurückgenommen (Namenwirth & Weber, 1987). Dennoch zeigt sich in der Organisation auch weiterhin der Bezug zu Lasswell und Kaplans politisch-soziologischem Ansatz. So wird von acht Hauptkategorien ausgegangen, die zu zwei Gruppen zusammengefaßt werden können: die "Deference Values" (POWER, RECTITUDE, RESPECT, AFFECTION), die als Indikatoren für Interaktionen zwischen Handelnden genommen werden, und die "Welfare Values" (WEALTH, WELL BEING, ENLIGHTENMENT, SKILL), die als Indikatoren für die Aktivitätserhaltung (notwendige Ressourcen der Individuen) genommen werden. Zu diesen acht Hauptkategorien treten, ähnlich den Second Order Kategorien des Harvard IV-4, weitere Nebenkategorien, wie die Gruppe der "general value transaction indicators" (z.B. transaction gains, transaction losses), einer Kategorie "ANOMIE", Indikatoren für Gefühle (Affect), und Indikatoren für Raum und Zeit. Dazu kommen noch die Markerkategorien für die Disambiguierung.

5.0 Programme des General Inquirers

Die Programme des General Inquirer lassen sich in drei Hauptgruppen unterteilen:

1. Programme, die die Texte aufbereiten und bearbeiten (Disambiguieren und Vercoden);
2. Programme, die auf Grund der vorher erstellten Klassifikationen Häufigkeiten erstellen, die Vercodungen im Satzzusammenhang darstellen und die Vercodungen für die weitere Analyse aufbereiten;
3. Programme, die als Hilfsmittel zur Wörterbucherstellung und deren Überprüfung dienen.

Zur ersten Gruppe gehören die Programme TEXTREAD und TAGGER, zur zweiten TALLY, RETRIEVE und CONNECT, und die dritte setzt sich aus den Programmen PARSER, DICTMERG, PEEL und DOCUMENT zusammen.

⁸ Lasswell, H. D., and A. Kaplan. 1963 (1950). *Power and Society: A Framework for Political Inquiry*. New Haven: Yale University Press.

5.1 Programm TEXTREAD

TEXTREAD liest den Eingabetext des Anwenders und erstellt daraus einen internen "Systemfile", der für die Disambiguierung und Vercodung mit TAGGER erforderlich ist. TEXTREAD hat dabei zwei wesentliche Aufgaben. Zum einen zerlegt es den Text abhängig von Satzzeichen in einzelne Sätze. Die Sätze bilden dann die Basis für die weitere Vercodung, d.h. der General Inquirer arbeitet dabei immer auf Satzebene. Zum anderen werden von TEXTREAD bereits Kategorien für Satzzeichen und Zahlen vergeben und Abkürzungen überprüft (z.B. wird Mr. als Abkürzung erkannt, gekennzeichnet und entsprechend vercodet).

Der General Inquirer kann alle (und nur) englischsprachige Texte verarbeiten. Bei der Verschriftung bzw. Textvorbereitung müssen dabei nur einige Grundregeln beachtet werden:

- Die Texte können in Groß-/Kleinschreibung verschriftet sein. Zur Vercodung werden sie jedoch immer in Großbuchstaben umgesetzt, da alle Regeln in den zur Verfügung stehenden Wörterbüchern nur in Großschreibung gespeichert sind. Diese Einschränkung stellt bei englischsprachigen Texten (im Gegensatz zu deutschen) keinen Informationsverlust dar.

- Sätze dürfen maximal 2000 Zeichen lang sein, Wörter 20 Zeichen. Längere Sätze werden vom Programm automatisch in kürzere zerlegt, längere Wörter werden abgeschnitten. Das Programm druckt in solchen Fällen eine entsprechende Warnung.

- Der Text kann durch Identifikatoren in logische Blöcke unterteilt werden, die aber auf die Analyse keinen Einfluß haben, sondern nur der Dokumentation dienen. Dasselbe gilt für Kommentare, die in den Text eingefügt werden können. Während die Identifikatoren als Dokumentation auch für TAGGER und die nachfolgenden Programme erhalten bleiben, werden die Kommentare im Text von TEXTREAD ausgedruckt, danach aber aus dem Text entfernt.

5.2 Programm TAGGER

Nachdem TEXTREAD den Eingabetext aufbereitet hat, übernimmt TAGGER die entgeltliche Disambiguierung und Vercodung. TAGGER vercodet auf Grund der Regel eines Wörterbuchs (siehe oben). Wie diese Regeln aufgebaut sind und wie die Vercodung erfolgt, wurde im Kapitel 3.0 ausführlicher beschrieben. Jedem Wort werden ein oder mehrere syntaktische und

semantische Codes zugewiesen. Bei nicht eindeutigen Wörtern wird versucht mit Hilfe des Satzzusammenhangs, in dem sie stehen, festzustellen, welche Bedeutung sie an dieser Stelle haben (Disambiguierung). Diese Wörter werden in der Ausgabe in der Form Wortstamm, Bedeutungscode und gegebenenfalls Endung (z.B. ARMS wird ARM1S) gespeichert. Die nachfolgende Tabelle zeigt anhand eines Satzes die Vercodung durch TAGGER. Sie ist ein Auszug aus der Druckerausgabe von TAGGER. In der letzten Spalte sind alle zugewiesenen syntaktischen und semantischen Kategorien aufgeführt.

Tabelle 5.2.1: Vercodung durch TAGGER zu dem Satz "Systems of thought have been jarred, ways of life have been uprooted, institutions are under siege."

```

****TRACE****SENTENCE 10 *****DOCUMENT 1 *****IDENTIFICATION AD1968
1: SYSTEM      S  0  1  NOUN S B ABS ABS* THINK KNOW ORDER
2: OF           0  1  PREP ROOT
3: THOUGHT     4  4  1  NOUN ED ABS ABS* THINK KNOW ACTVI
4: HAVE        3  4  1  SUPV ROOT VERB HAVE
5: BEEN        3  2  1  SUPV ROOT VERB BE ED PASSIVE
6: JAR         2 RED 4 1  SUPV ED MOVE EXERT NGTV) ACTV2 STRNG2
7: ,           0  1  PUNC COMMA
8: WAY         1 S  3  1  NOUN S THINK EVALU MEANS PFREQ
9: OF           0  1  PREP ROOT
10: LIFE        0  1  NOUN ROOT NATURE
11: HAVE       3  4  1  SUPV ROOT VERB HAVE
12: BEEN       3  2  1  SUPV ROOT VERB BE ED PASSIVE
13: UPROOT     ED 0  1  ED X
14: ,           0  1  PUNC COMMA
15: INSTITUTION S 0  1  NOUN S PLACE SOCIAL POLIT ECON STRNG2
16: ARE        1  2  1  SUPV ROOT VERB BE
17: UNDER     1  1  1  PREP ROOT POWER -
18: SIEGE      0  1  E X ROOT
19: .          0  1  PUNC PER

```

Alle Wörter, die nicht im Wörterbuch definiert sind und nicht vercodet werden können, werden in einer Leftover-Liste gedruckt und in einer Datei (als Wortstamm) abgelegt. Sie müssen, sofern sie für die Analyse von Bedeutung sind, nachträglich ins Wörterbuch aufgenommen oder anderweitig nachvercodet werden.

Am Ende der Vercodung erhält der Anwender einige Statistiken zum eingesetzten Wörterbuch und zum Text wie z.B. durchschnittliche Länge der Regeln im Wörterbuch, Zahl der Sätze im Text, Anzahl aller Wörter und Zahl der Satzzeichen.

5.3 Programm RETRIEVE

Nachdem die Vercodung beendet ist, stehen dem Anwender einige Möglichkeiten der Überprüfung der Vercodung und der Weiterverarbeitung zur Verfügung. Eines dieser Programme ist RETRIEVE, das dem Anwender die Möglichkeit gibt, Sätze auf Basis ausgewählter Codes auszudrucken. Eingabe in RETRIEVE ist die von TAGGER erstellte Ausgabedatei. Tritt ein vorgegebener Code in einem Satz auf, wird der ganze Satz gedruckt. Dabei sind alle dieser Kategorie zugeordneten Wörter unterstrichen. Bei großen Texten kann eine Satzauswahl der Form "jeder n-te Satz wird gedruckt" getroffen werden.

RETRIEVE erlaubt drei verschiedene Auswahlmöglichkeiten: Nach der ersten werden alle Sätze gedruckt, in denen die vom Anwender vorgegebene Kategorie auftritt (z.B. alle Sätze, in denen Kategorie POLIT vorgegeben wurde).

Eine zweite Möglichkeit der Auswahl ist die Definition zweier Kategorien. Das Auswahlkriterium heißt, daß ein Satz dann gedruckt wird, wenn in ihm Kategorie A und B zusammen auftreten (AND Verknüpfung) (z.B. alle Sätze, in denen die Kategorien POLIT und POWCON (power conflict) auftreten).

Ein drittes Auswahlkriterium bezieht sich auf Mehrfachvercodungen eines Worts (z.B. Sätze, in denen für ein Wort sowohl NOUN als auch POLIT vorgegeben wurden).

5.4 Programm TALLY

TALLY erlaubt einfache Häufigkeitsauszählungen aller syntaktischen und semantischen Kategorien. Als Grundlage dient dazu die TAGGER Ausgabedatei. Die Häufigkeiten werden pro Dokument berechnet und gespeichert, d.h. immer wenn die Identifikation wechselt, werden neue Häufigkeiten errechnet.

Da es verschiedene Möglichkeiten der Prozentuierung gibt, erstellt TALLY nur einfache Häufigkeiten. Ein Beispiel für verschiedene Vorgehensweisen bei der Prozentuierung ist die Kategorie der N-Type-Wörter im

Lasswell Value Diktionär (siehe dort). Da diese Wörter im allgemeinen keine semantische Bedeutung haben, sollen sie nicht immer in die Prozentuierung eingehen. Nutzer der Harvard IV Diktionäre dagegen möchten die Summe aller Wörter zur Prozentuierung verwenden. TALLY erstellt immer eine Datei mit allen Häufigkeiten, die dann zu weiteren Analysen mit Statistiksoftware-Paketen verwendet werden können.

5.5 Programm CONNECT

CONNECT ermöglicht den Übergang von General Inquirer zu anderen Textanalyse und Statistikprogrammen (z.B. zu TEXTPACK V oder zu SAS, SPSS usw.). Um mit den Vercodungen oder dem disambiguierten Text weitere Analysen machen zu können, stellt CONNECT geeignete Dateien zur Verfügung. Alle Ausgabesätze können mit Identifikatoren und Satznummern versehen werden. Dabei sind drei Formen der Ausgabe möglich.

Die erste Form ist die Ausgabe des disambiguierten Texts. Dieser Text kann so formatiert ausgegeben werden, daß er mit anderen Textanalyse-systemen wie z.B. TEXTPACK V⁹ weiterverarbeitet werden kann. Alle nicht eindeutigen Wörter, für die Disambiguierungsregel im Diktionär vorhanden waren, werden in der Textdatei so gespeichert, daß nach dem Wortstamm der Bedeutungscode folgt (siehe Disambiguierung).

Eine weitere Möglichkeit ist die Ausgabe einer numerischen Datei, die pro Satz für jede Kategorie die Kennzeichnung vergeben/nicht vergeben enthält (Dummy Variablen).

Die dritte Form der Ausgabe ist ebenfalls eine numerische Datei, die pro Kategorie die Häufigkeit des Auftretens in diesem Satz enthält. Die beiden numerischen Dateien können mit Statistiksoftware-Paketen weiterverarbeitet werden.

5.6 Programm PARSER

PARSER ist das zentrale Programm zum Bearbeiten von Diktionären. Das Programm liest die Vercodungs- und Disambiguierungsregeln des Benutzers, die in einer eigenen Sprache abgefaßt sind (siehe Wörterbücher, Kapitel 3.0). Die Regeln werden auf syntaktische Fehler überprüft und in ein internes Format umgesetzt, so daß TAGGER damit seine Disambiguierung

⁹ TEXTPACK ist ein von ZUMA entwickeltes und betreutes Programmsystem zur computerunterstützten Inhaltsanalyse. Es ist für Mainframe-Rechner und PCs unter MS-DOS erhältlich.

und Vercodung durchführen kann. Das Programm kann verwendet werden, um neue Regeln aufzubauen oder bestehende Regeln zu ändern. Soll ein bestehendes Wörterbuch verändert bzw. erweitert werden, so werden alle neuen bzw. geänderten Regeln zunächst mit **PARSER** umgesetzt und anschließend mit dem im folgenden beschriebenen Programm **DICTMERG** an das bestehende Wörterbuch, bereits mit **PARSER** bearbeitete Wörterbuch angefügt.

5.7 Programm **DICTMERG**

Mit **DICTMERG** können neue oder geänderte Regeln an ein bestehendes Wörterbuch angefügt werden. Sowohl das bereits bestehende Wörterbuch wie auch die neu dazuzufügenden Regeln müssen vorher mit **PARSER** geprüft und umgesetzt worden sein. **DICTMERG** ersetzt dann geänderte Regeln, löscht entsprechend markierte Regeln und fügt alle neuen Regeln in das Diktionär ein.

5.8 Programm **PEEL**

PEEL stellt ein Hilfsmittel dar, um aus einem Diktionär im sogenannten Source-Format für einzelne Wörter Regeln auszuwählen. Will man Regeln verändern, ist der einfachste Weg, die bestehende Form mit Hilfe eines Editors oder eines Textverarbeitungssystems (z.B. **WORDMARC**, **WORD**, **NORTON EDITOR**) zu editieren. Da die Regeln im Source-Format eine Satzlänge von bis zu 3600 Zeichen haben, kann es mit einigen Textverarbeitungsprogrammen Probleme beim Bearbeiten dieser langen Sätze geben. Mit **PARSER** bereits bearbeitete Diktionäre können in keinem Texteditor verändert werden. **PEEL** unterteilt deshalb alle Regeln in Zeilen mit nicht mehr als 80 Zeichen, die dann einfach zu bearbeiten sind. Diese Regeln können dann wieder mit **PARSER** und **DICTMERG** oder mit **DOCUMENT** an ein bestehendes Wörterbuch angefügt werden oder die alte Regel im Wörterbuch ersetzen.

5.9 Programm **DOCUMENT**

Um bestehende Diktionäre zu drucken oder zu dokumentieren, kann **DOCUMENT** verwendet werden. **DOCUMENT** druckt die Regeln eines Wörterbuchs oder Listen aller im Wörterbuch vorhandenen Einträge. Optional können entweder alle Einträge mit semantischer Bedeutung oder alle

Einträge, die nur syntaktische Bedeutung haben (z.B. Artikel) oder auch alle Einträge (semantische und syntaktische) aufgelistet werden. Alle Listen werden in drei weitere syntaktische Gruppen unterteilt: Substantive, Verben und anderes.

Eine zusätzliche Möglichkeit von DOCUMENT ist das Ändern und Erweitern von Source-Format Diktionären. Der Anwender gibt seine Regeln ein. Diese Regeln werden dann an das bestehende Wörterbuch angefügt oder ersetzen dort eine bestehende Regel.

6.0 Ausblick

Der General Inquirer mag wie ein Dinosaurier im weiten Feld der heutigen Textmanipulationsprogramme wirken. Aber, trotz aller Fortschritte in der Hard- und Software und vor allem in der Benutzeroberfläche solcher Programme bleibt die Leistung des General Inquirers, große und größte Textmengen, zu disambiguieren und zuverlässig zu klassifizieren, unerreicht. Dies liegt vor allem an den Disambiguierungsregeln und den umfassenden Klassifikationsschemata des Harvard IV sowie des Lasswell Value Diktionärs, denen nichts vergleichbares im englischen Sprachraum entgegensteht. Für die historische Textanalyse eröffnet der General Inquirer, wie wir anzudeuten versuchten, neue Dimensionen der Massendatenanalyse. Für den deutschen Sprachraum empfiehlt sich TEXTPACK¹⁰ mit seiner Menüführung, sofern man keine Disambiguierung benötigt. Es ist zu hoffen, daß mit der allgemeinen Verfügbarkeit des General Inquirers, der Klassifikationsschemata und von TEXTPACK solche Untersuchungen nicht mehr nur wenigen Spezialisten vorbehalten bleiben.

¹⁰ Züll, C., P. Ph. Mohler, A. Geis, 1991. Computerunterstützte Inhaltsanalyse mit TEXTPACK PC. Stuttgart: Gustav Fischer.

Erschließung historischer Texte mit Hilfe linguistischer Methoden

von Klaus Mudersbach

1. Interesse des Historikers am Text-Verstehen und historischer Stellenwert eines Textes

Was interessiert einen Historiker an einem geschichtlichen Text und wie können linguistische Methoden der Texterschließung diesem Interesse dienlich sein? Von diesen beiden Fragen ist die erste von einem Linguisten nicht kompetent beantwortbar. Um dennoch einen Ausgangspunkt zu haben für die zweite Frage, möchte ich von einer Modellvorstellung des Historiker-Interesses ausgehen und es dem Historiker überlassen, ob er sich in dem Modell ganz oder zum Teil wiedererkennen kann und aufgrunddessen dann auch an der Antwort auf die zweite Frage interessiert ist.

1. 1. Modellvorstellung zum Historikerinteresse an einem Text

Zunächst müssen einige Parameter charakterisiert werden: Angenommen ein Historiker H sei mit einem bestimmten Text T konfrontiert, er habe außerdem Zusatzinformationen ZI zu dem Text, die es ihm gestatten, einige geschichtliche Koordinaten zuzuordnen.

Anmerkung zu "Text": Unter "Text" kann hier sicher nicht ein Schriftstück verstanden werden, das per Konvention einem einheitlichen Zweck dient oder das von der Form her vom Schreiber als eine Einheit deklariert wird. Hier soll "Text" soweit gefaßt sein, daß auch ein "vermeintlicher Text" darunterfällt, also ein Text, von dem sich eventuell herausstellt, daß er aus heterogenen Stücken verschiedener Autoren bzw. zu verschiedenen Zeiten zusammengesetzt ist.

Anmerkung zu "Geschichtliche Koordinaten": Darunter möchte ich Parameter verstehen, die die Geschichtsschreibung zur Einordnung von Text-Informationen entwickelt hat. Ich stelle mir darunter Angaben, wie Zeitabschnitt, Ort, Kulturraum, Zweck des Textes, Umstände der Entstehung des Textes etc. vor. Im folgenden verwende ich den Ausdruck *geschichtlicher Koordinaten-Bereich* und die Abkürzung GKB für einen Epochen-Ausschnitt, der durch eine geeignete Folge von Koordinaten geschichtlich sinnvoll ausgrenzbar ist.

Der Historiker H hat nun ein bestimmtes **Interesse** an dem Text T und dieses Interesse bestimmt den **Zweck der Lektüre**. Z.B.: der Historiker kennt einige geschichtliche Koordinaten des zu lesenden Textes, sein Interesse gilt

der Bestimmung weiterer Koordinaten, die zur Einordnung des Stellenwertes des Textes relevant sind. Der Zweck der Lektüre besteht dann darin, diese Koordinaten aus dem Text zu erschließen.

Ein anderes Beispiel: H kennt den geschichtlichen Koordinaten-Bereich GKB eines Textes und interessiert sich dafür, ob aus dem Text Informationen zu gewinnen sind, die das bisherige Wissen des H über GKB ergänzen oder bestätigen.

Das **Interesse des H an T** läßt sich nun vorläufig folgendermaßen kennzeichnen: H will etwas aus dem Text T erfahren über Koordinaten des geschichtlichen Bereich GKB oder über einen Aspekt bzw. eine bestimmte Detailfrage zu dem GKB (z.B. über die kulturellen Vorstellungen, über die Werte und das Wertsystem, über die Darstellung oder Einschätzung einer Person oder Aktion usw.). Durch das Interesse wird der textspezifische Zweck der Lektüre bestimmt.

Die Vorgabe des Interesses und damit des Zwecks der Lektüre ist deswegen essentiell wichtig, weil sich aufgrunddessen entscheiden läßt, welche Methode der Texterschließung am ehesten etwas zum Erreichen des Zwecks beitragen kann (siehe 3) und welche Gesichtspunkte bei der Lektüre besonders relevant sind.

Das **Verstehens-Problem des Historikers** bei dieser Ausgangslage läßt sich nun allgemein so formulieren:

Der Zugang zu dem jeweiligen geschichtlichen Bereich ist aufgrund der zeitlichen Entfernung mit Hilfe eines Wissens, das auf die Gegenwart bezogen ist, nicht gewinnbar, wenn dieses *gegenwartsbezogene Wissen* zu verschieden ist von dem Wissen in der betrachteten Epoche. Das gegenwartsbezogene Wissen kann den Zugang eventuell sogar verhindern und zu Fehleinschätzungen führen.

Anmerkung zu "gegenwartsbezogenes Wissen": Damit soll dasjenige Wissen gemeint sein, das uns z.B. ermöglicht, heutige gegenwartsbezogene (also nicht historisch orientierte) Texte zu verstehen.

Das Interesse des Historikers betrifft also auch die Frage: wie ist überhaupt erkennbar, wo eine Diskrepanz zwischen dem "damaligen" Wissen und Werten und dem heutigen besteht, und danach die Frage: wie ist diese Diskrepanz beim "Verstehen einer geschichtlichen Epoche" zu überwinden. Mit diesen Fragen wendet sich der Historiker an den Linguisten, wenn es um das Verstehen historischer Texte geht. Der Laie würde vielleicht seine heutigen Kenntnisse über Sprache, Welt und Werte an den Text herantragen und ihn

so zu verstehen trachten. Sollte er dabei tatsächlich etwas "verstehen", so wäre dies eher hinderlich als nützlich für eine geschichtsgerechte Einschätzung des Textes, denn der Laie weiß ja nicht, wo er *seine* Kenntnis an Wortbedeutungen und *seine* Sicht der dargestellten Sachverhalte einbringen darf und wo nicht. Selbst wenn ein Wort seine Bedeutung tatsächlich nicht verändern haben sollte, so müßte dies eben erst sichergestellt sein, bevor es in den Verstehensprozeß einbezogen werden kann.

Die Möglichkeiten der geschichtlichen Veränderungen von Sprache, von deren Gebrauch in Texten und von der Art, etwas in Texten darzustellen, sind vielfältig. Was sich verändert haben kann, ist unter anderem:

- die Ausdrucksseite eines Wortes, seine lautliche oder graphische Gestalt (Sprachwandel), - eventuell sogar soweit, daß es mit einem heute gebräuchlichen Wort anderer Herkunft zusammenfällt (Homonymie),
- die Bedeutung eines Wortes (Bedeutungswandel); Wörter können verschwinden, neue hinzukommen (Bezeichnungswandel),
- die stilistischen und rhetorischen Mittel der Sprachverwendung,
- die Gestaltung und Gewichtung von bestimmten Aspekten im Dargestellten,
- die im Text angesprochenen oder dahinterstehenden Kultur- und Wertesysteme,
- die Wahrnehmung bzw. Kognition in Verbindung mit dem wissenschaftlichen Wissen der Zeit,
- die Bewertung der Handlungen der dargestellten Personen, usw.

Von dieser Liste sind nur die ersten drei Parameter auf eine bestimmte einzelne Sprache beziehbar, aber die anderen benötigten ebenfalls eine linguistische Aufarbeitung, weil die Gestaltung eines Sachverhaltes mit sprachlichen Mitteln abhängig ist vom jeweiligen Wissensstand, von den jeweiligen sprachlichen Mitteln und von dem, was in der jeweiligen Epoche als selbstverständlich vorausgesetzt wurde.

Das Problem, ich möchte es das Stellenwert-Problem nennen, läßt sich also genauer so formulieren:

STELLENWERT-PROBLEM: ¹

wie kann ein Text im Rahmen seiner Zeit und seiner Funktion verstanden werden (d.h., welches ist sein bereichs- bzw. epoche-spezifischer Stellenwert), ohne daß der Filter des heutigen gegenwartsbezogenen Verstehens zu Verfälschungen führt.

Davon zu trennen ist die bewußt externe Bewertung des Textes aufgrund des heutigen historischen Wissens über den entsprechenden Bereich (Tele-skopsicht).

1.2. Kriterium der individuenorientierten Wissenschaftlichkeit

Wie kann nun eine linguistische Texterschließung zur Lösung des Stellenwert-Problems beitragen? Bevor hierauf eine Antwort gegeben wird, muß noch die Abhängigkeit der Sicht des Stellenwert-Problems vom *individuellen Text-Betrachter* angesprochen werden.

Bei der Charakterisierung dieses Problems wurde so getan, als ob einem Text durch eine bestimmte Methode der Texterschließung ein bestimmter Stellenwert hinsichtlich seiner Zeit und Funktion sozusagen objektiv zugeschrieben werden könnte. Wenn dies so wäre, dann dürfte es keinen Meinungsstreit unter Historikern hinsichtlich der Einschätzung des Stellenwertes eines Textes, der Darstellung eines Ereignisses etc. geben. Dann könnte auch, - wenn die Methode genügend präzise formuliert wäre, - ein Computer die Bestimmung des Stellenwertes übernehmen.

Da dem aber nicht so ist, muß noch die Individualität des Historikers berücksichtigt werden und damit ein Spezifikum der geisteswissenschaftlichen Arbeit: die Sichtweise des beobachtenden bzw. interpretierenden Individuums geht in die Beobachtung bzw. Interpretation ein. Mit anderen Worten: eine Beobachtung oder eine Interpretation in einer geisteswissenschaftlichen Disziplin kann nicht ohne Angabe des interpretierenden Individuums gesehen werden, - und dies gilt auch für die Geschichtswissenschaft. Auch wenn es möglicherweise in Teilbereichen einer Disziplin auch "objektive" und "wiederholbare" Beobachtungen und Interpretationen geben kann (wie es für die Naturwissenschaften gefordert wird), so können diese nicht zum Maßstab der geisteswissenschaftlichen Tätigkeit erhoben werden.

Vor der Beantwortung der Frage, was eine linguistische Texterschließung einem individuellen Historiker bei seiner geschichtsbezogenen Tätigkeit des Registrierens und Interpretierens nützen kann, muß noch geklärt werden, nach welchen Kriterien das Resultat einer Texterschließung überhaupt wis-

¹ Mit "Stellenwert" lehne ich mich an Saussures Konzept des "valeur" eines sprachlichen Zeichens (Saussure, 1916) an. Dieses Konzept wird hier dahingehend verallgemeinert, daß jede Einbettung eines Parameters in ein holistisches System damit verbunden ist, daß dem Parameter eine unverwechselbare Rolle bzw. Funktion im Ganzen des Systems zugeordnet wird. Diese systembezogene Rolle des Parameters ist sein Stellenwert, sein "valeur". Zur formalen Beschreibung dieses Begriffs finden sich Vorschläge in Mudersbach (1983a und b).

senschaftlich akzeptabel sein kann. Wie schon gesagt, kommt hier die individuellenfreie *Wiederholbarkeit* des Beobachtens bzw. Registrierens und die *Objektivität* des Resultats nicht in Frage. Um nicht die wenig ergiebige Konsequenz ziehen zu müssen, daß das (linguistische bzw. historiographische) Tun demnach nicht (natur-)wissenschaftlich sei, müssen eigene geisteswissenschaftliche Maßstäbe zugelassen sein und zwar solche, die die obengenannte Tatsache des individuengebundenen Forschens berücksichtigen, aber dennoch eine Kontrollierbarkeit bzw. Überprüfbarkeit gestatten. Diese beiden Gesichtspunkte scheinen miteinander unvereinbar zu sein, aber ich meine, mit dem folgenden

**individuenorientierten Kriterium
der geisteswissenschaftlichen Wissenschaftlichkeit**

lassen sie sich vereinbaren:

Ein Forscher hat dann ein (geistes-)wissenschaftlich akzeptables Resultat erzielt, wenn die Methode des Vorgehens, die Entscheidungen auf dem Weg zum Resultat und das Resultat selbst explizit dargestellt werden und wenn sie für Experten als mögliche Methode, als mögliche Entscheidungen und als mögliches Resultat vertretbar sind, d.h., wenn sie *transparent* und *intersubjektiv* nachvollziehbar sind.

Die im folgenden vorgeschlagenen Texterschließungsmethoden sollen gerade diesem Kriterium hinsichtlich der Methode und der Transparenz der Entscheidungen Rechnung tragen und das interpretierende Individuum in seinem Stellenwert berücksichtigen. Das dahinter stehende linguistische Programm, das Individuum in seiner individuellen (aber nicht subjektiven) Wahl zu berücksichtigen, wird im folgenden auch *individually linguistisches Programm* genannt.

Zusammenfassend können wir also

**die Modellvorstellung zum Interesse des Historikers
an Texten und an der Texterschließung**

folgendermaßen formulieren:

Ein **Historiker H** mit einem bestimmten Interesse an einem Text **T** (zu dem er bestimmte **Zusatzinformationen ZI** vorliegen hat) macht es sich zum **Zweck ZW**, an dem Text **T** als individueller Experte in Kenntnis des **Stellenwert-Problems** * bestimmte, ihn bzgl. der **Koordinaten** eines geschichtlichen **Bereichs GKB*** zu registrieren und zu interpretieren und zwar so, a) daß das **Verstehensproblem*** berücksichtigt ist und b) daß aufgrund der angewandten Methode die im einzelnen gefällten Entscheidungen und das Gesamtergebnis

* die so bezeichneten Termini wurden im vorausgehenden Text eingeführt.

transparent und intersubjektiv (besser: interindividuell) **nachvollziehbar** sind (d.h. das **individuenorientierte Wissenschaftskriterium*** wird berücksichtigt).

1.3. Zu Inhaltsanalyse (Content Analysis) und linguistischen Texterschließungsverfahren

Wenn der Historiker sein Interesse primär auf den geschichtlich relevanten Inhalt eines Textes richtet und nicht auf "linguistik-interne" Texteigenschaften, so helfen ihm Texterschließungsverfahren, die primär auf das Vermitteln von Inhalten und Herstellen von Zusammenhängen im Text ausgerichtet sind (wie die drei Methoden, die später dargestellt werden sollen). Ihm helfen dagegen keine textverarbeitenden Prozeduren - und hier komme ich auf das Angebot in Computerlinguistik, Künstlicher Intelligenzforschung und in der Inhaltsanalyse (Content Analysis) zu sprechen - folgender Art:

- **Wortstatistik:** Texte werden zerstückelt in auszählbare Vorkommen von Wörtern oder morphologisch relevanten Buchstabenkombinationen. D.h. es werden statistische Untersuchungen vorgenommen, die zu einem Text Häufigkeitsaussagen liefern, die sich genausogut aus einem Worthaufen mit derselben Verteilung ergeben hätten (!). Damit lassen sich keine sinnrelevanten Zusammenhänge erfassen, es sei denn, man *interpretiert* die Statistik so, daß bestimmte Häufigkeitsmuster mit Eigenschaften des Autors, des Rezipienten oder der Sachverhalte in Verbindung gebracht werden. Diese Vorgehensweise charakterisiert gerade die Inhaltsanalyse (Content Analysis), auf die ich im Anschluß noch kurz eingehen werde.

- **Syntax-Statistik:** ebenfalls nicht nützlich sind rein linguistische Analysen auf Satzebene (z.B. Länge von Sätzen, Komplexität von Sätzen etc.), die nicht funktional auf das Textverstehen bezogen sind. Hingegen ist z.B. das Auflösen von pronominalen Bezügen zwar eine Untersuchung von Einheiten der Wortebene (Pronomina) und deren Eigenschaften, aber durch das Zuordnen von Referenzobjekten innerhalb von Texten und mit Bezug auf die Herstellung von Zusammenhängen entsprechend nützlich für das Textverstehen.

- **Klassifizierung von außen:** Vorgabe eines Kategorienkatalogs zur Klassifizierung von (Wort-)Bedeutungen, bei dem die Kategorien nicht aus dem Text heraus gebildet werden, sondern von außen und unabhängig von diesem Text vorgegeben werden. So ist z.B. ein Kategoriensystem, das einen bestimmten Bereich von Werten, z.B. Hochzeitsriten des Mittelalters, klassifizieren soll, wenig sinnvoll, wenn es nach heute unterscheidenswerten Kategorien eingeteilt wird. Sinnvoll ist es dagegen, Kategorien zu verwenden, die unter Berücksichtigung des Textes, der Zusatzinformation und des Zwecks des Historikers und aus der Kenntnis des geschichtlichen Bereichs die

dort relevanten Unterscheidungen verwenden und danach klassifizieren (d.h. die das Stellenwert-Problem mitberücksichtigen).

- **Skalierung von Textinformation:** Ähnliches wie für die Klassifikation gilt für die Skalierung. Wenn z.B. emotive Ausdrücke in einem Text zu skalieren sind, so darf der Leser nicht einfach *seine* emotiven Unterschiede als Skala verwenden, er muß vielmehr ähnliche Texte desselben Autors oder Texte zum selben Thema bei verschiedenen Autoren derselben Zeit miteinander vergleichen, um eine geeignete Skalierung zu finden. Denn der Stellenwert des einzelnen emotiven Ausdrucks auf der Skala kann zu verschiedenen Zeiten und für verschiedene Personen verschieden sein. Es muß versucht werden, die individuellen Bedeutungen bei der betreffenden Person bzw. die "soziale" Bedeutung zur selben Epoche zu modellieren.²

Um Mißverständnissen vorzubeugen: die oben genannten Verfahren können dann nützlich sein, wenn der Historiker z.B. den Zweck verfolgt festzustellen, ob zwei Texte vom selben Autor stammen und dazu gezielt linguistische Wort-, Satz- oder Texteigenschaften untersucht, d.h. wenn er glaubt, dies anhand der Häufigkeit von bestimmten Wörtern bzw. Satzbaueigenschaften etc. nachweisen zu können. Ebenso können globale Kategorienvorgaben dann sinnvoll sein, wenn das Ziel des Historikers gerade ist, quer durch die Epochen hindurch Verteilungen festzustellen.

Auf die Inhaltsanalyse (Content Analysis) soll hier noch kurz eingegangen werden: sie zeigt ein vielfältiges Spektrum an hauptsächlich statistischen Verfahren, deren Ziel - zusammengefaßt - darin besteht, von *manifesten* Texteigenschaften auf *latente* außertextliche Eigenschaften (Eigenschaften des Autors, des Rezipienten bzw. der Kommunikationssituation) zu schließen.³ Als manifeste Texteigenschaft zählt nur das, was behavioristisch gesehen an den Ausdrucksgestalten erkennbar ist. Diese Ausdrucksgestalten haben zwar

² Auch wenn die Übertragbarkeit auf historische Fragestellungen offen bleiben muß, so möchte ich doch auf die Skalierung der Gefühle der Angst bzw. des Hoffens nach dem Gottschalk-Gleser-Verfahren hinweisen. (Gottschalk, Gleser 1969, Koch, Schöfer 1986). Hiernach werden die Äußerungen einer Person, die Ausdrücke der Angst enthalten, auf einer vom Verfahren vorgegebenen Skala situiert, ohne daß geklärt wird, wie die individuelle Skalierung der Person angelegt wäre. Als erstes, aber unausgesprochenes Postulat geht in solche Testverfahren die ungeprüfte Annahme ein, daß die Bedeutung der verwendeten sprachlichen Ausdrücke bei der interviewten Person, beim Interviewer und beim Text-Auswerter immer dieselbe sei (vgl. die Kritik an der Content Analysis von Berelson (1952), S. 19). Zur Kritik am Gottschalk-Gleser-Verfahren aus linguistischer Sicht: siehe auch Stutterheim (1988).

³ Vergleiche dazu die Übersichtsdarstellung in Merten (1983).

Bedeutung, aber sie ist latent und damit nicht direkt zugänglich. Diese unreflektierte Bilateralität des Zeichenbegriffs, die hier zum Ausdruck kommt, zwingt außerdem zu der Annahme, daß die Bedeutung hinter einer Ausdrucksgestalt nicht individuell verschieden sein kann, sondern für alle Mitglieder einer Sprachgemeinschaft gleich sein muß. Denn nur so läßt sich aus dem Wortmaterial statistisch etwas über die Bedeutung der Wörter erschließen. Von linguistischer Seite ist dieser behavioristische Ansatz deutlich kritisiert worden (Fühlau 1982, siehe auch Berelson 1952 in Fußnote 2). Dagegen ermöglicht ein *relationales* Zeichenmodell (Mudersbach 1983b), wie es der nachfolgend dargestellten Methode RELATEX zugrunde liegt, die Bedeutung als *manifeste* und *textspezifische* Texteigenschaft zu betrachten (ohne Zuhilfenahme von Statistik). Statt der statistischen "Brücke" zwischen Texteigenschaft und Intention des Autors wird die textspezifische Bedeutung direkt als individuelle Bedeutung des Autors interpretiert. Das Konzept, Bedeutung als *manifeste* Texteigenschaft aufzufassen, wird dabei fast dem behavioristischen Interesse gerecht, genau genommen dem (historisch nachfolgenden) kognitivistischen Interesse: das sprachliche Bedeutungsraaster und das kognitive Raaster einer Person gleichzusetzen und so die Brücke der Interpretation überflüssig zu machen.

Was Arbeiten in der Computerlinguistik und Künstlichen Intelligenz Forschung angeht, so ist unter dem Gesichtspunkt des Historiker-Interesses, wie es hier zugrunde gelegt wurde, folgendes zu sagen. Es lassen sich zwei Gruppen von Arbeiten unterscheiden, die aber beide an dem spezifischen Interesse des Historikers (noch) vorbeigehen. Die eine Gruppe enthält an historischen Texten orientierte Arbeiten der Computerlinguistik (z.B. Lenders 1982): sie sind eher von linguistisch-statistischem Interesse, weil sie ihre Aufgaben in den oben genannten statistischen Analysen im Bereich von Morphologie und Syntax sehen. In der anderen Gruppe sind Arbeiten zur Textanalyse bzw. zum Text-Verstehen im allgemeinen. Hierzu gehören - um eine Auswahl zu nennen - z.B. Analysen von Wetterberichten (verschiedene Beiträge in Rollinger 1984), von einer Zeitungsnotiz (verschiedene Beiträge in Batori 1982) oder von Kriminalgeschichten (Neuhaus 1988). Hier fehlt aber das den Historiker interessierende Problem des Verstehens in zeitlicher Distanz völlig. Außerdem wird immer der Standpunkt des Computers eingenommen: wie kann er einen Text "verstehen" d.h., wie kann man zu einer *eindeutigen* Textanalyse mittels Computerprogramm gelangen. Die Dimension des individuellen Rezipienten bzw. des Produzenten, also die *individual-linguistische* Sicht eines Textes fehlt ganz, - verständlicherweise, weil sich hermeneutisch orientierte

Wissenschaftler noch nicht mit ihren Forderungen an die Computerlinguistik gewandt haben. Meiner Meinung nach gibt es keinen prinzipiellen Hinderungsgrund, solche Forderungen in der Computerlinguistik aufzunehmen. Allerdings würde dann die Mensch-Maschine-Interaktion oder -Kooperation anders zu gestalten sein. Ansätze in dieser Richtung ergeben sich aus den im nächsten Teil dargestellten Methoden des Textverstehens.

Damit will ich die einleitende Betrachtung des Modells des Interesses des Historikers und die Erörterung der sich für die linguistische Texterschließung ergebenden Fragen abschließen.

2. Individuallinguistische Texterschließungsverfahren

Im folgenden will ich mich also auf Methoden der Texterschließung bzw. des Textverstehens konzentrieren, die dem individuellen Historiker beim individuellen Textverstehen helfen können (individuallinguistische Methoden, dazu vgl. Mudersbach 1988).⁴

Der Leitgedanke soll dabei sein, das gegenwartsbezogene Wissen "abzuschalten" (und damit auch ein Textverstehen aus der Gegenwart bzw. aus dem individuellen gegenwartsbezogenen Bewußtsein des Individuums H heraus), andererseits sollen die *individuellen* Entscheidungen des Experten, d.h. des Historikers H bzgl. der historischen stellenwertorientierten Textbewertung ernst genommen werden. Und dies wiederum soll nicht ausschließen, daß stufenweise die heutigen Kenntnisse des Historikers über die frühere Epoche miteinbezogen werden können, indem er sie als explizit formulierte Hypothesen hinzunimmt.

Die Vielfalt der Möglichkeiten, einen Text zu betrachten, wird sicher nicht vollständig anhand einer einzigen Methode erfaßt werden können. Im folgenden sollen daher mehrere sich ergänzende bzw. aufeinander aufbauende Methoden vorgeschlagen werden.⁵ Ich möchte sie so darstellen, daß ein Historiker in der Lage sein sollte, sie ohne besondere linguistische Vorkennt-

⁴ Das Textverstehen steht nicht nur in der Spannung zum Nicht-Verstehen, sondern auch zum Mißverstehen. Was man als Leser nicht versteht, das kann man sich konstatieren; was man mißverstehen, bleibt einem selbst eine Zeitlang unentdeckt, weil man ja glaubt verstanden zu haben. Die vielfältigen Möglichkeiten, die Äußerung eines anderen mißzuverstehen (z.B. die zehn Möglichkeiten in Mudersbach 1987), sind auch auf das Gebiet der historischen Texte übertragbar.

⁵ Diese Methoden haben sich aus dem praktischen Interesse an Texten entwickelt. Sie sind aber auch in Verbindung mit unterschiedlichen linguistischen *Theorien* oder Modellen des Textes bzw. des Verhältnisses Text-Leser zu sehen. Die Darstellung dieser Modelle soll jedoch in einer linguistischen Arbeit zum Thema ausgeführt werden.

nisse zu verstehen und eigenständig anwenden zu können.

Es ist hier noch kurz auf die Frage nach dem *Einsatz von Computern* bei diesen Methoden einzugehen: die hier dargestellten Methoden sind bei kürzeren Texten prinzipiell auch ohne Computer anwendbar. Bei längeren Texten ist es aber aus Zeitgründen fast unmöglich, daß der Historiker das entsprechende Computerprogramm zum Registrieren und Auswerten der gewählten Entscheidungen zur Verfügung hat. Die hier vorgestellten Methoden des Textverstehens gehen bei der Computerunterstützung arbeitsteilig vor: bei jeder Methode muß der Historiker bestimmte Entscheidungen individuell fällen und dem Computer eingeben. Der Computer ordnet diese Entscheidungen nach bestimmten Prinzipien und berechnet gegebenenfalls ein sich daraus ergebendes Interpretationsresultat. Auf die Programme komme ich nach Darstellung der Methoden noch einmal zurück.

Im folgenden werden drei Methoden vorgestellt, die zusammengenommen meiner Meinung nach die Gesichtspunkte, unter denen ein Textinterpret einen Text versteht, aufnehmen und systematisieren können. Jede Methode betrachtet den Text aus einem anderen "Blickwinkel" und ist mit einer Arbeitsteilung in der Texterschließung verbunden: jede Methode schöpft ihren abgrenzbaren Arbeitsbereich aus, so daß sich die drei Blickwinkel insgesamt ergänzen zu einer individuellen Gesamtschau des Textes.

Es hängt vom Zweck des Historikers und vom jeweiligen Text ab, welche Methode vorrangig geeignet ist. Dies wird nach der Darstellung der Methoden noch erläutert. Im Anhang werden alle Methoden an einem Text skizzenhaft vorgeführt.

Zunächst soll der Blickwinkel jeder der drei Methoden kurz charakterisiert werden, dann die zur Anwendung erforderliche Kompetenz des Lesers und schließlich die Schrittabfolge zu jeder Methode.

* Die *erste Methode* berücksichtigt die *Gesichtspunkte*, die der individuelle Leser und Interpret *individuell als relevant für sein Verstehen des Textes* ansieht. Jeder Gesichtspunkt (nachher "Aspekt" genannt) enthält Merkmale (Ausprägungen, Werte), die der Textverstehende nach eigenem Ermessen zusammenstellen kann, um damit die *Differenziertheit seines Verstehens* explizit zu machen.

Die Methode heißt ASPECTEX (*Aspectives Lesen des Textes*).

* Die *zweite Methode* betrachtet die *informative Struktur* des Textes, sowohl in der zeitlichen linearen Abfolge als auch in der synchron-optischen Zusammenfassung des ganzen Textes. Hierbei werden die im Text enthaltenen *semantischen Relationen* in einem Netz zusammengestellt.

Die Methode heißt RELATEX (für *relationale Analyse des Textes*).

* Die *dritte Methode* berücksichtigt das *Hintergrundwissen*, das der Leser an die Textteile anlegen kann. Der Leser ergänzt so das im Text stehende um die allgemeinen Kenntnisse, die ihm zum Thema oder zur Epoche schon zur Verfügung stehen. Dieses Hintergrundwissen besteht aus einer Menge von Systemen, von denen jedes einen bestimmten thematischen Bereich (der Kultur, des Lebens, der Wissenschaften) abdeckt. Auch das Wissen über die Textgestaltung (z.B. einer Urkunde oder eines Briefes) gehört dazu. Jedes System besteht aus Relationen zwischen Begriffen. Den Begriffen wird dabei eine holistische Rolle bzgl. des ganzen Systems (dem HOLON) zugeschrieben.

Die Methode heißt daher HOLONTEX.

Allen drei Methoden ist gemeinsam, daß das interpretierende Individuum im Prinzip frei ist, seine individuelle bzw. textspezifische Ausfüllung des methodischen Schemas vorzunehmen und dennoch dieses Verfahren sich und ändern transparent und reflektierbar zu machen.

Jede dieser Methoden erfordert eine andere *Kompetenz des Lesers*:

ASPECTEX die Kompetenz des Lesers, Gesichtspunkte sprachlicher und inhaltlicher Art in differenzierter Form zur Verfügung zu haben, um Textstellen dementsprechend beurteilen zu können,

RELATEX die Kompetenz, Sätze in basale Informationseinheiten und deren Beziehungen untereinander umsetzen zu können, um das "Gewebe" des Textes (Vernetzung der Informationen) im ganzen überschauen zu können.

HOLONTEX die Kompetenz, an Textstellen den Bezug auf ein System erkennen zu können und dies beim weiteren Textverlauf einbringen zu können, um so die Folgerichtigkeit und Erwartbarkeit des Textverlaufs beurteilen zu können, aber auch um die Explizitheit des Textes bzw. seine implizite Stütze auf geeignetes Hintergrundwissen erfassen zu können.

Die Methoden sollen nun im einzelnen dargestellt werden. Da jede Methode so formuliert sein soll, daß sie im Prinzip auch von einem Computerprogramm in Interaktion mit dem Leser durchgeführt werden kann, ergibt sich pro Methode eine bestimmte Schrittfolge, ein Algorithmus in vier Schritten. Teilweise lassen sich beim Vergleich der drei Methoden diese Schritte parallel setzen.

VORGABEN: Für jede der Methoden sind einheitlich folgende Vorgaben erforderlich:

- der zu lesende Text,
- Eingangsinformationen d.h. Zusatzinformationen zum Text (die den Autor, den Adressaten, die Zeit, den Anlaß des Textes etc. betreffen),
- die Angabe, warum der Leser sich für den Text interessiert (der Zweck der Lektüre),
- als Voraussetzung: die Kompetenz des Lesers, die jeweiligen Parameter anlegen zu können.

2. 1. Erste Methode: Bewertung von Aspekten (ASPECTEX)

Die *Kern-Idee* dieser Methode besteht darin, den Text unter einem *einzigem Aspekt* ganz zu lesen, um sich maximal zu sensibilisieren für Erscheinungen im Text, die mit diesem Aspekt zusammenhängen.

Ich werde kurz die Schritte der Methode charakterisieren. Im Anhang befindet sich ein skizzenhaftes Beispiel dazu (vgl. Abb. 1a und 1b).

ASPECTEX-SCHRITTFOLGE (4 Schritte):

Vorgaben: Text, Eingangsinformationen zum Text, Zweck der Lektüre und Kompetenz des Lesers, Textaspekte differenzieren zu können.

1. Schritt: Erstlektüre

Die erste Lektüre des Texts enthält eine Möglichkeit, die bei jeder weiteren Lektüre verlorengeht: nämlich überrascht zu werden durch unerwartete Auffälligkeiten. Um diesen Effekt bei sich zu kultivieren und zu beobachten, sollte der Text in kleinen Portionen (z.B. satzweise) gelesen werden. Bei jedem gelesenen Textsegment soll jeweils schriftlich angegeben werden,

- ob die Textstelle Überraschendes, Auffälliges enthält oder ob Fragen offen geblieben sind,
- ob sich ein Gesichtspunkt (Aspekt oder System) herauskristallisiert, der in den weiteren Schritten eventuell systematisch zu bearbeiten wäre,
- ob sich eine Erwartung an das Nachfolgende ergibt und
- ob eine bis dahin vorhandene Erwartung bestätigt oder enttäuscht wird, oder ob die Erwartung nicht tangiert war und daher als Frage an den weiteren Text beibehalten wird (cf. Anhang, Abb. 1a).

Auf diese Weise kann an jeder Textstelle eine Erwartungsbilanz gezogen und geprüft werden, ob man den Text bis dahin intuitiv "verstanden" hat oder ob Fragen offen geblieben sind.

Es ist wichtig, die *Erwartungsbilanz* pro Textsegment aufzustellen, da nur bei der ersten Lektüre bestimmte Erwartungen in das Blickfeld geraten. Nach Kenntnis des ganzen Textes bzw. bei jeder weiteren Lektüre treten sie in dieser Form nicht wieder auf. Die Folge der Erwartungsbilanzen zeigt den Erwartungsverlauf. Er gibt Hinweise darauf, wo man stutzig geworden ist, wo bestimmte Erwartungen aufgrund des Hintergrundwissens nicht bestätigt werden, welche Erwartungen sich als fehlgeleitet ergeben haben und welche offen geblieben sind.

Da bei dieser Erstlektüre auch bestimmte Systeme aus dem Hintergrundwissen ins Blickfeld kommen, die eventuell nachher systematisch an den Text anzulegen sind (nach der Methode HOLONTEX), ist dieser Schritt auch zugleich eine Vorbereitung der dritten Methode.

2. Schritt: Aufstellen einer ASPEKTLISTE

Aspekte dienen dazu, sprachliche und inhaltliche Eigenschaften des Textes zu beurteilen. Ein Aspekt ist ein Rahmen, in dem sich der Leser bestimmte Differenzierungsmöglichkeiten (Aspektwerte) vorgibt, die es ihm gestatten, eine Textstelle zu beurteilen. Zum Teil haben sich die Aspekte und Differenzierungsmöglichkeiten schon im ersten Schritt ergeben.

Jetzt soll für jeden Aspekt die Liste der für den Text relevanten Aspektwerte explizit aufgestellt werden. Hierbei ist darauf zu achten, daß die Liste nach Möglichkeit genau die Differenzierungen enthält, die zur Beurteilung des Textes benötigt werden. Die einzelnen Differenzierungen sollen gegeneinander abgegrenzt sein, damit nicht zwei Aspektwerte bei derselben Textstelle als Bewertung in Frage kommen. Aspekte können Textstellen unterschiedlicher Länge bewerten. Der Erwartungsverlauf und die Auffälligkeiten, die im ersten Schritt aufgestellt wurden, sollen ebenfalls in die Aspekte eingearbeitet werden oder als eigene Aspekte angesetzt werden.

Wir erhalten also Tabellen der folgenden Form:

TABELLE T1: Formale Aspektliste (Aspectex, Schritt 2)

ASPEKT 1	ASPEKT 2	ASPEKT 3	(Aspektnamen)
Wert 1.1 Wert 1.2	Wert 2.1 Wert 2.2 Wert 2.3 Wert 2.4	Wert 3.1 Wert 3.2 Wert 3.3	("Aspekt- werte": Listen von unterschied- lichen Längen)

3. Schritt: Aspektives Lesen des Textes

Dies ist der zentrale Vorgang der Methode: man wähle *einen* Aspekt aus der Aspektliste aus und lese den ganzen Text nur unter diesem einen Aspekt. Die dafür geeigneten Textsegmente werden unter diesem Aspekt beurteilt, indem man jeweils einen zum Aspekt gehörenden Aspektwert in eine an den Text angelegte Liste einträgt.

Durch die Konzentration auf einen einzigen Aspekt bei der Lektüre sensibilisiert sich der Leser für diesen Aspekt und seine Werte. Dabei kann es vorkommen, daß man noch weitere Werte in die Liste der Aspektwerte aufnehmen möchte. In diesem Fall muß man darauf achten, daß man auch die vorausgehenden Textteile nach dem neuen Bewertungsrahmen bewertet, da der Bewertungsrahmen ja einheitlich an den ganzen Text angelegt werden soll.

Dieses Aspektive Lesen führe man mit jedem Aspekt der Aspektliste durch. D.h. der Text soll für jeden interessierenden Aspekt ganz gelesen werden. Das Ergebnis der Gesamtlektüre ist eine "Aspektmatrix", in der zu jeder Textstelle pro angelegtem Aspekt Aspektwerte eingetragen sind, sofern dies für diese Textstelle sinnvoll ist. Diese Matrix gibt dann das individuelle Verständnis des Textes auf der Mikroebene wieder, denn jede Eintragung gibt ja die Entscheidung des Lesers wieder, die Stelle soundso verstanden zu haben. Das Makro-Verstehen, das mit dem Anlegen eines Systems aus dem Hintergrundwissen zu tun hat, ist bei diesem Vorgehen nicht erfaßt. Das soll die Methode HOLONTEX leisten.

4. Schritt: Gewichtung des Zwecks der Lektüre

Zu dieser sozusagen "interessenfreien" Lektüre des Textes kommt in diesem Schritt jetzt die Auswertung des Lesezwecks hinzu. Und zwar in folgender Form: der Zweck der Lektüre wird interpretiert als das Präferieren bestimmter Aspekte gegenüber anderen. Daher werden die Aspekte nach dem Interesse des Lesers gewichtet. Diese Gewichtung kann qualitativ vorgenommen werden durch eine explizite Markierung der Aspekte mit Gewichten: besonders wichtig,..., weniger wichtig, nicht wichtig, etc. - Die Gewichtung dient zum Transparentmachen des eigenen Interesses (eventuell im Gegensatz zum Interesse eines anderen Lesers). Statt der qualitativen Angaben kann auch eine quantitative Angabe (z.B. auf einer Skala zwischen 0 und 1) die individuelle Wichtigkeit des Aspektes verdeutlichen. Diese quantitative Gewichtung ist auch geeignet, bestimmte Texteneigenschaften gegeneinander abzuwägen und ihre Relevanz im Gesamturteil zu markieren.

Man wird fragen: warum wurde der Zweck der Lektüre nicht schon beim Aufstellen der Aspektliste (im 2. Schritt) berücksichtigt. Denn dann erspart man sich praktisch die Auswertung der weniger relevanten Aspekte. Das stimmt, aber diese verkürzte Anwendung der Methode birgt die Gefahr in sich, daß dann bestimmte Aspekte, die eventuell erst bei der Auswertung ihre Relevanz zeigen, zu früh von einer systematischen und damit eventuell auch erhellenden Betrachtung ausgeschlossen werden.

Die Methode ASPECTEX ist sowohl zur methodischen Lektüre eines Textes geeignet als auch zur Rekonstruktion einer schon geleisteten Textinterpretation. In diesem Fall müssen aus dem interpretierenden Text die Aspekte herauspräpariert werden, die der Interpret seiner Auffassung zugrunde gelegt hat. Die Rekonstruktion zeigt, wie (un)systematisch er dabei vorgegangen ist.

Zu dieser Methode wurde ein Computerprogramm entwickelt,⁶ das zum einen bei längeren Texten die aspektive Bearbeitung erleichtert, zum anderen den Ablauf und die Konsistenz der Bewertungen kontrolliert.

2.2. Zweite Methode: Extraktion von Textinformation (RELATEX)

Die zweite Methode des Textverstehens, RELATEX, betrifft den Informationsgehalt eines Textes, der sich in Form von Relationen mit Argumenten ausdrücken läßt. Zur Anwendung der Methode muß der Textverstehende die Kompetenz haben, Relationen aus dem Text herauspräparieren zu können. Vorgaben sind wie bei der ersten Methode der Text, Eingangsinformationen und der Zweck der Lektüre.

Die *Idee* dieser Methode besteht darin: aus dem Text sollen Satz für Satz Relationen gewonnen werden, die zu einem semantischen Netz zusammengefaßt werden (vgl. Beispiel-Analyse im Anhang und Abb. 2).⁷ Aus dem Netz

⁶ Das Programm wurde von H.D. Koch entwickelt und ist außerdem für das Übersetzen nach der Aspektiven Methode eingerichtet. Es heißt ASPECTRA (Koch, Mundersbach 1988). - Ich möchte an dieser Stelle Herrn Koch danken für seine Einsatzfreude beim Erstellen des Programms und für die gute Zusammenarbeit.

⁷ Semantische Netze werden seit Collins/Quillian (1969) zu verschiedenen Zwecken in der Linguistik und der KI-Forschung verwendet. In der Textlinguistik gibt es unterschiedliche Interpretationsmöglichkeiten der Relationen, die aus einem Text extrahiert werden können, je nachdem wie sehr man die im Text vorkommenden Relationen zerlegt in Elementarbeziehungen (z.B. in Schank (1975) bzw. in Beaugrande, Dressler (1981). Der hier gewählte Ansatz unterscheidet sich von der zerlegenden Interpretation darin: um ein vorschnelles falsches Verständnis der Begriffe in einem Text zu vermeiden, werden in RELATEX die Begriffe eines Textes nicht analysiert, sondern als "Bedeutungs-Atome" angesehen.

lassen sich dann bestimmte Texteigenschaften ermitteln.

Was unter einer *Relation* zu verstehen ist, soll zunächst kurz an einem einfachen Beispiel erläutert werden: der Satz "Paul schenkte Petra ein Buch" wird aufgefaßt als eine *Relation* mit dem dreistelligen RELATOR "schenkte" und den drei *Argumenten* "Paul", "Petra" und "ein Buch". Die Relation kann durch einen Klammerausdruck ähnlich wie in der Prädikatenlogik notiert werden:

schenkte (Paul, Petra, ein Buch)

wobei mit der Reihenfolge der Argumente auch ihre *semantischen Rollen* in bezug auf den Relator festgelegt sind:

1. Stelle: der Schenkende (Paul),
2. Stelle: der Beschenkte (Petra),
3. Stelle: das Geschenk (ein Buch).

Ein Vorschlag zu einer graphischen Notation, die sich eher eignet zur Veranschaulichung von Relationsnetzen, findet sich im Anhang. Die Idee dabei ist in jedem Fall: die Informationen in einem Satz nur insofern zu berücksichtigen, wie es für die Kenntnis der Argumente, der zeitlichen und räumlichen Struktur des Textes und die Kennzeichnung der Relatoren (eventuell ergänzt durch Modifikatoren) erforderlich ist. Bei der Vernetzung wird dann hauptsächlich danach gefragt: welche Relationen enthält der Text zu einem bestimmten Begriff.

Die einzelnen Schritte der Methode betreffen hier nicht die Bearbeitung des ganzen Textes, sondern jeweils eine Textstelle. Neben dem Text wird ein *Lexikon LX* angelegt, in das alle neuen Wörter oder Wortgruppen der relevanten Kategorien aufgenommen werden.

RELATEX-SCHRITTFOLGE (4 Schritte)

Vorgaben: der Text, Eingangsinformationen, Zweck der Lektüre und Kompetenz des Lesers, Relationen aus Sätzen zu bilden.

Um die Methode ganz allgemein zu charakterisieren, soll nicht mit dem Textanfang begonnen werden, sondern mit einem Satz mitten im Text.

Angenommen: die ersten n Sätze des Textes seien schon nach der Methode analysiert. Demnach ist also das Lexikon $LX(n)$ schon mit Wörtern und deren Kategorien bis zu diesem Satz aufgefüllt und das semantische Netz zum Text bis zu dieser Stelle liegt schon vor ($NZ(n)$).

Der nächste Satz S soll nun analysiert werden.

1. Schritt: Ergänzen des Lexikons zur Textstelle (Klassifikation der neuen Wörter nach semantischen Kategorien)

Der Satz S wird in eine oder mehrere Relationen mit Argumenten segmentiert. Daraus ergibt sich, welche Wörter oder Wortgruppen zu den Kategorien Relator, Argument, Argumentmodifikator, Relatormodifikator gehören. Sind hierunter Einheiten, die noch nicht im bisherigen Lexikon LX(n) enthalten sind, dann werden sie in das Lexikon aufgenommen.

2. Schritt: Erstellen der Relationen zur Textstelle

Die Relationen zu S werden gebildet. Mehrfache Lesarten werden getrennt notiert. Der Leser kann auch *Zusatzhypothesen* formulieren, die aus seiner Sicht hinzugefügt werden müssen, oder evtl. aus der Sicht eines der beteiligten Personen. Diese Hypothesen müssen als Zusätze zur Textinterpretation markiert werden.

3. Schritt: Integration der Relationen in das semantische Netz

In das vorhandene Netz wird nun die Relation zu Satz S eingefügt (eventuell parallel dazu auch mehrere). Das Hinzufügen geschieht bei einer graphischen Darstellung dadurch, daß die Relation an das vorhandene graphische Netz so angefügt wird, daß von den schon vorhandenen Begriffen ausgehend die Relationen zum Satz S gezeichnet werden. Eventuell führt der neue Satz zu einer *Desambiguierung* des vorhandenen Netzes. D.h. S dient dazu, die Darstellung(en) der schon vorhandenen Sätze zu reduzieren. Andererseits führt ein Satz mit z.B. drei verschiedenen Lesarten zu einer *Auffächerung* der Netze: sie müssen in drei getrennte Netze aufgespaltet werden.

Falls der Text eine zeitliche Abfolge enthält, so ist der Zeitparameter entweder an der Relation insgesamt zu notieren oder an den Argumentstellen, falls Argumente sich zeitlich entwickeln.

4. Schritt: Zweckorientierte Teilnetzbildung

Unter dem Zweck der Lektüre können bestimmte Relationen bzw. Argumente stärker gewichtet werden als andere, so daß sich eine zweckspezifische Ausblendung eines Teilnetzes aus dem Gesamtnetz ergibt.

Die Schrittfolge 1 - 3 ist fortlaufend auf die Abfolge der Sätze im Text anzuwenden. Der Schritt 4 kann je nachdem entweder pro Textstelle angewandt werden, indem man entscheidet, ob man diese Stelle überhaupt aufnehmen will oder erst bei Vorliegen des Gesamtnetzes.

Zu dieser Methode gibt es ein gut ausgearbeitetes Computerprogramm RELATAN:⁸ nach Eingabe der Relationen in eine Datenbank sortiert das Programm die Argumente nach verschiedenen Gesichtspunkten und erstellt ein virtuelles semantisches Netz, zu dem verschiedene Eigenschaften abgefragt werden können wie z.B. die Anzahl der Teilnetze (Inseln), die am stärksten ausgeprägten Begriffe (d.h. die, die mit den meisten Relationen versehen sind); der längste Weg durch die Insel; der Begriff, der in der Mitte der Insel liegt usw. Außerdem wird analysiert, ob zwischen den Inseln "Brücken" bestehen, d.h. ob Inseln nur über einen oder wenige Begriffe miteinander zu einem einheitlichen Netz verbunden sind. Diese Analyseform bietet die Möglichkeit, neue strukturelle Texteneigenschaften zu untersuchen, die je nach Herkunft des Textes verschiedenartig interpretiert werden können.

Das Ergebnis nach der Interpretation der n-ten Textstelle ist ein (oder mehrere) semantische Netz(e) zu dem Textstück bis zu dieser Textstelle. Nach der relationalen Interpretation des ganzen Textes liegen ein oder mehrere Netze vor je nach Anzahl der nicht aufgelösten Lesarten.

An dem Gesamtnetz ist folgendes ersichtlich:

G1. Das Netz kann eventuell aus unverbundenen Teilnetzen bestehen. Anzahl und Größe der Teilnetze geben Aufschluß über die gedankliche (Un-)Einheitlichkeit des Textes.

G2. Diese Methode scheint mir für die Erfassung von historischen Texten besonders gut geeignet zu sein aus folgendem Grund: sie liefert die *textspezifische Bedeutung* (siehe G3.) der im Text vorkommenden Begriffe, wenn man folgendes betrachtet: in einem Text können sowohl *kontingente* Sachbehauptungen vorkommen als auch *gesetzesartige* Aussagen. Diese sind üblicherweise daran zu erkennen, daß sie eine allgemeine Formulierung erhalten, die unabhängig ist von kontingenten Veränderungen. Wenn man nur diese gesetzesartigen oder regelhaften Aussagen als Relationen formuliert und in ein semantisches Netz zusammenfaßt, dann erhält man zu einem darin vorkommenden Begriff die sich aus dem Text ergebenden Information über das gesetzesartige *Sprach- und Sachwissen*. Diese lassen sich ergänzen durch die Informationen über den *Möglichkeitsraum* der Begriffe, der aus den kontingenten Äußerungen des Textes gewinnbar ist. Z.B. enthält die Äußerung:

⁸ Das Programm zu dieser Methode wurde von L. Schönherr entwickelt und heißt RELATAN (Schönherr, Mudersbach 1988). - Herrn Schönherr möchte ich meinen Dank aussprechen für die Begeisterung, mit der er das Programm erstellt und ständig perfektioniert hat.

"dieser A hat - wie jeder A - ein B, aber auch ein C".

drei Arten von Informationen:

- eine gesetzesartige, nämlich:

"A-sein impliziert: ein B zu haben"

- eine kontingente

"dieses A hat ein B und ein C"

- und auch eine Information über den Möglichkeitsraum der Begriffe A und C.

"Möglich ist für einen A-Gegenstand: er hat ein C."

Trivialerweise gehört zum Möglichkeitsraum von A bzw. B auch:

"Möglich ist für einen A-Gegenstand: er hat ein B."

und zwar wegen des schon oben formulierten Gesetzes.

Ich habe hier absichtlich Buchstaben A, B, C verwendet, um deutlich zu machen: selbst ohne Kenntnis der Bedeutungen dieser Begriffe kann man über ihre *Bedeutungsbeziehungen* untereinander schon einige Schlüsse aus dem Text ziehen, ohne diesen Begriffen selbst eine Bedeutung zuzuordnen zu müssen, die unserem heutigen Verständnis entspricht (denn bei inhaltlich aufgefüllten Begriffen wie "Ehre", "Vaterland", "Kirche" ist es gar nicht einfach, von unserem heutigen Verstehen konsequent zu abstrahieren und sie als unbekannte Konstanten zu lesen, deren Bedeutung erst der Text ergeben soll!). Nur wenn wir ein Wort des Textes zunächst als eine Unbekannte betrachten, die zu analysieren ist im Rahmen der Gesetzesanalyse des Textes, halten wir uns offen für die "uns fremden" Bedeutungen, die sich aus dem Text ergeben. Oft ist uns gar nicht bewußt, was wir unkontrolliert aufgrund der intuitiven Zuordnung von Ausdruck und Inhalt einem Wort im Text aus *unserer* Sprachkenntnis zuzuordnen. Bei RELATEX wird nicht Gebrauch gemacht von unserer Fähigkeit, Texte zu verstehen (und wohlwollend verstehen zu wollen), sondern nur von der Fähigkeit, die syntaktische Struktur der Sätze erkennen zu können und die kontingenten von den gesetzesartigen Aussagen aufgrund der syntaktischen Form der Aussage unterscheiden zu können.

Erst in einem zweiten Durchgang soll der Textinterpret seine Hypothesen in Form von Relationen formulieren und zu dem Textnetz hinzufügen. Dabei muß immer klar erkennbar bleiben, was aus dem Text stammt und was hypothetisch hinzugefügt wurde.

G3. Die *textspezifische Bedeutung eines Begriffswortes* W läßt sich nun anhand der unter G2. gemachten Unterscheidung zwischen gesetzesartigen und kontingenten Informationen in einem Text definieren als die Menge aller ge-

setzesartigen Relationen, in denen *W* im Text vorkommt. Diese Relationenmenge heie die *erste Umgebung* um *W*. - Dieser Bedeutungsauffassung liegt eine Bedeutungstheorie zugrunde (die hol-atomistische Bedeutungsauffassung der Leksemantik, Mudersbach 1983), bei der ein Wort *W* nicht als ein bilaterales Zeichen (bestehend aus Ausdrucksseite und Inhaltsseite) verstanden wird, sondern als ein mehrstufiges Bedeutungs-Gebilde (als "Leksem" zu *W*), dessen erste (Bedeutungs-)Stufe die eben definierte erste Umgebung des Wortes *W* ist. Und jede weitere Stufe ergibt sich aus der vorhergehenden, indem alle ersten Umgebungen der in der vorangehenden Stufe erreichten Begriffe zusammengenommen werden. Diese Auswertung der Umgebung um *W* lt sich soweit ausdehnen, bis das ganze Netz als die maximale Bedeutungsstufe zu *W* erscheint. Diese maximale Stufe charakterisiert den unverwechselbaren STELLENWERT (valeur) des Wortes *W* in diesem (Text-)Netz. Dies ist die *holistische Bedeutung* zu *W*. Jede Stufe darunter wird als *hol-atomistische* Bedeutungsstufe bezeichnet. Der *atomistische* Teil besteht aus dem Bedeutungsatom *W* (nullte Stufe) und der ersten Umgebung um *W*.

G4. Die Methode RELATEX erfat auch traditionelle Textordnungs-Parameter wie die *Thema-Rhema-Gliederung* eines Textes oder das Aufzeigen von *Isotopie-Linien* im Text. Beides sind Techniken, um die Kohrenz eines Textes zu berprfen und damit auch die eigene Verstehensmglichkeit vorzubereiten oder zu flankieren.

Die Thema-Rhema-Gliederung soll dazu dienen zu unterscheiden, was in einem Text als fr den Hrer schon bekannte Informationen vorausgesetzt wird und was vom Sprecher als neue Information intendiert wird. Diese Gliederung ergibt sich beim Anschlieen eines Satzes an das vorher analysierte Textnetz (also im 3. Schritt): die Begriffe, die im vorausgehenden Satz (bzw. der dazu erstellten Relation) schon vorhanden sind und die daher geeignete Anknpfungspunkte fr die neuen Relationen sind, sind die *Thematischen Teile* im neuen Satz, und diejenigen Teile der Relation, die dann brig bleiben, sind die *Rhematischen Teile*. Da bei RELATEX an jedem bisher vorhandenen Begriff im Text angeknpft werden kann, ergeben sich auch Anknpfungen an weiter zurckliegende Begriffe, so da schlielich Bgen gespannt werden durch den ganzen Text hindurch. Solche Linien zu einem bestimmten Begriff oder Thema machen die *Isotopielinien* aus. Je mehr unter irgendeinem gewhlten Aspekt gleichartige Begriffe zusammengefat werden, um so mehr werden die Isotopie-Linien zu einer *Isotopie-Ebene*. Diese Betrachtungstechniken sind geeignet, grere Zusammenhnge in einem Text zu (re)konstruieren. Damit sind in RELATEX auch Mglichkeiten angelegt, ho-

listische Gesichtspunkte mitzuerfassen. Gleichzeitig sieht man, daß auch ASPECTEX solche Isotopien aufzeigen kann, wenn ein Aspekt eine gleichmäßige Verteilung über den ganzen Text aufweist. Dies zeigt, daß der im folgenden zentral zu behandelnde Blickwinkel der holistischen Betrachtung auch an die anderen Methoden schon angelegt werden kann.⁹

2.3. Die dritte Methode: Analyse der Text-Systeme (HOLONTEX)

Diese Methode richtet sich auf die ganzheitlichen Gesichtspunkte, die mit einem Text verbunden sein können. Sie betreffen zum einen die Textgestalt und -gestaltung, so wie sie sich z.B. in verschiedenen Briefformen ausdrückt, - zum anderen betreffen sie die Gesamtvorstellungen (Systeme), die in einem Text vom Autor angesprochen werden, ohne jedoch im Text expliziert zu werden.¹⁰ Diese Systeme, die der Leser, für den der Text bestimmt war, aus seinem jeweiligen Hintergrundwissen aktivieren muß, um den Text zu verstehen, müssen vom Textinterpreten an den Text herangetragen werden als Hypothesen über die angedeuteten Hintergrundsysteme. Jedes dieser Systeme sowie auch die Textgestalt läßt sich wieder als Relationsnetz formulieren bzw. aus gesetzesartigen Texten nach der Methode RELATEX gewinnen.

Der Textverstehende muß die Kompetenz mitbringen: zum einen die Textgestaltung erbringen zu können, d.h., er muß verschiedene Formen von Textganzen gegeneinander abgrenzen und unterscheiden können, zum anderen muß er zum Textinhalt passende Hintergrundsysteme schon zur Verfügung haben oder hypothetisch formulieren können. Daraus ist ersichtlich, daß diese Methode die meiste Aktivität des Lesers erfordert und am stärksten von außen her die Textinformation mitbestimmt. Sie wird als Holistische Methode des Textverstehens bezeichnet (abgekürzt: HOLONTEX), weil verschiedene Arten von Ganzheiten (Ganzheit: HOLON) mit dem Text insgesamt in Verbindung gebracht werden sollen.

HOLONTEX ergänzt die beiden anderen Methoden gerade um die *Inter-*

⁹ Ein Beispiel der Anwendung von RELATEX findet sich in Mudersbach (1979), eine weitergehende Verwendung der Methode zur Analyse von Isotopien eines Textes in (Mudersbach, Gerzymisch 1989). Mehr zur Analyse von Gesetzen mit RELATEX findet sich in Mudersbach (1988).

¹⁰ Der hier verwendete System-Begriff hat Ähnlichkeiten mit den in der Künstlichen Intelligenz Forschung verwendeten Konzepten wie Frame (Minsky 1975) oder Script (Schank/Abel 1977). "System" soll hier jedoch nicht in der starren reglementierten Weise verstanden werden, wie es für die Programmierung erforderlich ist, es soll vielmehr der Bewußtmachung von Hintergrundwissen dienen, auch wenn dies individuell verschieden ausgeprägt sein kann.

pretationsdimension, bei der der Textinterpret am besten seine Beziehung zum Text ausdrücken und seine Sicht an den Text anlegen kann, eine Sicht, die natürlich vom *geschichtlichen Standort* des Lesers stark geprägt ist. Die Schrittfolge bezieht sich hier, wie bei ASPECTEX, auf den ganzen Text, nicht auf einzelne Textstellen.

HOLONTEX-SCHRITTFOLGE (4 Schritte):

Vorgaben: der Text, Eingangsinformationen, Zweck der Lektüre und die Kompetenz des Lesers, Systeme erkennen zu können.

1. Schritt: Erstlektüre des Textes unter holistischen Gesichtspunkten

In Verbindung mit der Erstlektüre-Empfehlung zu ASPECTEX (erster Schritt dort) versucht der Textinterpret beim ersten Lesen des Textes die für den Text relevanten Systeme zu ermitteln, sei es Systeme, die die Textgestalt betreffen, sei es solche, die die inhaltliche Interpretation betreffen.

2. Schritt: Aufstellen der Liste der Systeme

Die so vermuteten Systeme werden aufgrund des eigenen Kenntnisstandes formuliert. Das explizite Bereitstellen von solchen Systemen ist deswegen besonders wichtig, weil sich erst durch den Kontrast zwischen der systemhaft angesetzten Information aus dem Hintergrundwissen ("System-Skelett") und der Information, die im Text enthalten ist ("System-Auffüllung"), der Stellenwert der Textaussagen ergibt.

3. Schritt: Holistisches Lesen des ganzen Textes

Ähnlich wie in ASPECTEX wird jetzt aus der Liste der Systeme *ein* System ausgewählt und an den Text angelegt. Es werden die Textstellen markiert, die einen bestimmten Systemteil entweder ansprechen und ergänzen oder ihm widersprechen. Dieser Durchlauf wird nacheinander für alle Systeme durchgeführt.

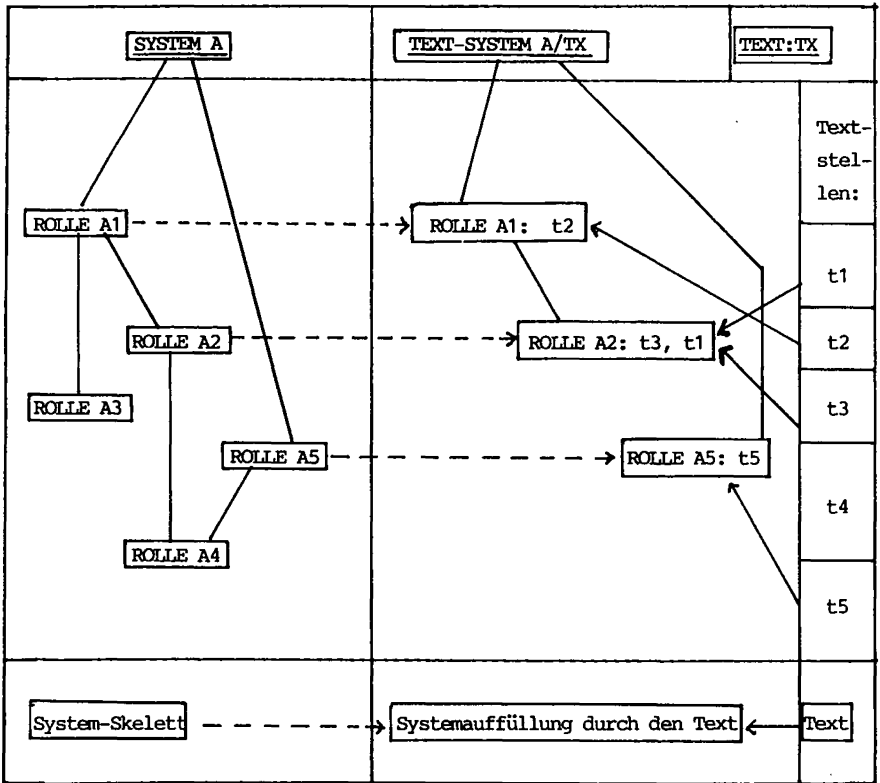
4. Schritt: Gewichtung der System(teil)e nach dem Zweck

Wird der Zweck der Lektüre interpretiert als eine Auswahl unter den Systemen, die für die Lektüre relevant sein sollen, dann läßt sich wie bei den vorhergehenden Methoden eine Präferenzreihenfolge angeben für die Systeme, die primär zu betrachten sind.

Resultat:

Betrachten wir zunächst einmal das Resultat der Anwendung dieser Methode für nur ein gewähltes System: wir erhalten ein textbezogenes System, das entstanden ist aus dem Systemskelett, das aus dem Hintergrundwissen stammt, und der System-Auffüllung an den jeweilig geeigneten Stellen um die textspezifischen Informationen (vgl. Abb. A).

Abb.A: Beispiel für das Zusammenspiel von System-Skelett und System-Auffüllung (HOLONTEX, Resultat)



Verschiedene textbezogene Systeme können ineinander greifen, weil sie an derselben Textstelle ansetzen. Während bei den beiden anderen Methoden der Text in seinen verschiedenen Aspekten oder im informativen Charakter betrachtet und analysiert wurde, die Betonung also mehr auf der Sicht des Textes als auf der Interaktion mit dem Wissen des Lesers lag, sind bei HO-

LONTEX die verschiedenen textbezogenen Systeme im Text miteinander verflochten in Konfigurationen, die ein Gesamtsystem ausmachen (vgl. die Beispielanalyse im Anhang und Abb. 3).

Damit ist die Darstellung der Methoden abgeschlossen. Es folgen noch Bemerkungen zur Berücksichtigung des Interesses des Lesers (2.4) und zur Computerunterstützung (2.5).

2.4 Zur Gewichtung nach Interessen (Lektürezweck) und zu den Anwendungsmöglichkeiten

Für jede der drei Methoden gilt hinsichtlich der Einbeziehung des Lektürezwecks bzw. der weiteren Aufgaben wie *Textbeurteilung* oder *-bearbeitung*: sie führen jeweils zu einer Gewichtung der methodenspezifischen Parameter bzw. der dazugehörigen Aussagen. Also:

für die 1. Methode: die Gewichtung der Aspekte bzw. bestimmter präferierter Aspektwerte,

für die 2. Methode: die Gewichtung bestimmter Relationen bzw. bestimmter Begriffe oder Teilnetze im Gesamtnetz,

für die 3. Methode: die Gewichtung bestimmter holistischer Systeme bzw. deren inhaltlicher Auffüllung durch den Text.

Je nach Gewichtung ergibt sich eine Teilansicht des Textes, nämlich die, die der vorgegebenen Interessenverteilung beim Leser (dem Historiker) entspricht bzw. die die Grundlage eines Interpretationsvergleiches, einer Textbeurteilung oder Textbearbeitung (inklusive Übersetzung)¹¹ bilden kann.

Zu diesen verschiedenen Anwendungsmöglichkeiten noch eine Bemerkung: *eine* Vorgehensweise für den Leser/Historiker ist die, daß er einen neuen Text *von vomherein* nach einer der Methoden zu lesen beginnt (das ist ja besonders wichtig für die Erstlektüre), d.h. daß er sein Textverstehen gleich systematisch flankiert und kontrolliert (*methodengeleitetes Verstehen*). Eine weitere Anwendung ist die, eine vorhandene (fremde oder eigene) Interpretation des Textes nachzuvollziehen anhand einer der Methoden. Hierbei kann überprüft werden, wie systematisch der betreffende Interpret tatsächlich vorgegangen ist (*Interpretationsbeurteilung*). Die weitere Anwendung beim Interpretationsvergleich ergibt sich daraus.

¹¹ Die Methode ASPECTEX ist wie schon erwähnt zu einer Übersetzungsmethode (ASPECTRA) erweitert worden. Die anderen beiden Übersetzungsmethoden sind in Arbeit.

Die Gewichtung ist bei jeder Methode erst in einem vierten Schritt angegeben. Dadurch wird die Spannweite der Methode erst einmal maximal entfaltet und danach erst durch den Lese-Zweck eingeschränkt. Es ist aber durchaus verständlich, wenn der praktisch denkende Leser sich sofort auf seinen Lektüre-Zweck konzentrieren will und von vornherein nur die dafür relevanten Gesichtspunkte ins Auge faßt. Er nimmt dabei in Kauf, daß eine Reihe von Gesichtspunkten des Textes, von denen er *vermutet*, daß sie für seinen Zweck eventuell keine Rolle spielen, überhaupt nicht ins Blickfeld kommen, auch wenn sie unerwartet vielleicht doch etwas zum Lektürezweck beitragen könnten. Für diesen praktisch denkenden Leser ist jeweils eine "*schnelle*" *Alternative* in den Methoden denkbar, bei der der vierte Schritt schon vor dem zweiten eingebaut wird. Dadurch wird der Aufwand bei den beiden nachfolgenden Schritten verringert (vgl. Anhang, ASPECTEX, 4. Schritt und HOLONTEX, 2. Schritt).

2.5. Zur Computerunterstützung der Methoden

Zur Methode ASPECTEX gibt es, wie schon erwähnt, ein unterstützendes Computerprogramm (ASPECTRA); zu RELATEX ebenfalls (siehe dort 3. Schritt); das zu HOLONTEX ist noch in Planung.

Das jeweilige Computerprogramm dient dazu, folgende Aufgaben zu erfüllen:

- es gibt je nach Methode einen Entscheidungspfad vor und zeigt die Alternativen, aus denen der Historiker dann eine Entscheidung auswählen kann,
- es führt eventuell quantitative Auswertungen durch, die sich aus den Gewichtungen und Präferenzen ergeben, und
- es zeigt die Entscheidungen "synchron" systematisiert an und stellt evtl. graphische Veranschaulichungen bereit (um dem Kriterium der Transparenz des Resultates zu genügen).

Bei dieser Art der Computernutzung, die stark von der Individuenabhängigkeit der Entscheidungen geprägt ist, kommt der Computer *nicht als Ersatz* für die Erschließung des Textes zum Tragen, sondern als Unterstützung und Ergänzung der menschlichen Entscheidung.

Computerprogramme, die als "Ersatz" fungieren sollen, können entweder nur Aufgaben lösen, die "objektiv" durchführbar sind, wie das Auszählen von Wörtern (siehe 1.3) oder Entscheidungen treffen, deren Objektivität schon garantiert ist. Bei dieser Funktion des Computers wird der Historiker unzufrieden sein, wenn er die genuine Aufgabe darin sieht, sich selbst als Wissenschaftler mit dem historischen Text-Material zu konfrontieren, statt der individuenfreien Starrheit einer vermeintlichen Objektivität zu huldigen.

3. Zusammenspiel und Vereinheitlichung der drei Methoden

Wie verhalten sich die drei Methoden zueinander? Und zu welchem Zweck sollten die Methoden jeweils eingesetzt werden? - Die am genauesten auf den Text bezogene ist die Methode RELATEX. Sie erfaßt jedoch nur den *semantisch erfäßbaren Informationsgehalt* des Textes. Eine Beurteilung weiterer Texteigenschaften geschieht im Rahmen von ASPECTEX. Hier sind *alle relevanten Gesichtspunkte* anzulegen, die sich für den Textinterpreten aus dem Text ergeben. Die dritte Methode, HOLONTEX rundet das Bild schließlich ab, indem sie aus dem Text ein *Ganzes* macht, aber nicht nur *ein Ganzes*, sondern ein Ganzes mit verschiedenen ganzheitlichen Dimensionen, je nachdem wieviele *Systeme* bzw. *Gestaltungsaspekte* betrachtet werden. Daher ergänzen sich die drei Methoden. Der Historiker kann nun sein Interesse daraufhin befragen, ob er mehr an den konkreten inhaltlichen Informationen eines Textes interessiert ist (RELATEX) oder ob er den Text einer Detailbeurteilung unterziehen will, bei der es auf die verschiedenen Nuancen der Darstellung ankommt (ASPECTEX), oder ob er den Text in Gestaltung und Funktion unter einem oder mehreren von ihm herangetragenen Systemen einzustufen wünscht (HOLONTEX).

Wie könnte die Vereinheitlichung der drei Methoden aussehen? Sicher kann sie nicht in einer inhaltlichen Gleichschaltung bestehen, denn dann würde gerade das ergänzende Moment verloren gehen. Man kann sie aber formal subsumieren unter die Methode, Aspekte an den Text anzulegen, wenn man neben den Aspekten, die bei ASPECTEX angelegt werden, auch noch die für die beiden anderen Methoden relevanten Gesichtspunkte zu Aspekten erklärt. Wenn man also den relationalen Charakter des Textes als einen Aspekt ansieht, wird die Methode RELATEX formal unter die Aspekt-Methode integriert. Ebenso läßt sich HOLONTEX in die aspektive Methode formal einbeziehen, wenn man die Systemhaftigkeit bzw. die einzelnen Systeme jeweils zu Aspekten erklärt.

Wie können die drei Methoden konzentrierend, d.h. gleichzeitig und sich ergänzend, angewandt werden? Sie können nicht pro Textstelle nebeneinander durchgeführt werden, da dies gerade dem Aspektiven Lesen des ganzen Textes und dem Holistischen Lesen widersprechen würde, nicht jedoch der RELATEX-Methode. Erst die Ergebnisse aus den drei Methoden können so zusammengestellt werden, daß *pro Textstelle* die Informationen aus allen drei Betrachtungsweisen zusammengenommen werden können. Damit ergibt sich auch eine umfassende Grundlage zur Beurteilung dieser Textstelle im Rahmen einer Gesamtbeurteilung.

4. Zusammenfassung und abschließende Bemerkung

Wir sind ausgegangen von der Frage nach dem Interesse des Historikers an Texten und welche linguistischen Methoden der Texterschließung dabei dienlich sein können. Die Beantwortung der Frage wurde von vornherein unter zwei Randbedingungen gestellt, die meines Erachtens für die Tätigkeit des Historikers wichtig sind:

- einmal die durchgängige Reflektierung des Stellenwertproblems, d.h. einen Text aus seiner Zeit heraus zu verstehen (unter Vermeidung einer "Teleskopsicht" von heute aus), und
- zum anderen die fruchtbare Verbindung der individuellen Sicht des einzelnen Historikers mit den geisteswissenschaftlich angemessenen Kriterien von Wissenschaftlichkeit.

Die hier vorgeschlagenen Methoden versuchen, diesen Randbedingungen gerecht zu werden und dem Interesse des Historikers zu dienen. Jede Methode enthält sowohl einen Vorschlag, wie der individuelle Textverstehende sein Verstehen in Form von individuellen Entscheidungen umsetzen kann als auch wie durch die Explizitheit der Methode gleichzeitig gewährleistet ist, daß die Entscheidungen transparent und nachvollziehbar bleiben.

Hier ist noch eine Bemerkung angebracht hinsichtlich des Arbeitsaufwandes, den diese Methoden mitbringen. Es gibt sicher Wege, einen Text schneller zur Kenntnis zu nehmen. Bei den hier dargestellten Methoden geht es nicht um zeitliche Effizienz, wie es häufig in der Praxis (bei Gutachten usw.) verlangt wird. Hier geht es primär darum, vorzuführen, daß das Textverstehen nicht nur "hermeneutisch" in einem *schlechten* Sinn vonstattengehen muß, nämlich als eine *subjektive*, fast private Aneignung des Textes, verbunden mit einer Meinung über das Gelesene, die nur bei genügend Autorität unbesehen respektiert werden wird. Hier geht es also darum, ob es *möglich* ist, nach *wissenschaftlichen* Methoden zu einer fundierten und nachvollziehbaren Beurteilung von Texten zu gelangen. Daß ein solches Vorgehen mit einem gewissen Mehraufwand an Gründlichkeit verbunden ist, ist zu erwarten. Dafür hat man aber auch die Gewähr, alles einem selbst zu Gebote stehende Wissen auf systematische Weise einbringen zu können.

Um jedoch einem Mißverständnis vorzubeugen: die Methoden können nicht garantieren, daß ein Problem nun ohne Kontroverse und Alternativen behandelbar ist (ein Hammer kann auch nicht garantieren, daß ein Nagel richtig in die Wand eingeschlagen wird), wohl aber können diese Methoden dazu dienen, die unterschiedlichen Ansätze und Argumentationen leichter und unpolemischer miteinander zu vergleichen. Wenn man den Nutzen eines Instrumentes erkennt, kann man dafür auch Nägel mit Köpfen machen... -

ANHANG: Drei Beispielanwendungen

Im folgenden sollen nun die drei Methoden auf einen Text angewandt werden, um die Funktionsweise der einzelnen Methoden und ihr Zusammenspiel zu veranschaulichen. Ich habe dazu einen Text ausgewählt, der sowohl historisch als auch sprachlich interessant sein dürfte, und der genügend Möglichkeiten bei allen drei Methoden eröffnet: Kleists Anekdote "Franzosen-Billigkeit" aus den Berliner Abendblättern vom 3. 10. 1810.

Wo die Analysen aus Platzgründen nicht ganz so ausführlich dargestellt sind, wie es für eine exemplarische Darstellung der Methoden erforderlich wäre, weise ich darauf hin. Nach den Vorgaben, die für alle drei Methoden einheitlich sind, werden die einzelnen Methoden veranschaulicht.

VORGABEN:

Text:

Franzosen-Billigkeit
(werth in Erz gegraben zu werden.)

Zu dem französischen General H u l i n kam, während des Kriegs, ein.... Bürger, und gab, Behufs einer kriegsrechtlichen Beschlagnahme, zu des Feindes Besten eine Anzahl, im Pontonhof liegender, Stämme an. Der General, der sich eben anzog, sagte: Nein, mein freund; diese stämme können wir nicht nehmen. - "Warum nicht?" fragte der Bürger. "Es ist königliches Eigenthum." - Eben darum, sprach der General, indem er ihn flüchtig ansah. Der König von Preußen braucht dergleichen Stämme, um solche Schurken daran hängen zu lassen, wie er. -

Eingangsinformation:

- Autor des Textes: H.v. Kleist,
- Erscheinungsort: erschienen in den Berliner Abendblättern vom 3. 10. 1810, S. 13,
- Quelle: zitiert nach Sembdner (1982),
- Darstellung: in Schreibweise, Interpunktion und Zeilenbruch wie im Original.

Zweck der Lektüre: Exemplifizierung dreier Texterschließungsmethoden.

Anwendung der Methode ASPECTEX (vgl. Abb. 1a und 1b):

1. Schritt: Erstlektüre

In der Abb. 1a ist zur Veranschaulichung am ersten Satz eine Liste angelegt mit den Spalten:

- Erwartungsaufbau: was aufgrund der gelesenen Zeile erwartet wird,
- Erwartungsbilanz: welche der Erwartungen bestätigt oder enttäuscht werden und welche weiterhin beibehalten werden,
- Überraschungen: was überrascht am Text, was fällt auf,
- Welche Aspekte könnten von Interesse sein?

Abb. 1a: Kleists Anekdote "Franzosen-Billigkeit.": Tabelle zur Erstlektüre nach ASPECTEX (Titel und erster Satz).

	Erw.aufbau	Erw.Bilanz	Überschungen	Aspekte
Franzosen-Billigkeit. (werth in Erz gegraben zu werden.)	Bericht über Ausländer von exemplarischem Charakter von moralische Wert.		"Billigkeit" ein veraltetes Wort.	Interpunktion sehr genau und ungewöhnlich
Zu dem franz. General Hulin kam, während des Kriegs, ein...Bürger und gab, Behufs einer kr.r. Beschl.nehmung zu des Feindes Besten, eine Anzahl im Pontonhof liegender Stämme an	Es handelt sich um militärische Aktionen? der Bürger ist suspekt. Ist er Preuße? überreicht er ein geheimes Dokument? Jetzt muß sich bald klären, was er ihm gab.	Hulin ist der Franzose Moralischer Wert noch nicht zu erkennen - Kein Dokument! Holz! enttäuschend! -	warum "...?" durch die Einschübe wird die Spannung immer größer! ungewohntes Fremdwort	Interpunktion zu verschiedenen Zwecken Position der Akteure: Hulin und Bürger kontrollieren.
	Hulin wird die Stämme zur Flußüberquerung brauchen. Der Pontonhof wird noch eine Rolle spielen	Es war nicht "geben", sondern "angeben" - - beide Erwartungen werden enttäuscht werden	Aber das ist ja Landesverrat! Ohne daß es klar ausgesprochen wird! Warum?	

2. Schritt: Aufstellen der Aspektliste:

Unter dem Gesichtspunkt, daß die Anekdote nicht primär unter linguistischen Gesichtspunkten zu betrachten ist, habe ich folgende mehr inhaltlich orientierte Aspekte gewählt:

- Sicht des Bürgers vom Geschehen,
- Sicht des Generals vom Geschehen,
- Perspektive des Autors und des Lesers,
- Erkenntnisse über die Situation
- Anekdoten-Merkmale
- Kleists Sprache

Die Aspektwerte zu den einzelnen Aspekten lassen sich für die hier gewählten Aspekte nicht in reglementierter Form vorgeben, da sie teilweise dem Gang der Handlung folgen und diesen ad hoc unter dem Aspekt kommentieren. Für linguistische Aspekte wie den der Kleistschen Sprache oder die Anekdoten-Merkmale (siehe auch unter 4. Schritt), läßt sich jedoch das Feld der Aspektwerte vorgeben. Hier wird zur Exemplifizierung folgender gewählt:

ASPEKT: Kleists Sprache:

Werte:

- "Behufs" + Verbalsubstantiv auf -ung
- in Kommata eingeschlossene adv. Bestimmung
- in Kommata eingeschlossene attrib. Bestimmung
- Begleitumstände der Antwort werden charakterisiert
- Nur die Rede von einer Person (hier des Bürgers) steht in Anführungszeichen. Die des andern nicht.
- absolute Endstellung der Pointe

3. Schritt: das Aspektive Lesen ist anhand der gewählten Aspekte durchgeführt und in Abbildung 1b dargestellt.

Abb. 1b: Kleists Anekdote "Franzosen-Billigkeit."
 Aspektives Lesen bestehend aus Abb. 1b1. und 1b2.

Aspekt: Perspektiven von: A: Autor, L: Leser	<u>Text der Kleistschen Anekdote</u> <u>"Franzosen-Billigkeit."</u>
0 A bewertet für den Lesert schon im Titel die Anekdote 0.1 A's Bewertung wird noch deutlicher im Untertitel mit dem Appell an L zu derselben Einstellung	0 Franzosen-Billigkeit. 0.1 (werth in Erz gegraben zu werden.)
1.1 neutral	1.1 Zu dem französischen General H u l i n
1.2 neutral	1.2 kam,
1.3 A setzt die Kenntnis des Krieges voraus	1.3 während des Kriegs,
1.4 A macht deutlich, daß er hier eine Information wegläßt	1.4 ein Bürger,
1.5 A läßt L erwarten: der Bürger gibt Hulin etwas.	1.5 und gab,
1.6 A wählt einen absichtlich "umständlichen" Ausdruck	1.6 Behufs
1.7	1.7einer kriegsrechtlichen Beschlagnehmung
1.8	1.8 zu des Feindes Besten
1.9	1.9 eine Anzahl,
1.10 Während der Bürger Stämme zur Übernahme anbietet, übernimmt der	1.10 im Pontonhof liegender,
1.11 Autor das "Ponton" aus der Sprache des Feindes, vielleicht um	1.11 Stämme
1.12 die Anbiederung auch sprachlich perfekt zu machen.	1.12 an.
2.1	2.1 Der General,
2.2 A betont, wie unwichtig die Sache für den General ist.	2.2 der sich eben anzog,
2.3	2.3 sagte:
2.4	2.4 Nein, mein Freund;
2.5	2.5diese Stämme können wir nicht nehmen. -
3.1	3.1 "Warum nicht?"
3.2	3.2 fragte der Bürger.
4.1	4.1 "Es ist königliches Eigenthum." -
4.2	4.2 Eben darum, sprach der General,
4.3 A betont(s.o.), wie unwichtig die Sache für den General ist.	4.3 indem er ihn flüchtig ansah.
5.1	5.1 Der König von Preußen
5.2	5.2 braucht dergleichen Stämme,
5.3	5.3 um solche Schurken daran hängen zu lassen
5.4	5.4 wie er. -

Abb.1b:
 Aspektives Lesen, veranschaulicht
 an der Kleist-Anekdote (jeweils
 2.Spalte in den beiden Teilen).

<u>Aspekt: Sicht des Bürgers B</u> <u>vom General G</u>	<u>Aspekt: Sicht des Generals G</u> <u>vom Bürger B</u>
0	
0.1	
<p>1.1 B wird sich ein Bild von G machen</p> <p>1.2 B sucht G auf</p> <p>1.3</p> <p>1.4 Setzung der Sicht des B</p> <p>1.5</p> <p>1.6 B: wichtiguerische Sprache (falls seine Formulierung)</p> <p>1.7 B: G hat das Recht der Beschlagnehmung</p> <p>1.8 B glaubt, daß (nach Kleist: der Feind) G es so sehen wird</p> <p>1.9</p> <p>1.10</p> <p>1.11 B: für G wertvolles Material im Krieg</p> <p>1.12 B: "angeben" ist nicht "verraten"?</p> <p>2.1 B erwartet eine Belohnung.</p> <p>2.2</p> <p>2.3</p> <p>2.4 B wird als Freund angesprochen</p> <p>2.5</p> <p>3.1 B versteht die Ablehnung des G nicht</p> <p>3.2</p> <p>4.1 Königl. Eigentum gibt eher die Berechtigung zum Beschlagnehmen als anderer Eigentum.</p> <p>4.2</p> <p>4.3</p> <p>5.1</p> <p>5.2</p> <p>5.3 B sieht sich plötzlich als schurke beurteilt</p> <p>5.4 und als schändlich zurückgewiesen</p>	<p>Setzung der Sicht des Generals</p> <p>)</p> <p>:</p> <p>G: fährt trotz der "wichtigen" Mitteilung -im Alltagsgeschäft fort</p> <p>G meint die Anrede ironisch. G gibt eine entschlossene Antwort, ohne B zu bewerten. Aber das "nicht können" scheint ethisch motiviert zu sein, also nicht nur "nicht wollen".</p> <p>G bemerkt: B scheint sich seines Tuns nicht bewußt zu werden.</p> <p>G zeigt seine Geringschätzung.</p> <p>G ist überzeugt, auch ohne Hilfe durch Landesverrat den Krieg gewinnen zu können.</p> <p>G drückt endlich deutlich seine Verachtung aus.</p>

<u>Aspekt: Erkenntnisse über die Situation</u>	<u>Text der Kleistschen Anekdote</u> <u>"Franzosen-Billigkeit."</u>
0 man konnte auch in Berlin einen Franzosen loben 0.1 zumal es mit einer Warnung an die eigenen Bürger verbunden wurde	0 Franzosen-Billigkeit. 0.1 (werth in Erz gegraben zu werden.)
1.1 1.2 es war also möglich von einem feindlichen General 1.3 empfangen zu werden (als preuß.Bürger) 1.4 Dieser Bürger konnte also unbeobachtet ins feindliche Lager gelangen. 1.5 1.6 Er hatte Kenntnisse über kriegsdienstliches Material 1.7 das Kriegsrecht sieht Beschlagnahmen vor 1.8 1.9 1.10 Pontonhof: eventuell zum Brückenbau geeignete. 1.11 Stämme waren offensichtlich kriegswichtiges Material 1.12 "angeben" besagt hier: Verrat an den Feind 2.1 Der franz.General empfing d.Bürger offensichtlich am Morgen 2.2 ohne Angst vor dem Feind zu haben. 2.3 2.4 das Sprachproblem ist hier ausgelassen 2.5 3.1 3.2 4.1 Es besteht ein Unterschied zwischen Eigentum des Königs und dem der Bevölkerung(1). 4.2 4.3 5.1 5.2 5.3 Diese Tat wird in Preußen mit dem Tod durch 5.4 Erhängen bestraft.	1.1 Zu dem französischen General H u l i n 1.2 kam, 1.3 während des Kriegs, 1.4 ein Bürger, 1.5 und gab, 1.6 Behufs 1.7einer kriegsrechtlichen Beschlagnehmung 1.8 zu des Feindes Besten 1.9 eine Anzahl, 1.10 im Pontonhof liegender, 1.11 Stämme 1.12 an. 2.1 Der General, 2.2 der sich eben anzog, 2.3 sagte: 2.4 Nein, mein Freund; 2.5diese Stämme können wir nicht nehmen. - 3.1 "Warum nicht?" 3.2 fragte der Bürger. 4.1 "Es ist königliches Eigenthum." - 4.2 Eben darum, sprach der General, 4.3 indem er ihn flüchtig ansah. 5.1 Der König von Preußen 5.2 braucht dergleichen Stämme, 5.3 um solche Schurken 5.4 daran hängen zu lassen wie er.-

<u>Aspekt: Anekdoten-Merkmale</u>	<u>Aspekt: Kleists Sprache</u>
0 0.1	0 von Kleist bewußt eingesetzte Interpunktion. 0.1
1.1 direkter Einstieg in die Szene. Eigenname zur Plastizität 1.2 knappe Charakterisierung der Handlung 1.3 knapper direkter Bezug auf ein Ereignis 1.4 dagegen: merkwürdige Ungenauigkeit (eigentlich nicht-spezifisch für Anekdoten, daher auffallend im Kontrast) ebenso: merkwürdige Umständlichkeit muß einen Grund jenseits des anekdotischen Gestaltung: cf. 1.8 1.5 der Schwung der Anekdote wird gebremst 1.6 Die gegen den anekdotischen Duktus eingesetzte Umständlichkeit läßt sich als sprachliche Gestaltung der Umständlichkeit des Bürgers ansehen. 1.7 1.8 1.9 1.10 1.11 1.12 Das Verb wird "überdehnt". Ein verhältnismäßig harmloses Satzende. 2.1 2.2 dagegen: minimale Charakterisierungen der Situation. 2.3 Beschränkung auf das wesentliche in einer Situation. 2.4 Engführung der direkten Reden (im folgenden: 2.5, 3.1, 4.1, 4.2). 2.5 3.1 3.2 4.1 4.2 4.3 Vgl. 2.2 5.1 5.2 Die Pointe wird durch einen Satz + Nebensatz vorbereitet. 5.3 Pointe kommt erst im letzten Wort des Nebensatzes. 5.4	1.1 Dieser Satz "gestaltet" die umständliche, wichtig-tuerische Art des Bürgers, bei dem erst am Ende aus harmlosen "geben" ein Verraten wird 1.2 In Kommata eingeschlossene adv. Bestimmung 1.3 1.4 1.5 1.6 Behufs+ Verbalsubstantiv auf -ung (Kleist-spezifisch) 1.7 in Kommata eingeschlossene adv. Bestimmung 1.8 in Kommata eingeschlossene adv. Bestimmung 1.9 1.10 in Kommata eingeschlossene attrib. Bestimmung 1.11 1.12 Ende der umständlichen Klammer: euphemistisch pointiert 2.1 2.2 Begleitumstände der Antwort werden charakterisiert: die Nebenhandlung ist aber die für den General wichtigere. 2.3 Nur die Rede von einer Person (hier des Bürgers) steht in Anführungszeichen, die des andern nicht: deutliches Unterscheiden durch 2.4 Interpunktion mit dem Effekt weiterer Verdichtung, da die direkten Reden nicht eingeleitet werden müssen. 2.5 3.1 3.2 4.1 4.2 Begleitumstände der Antwort werden charakterisiert: die "Nebenhandlung" ist aber die für den General wichtigere. 4.3 5.1 5.2 5.3 5.4 meisterhaft gelungen ist die absolute Endstellung der Pointe

4. Schritt: Die Gewichtung unter dem gewählten Lektüre-Zweck geht hier unsichtbar mit ein: es sind nur die Aspekte aufgeführt, die unter meinem Zweck von mir stark gewichtet worden waren. Andere Aspekte, deren linguistische Auswertung unter anderen Lektüre-Zwecken durchaus sinnvoll wäre, wurden hier gestrichen, nämlich:

- Interpunktion und Orthographie
- Auffälligkeiten in der lexikalischen Bedeutung (auch Wortbildung)
- Syntaktischer Aufbau
- Sprachschicht
- argumentativer Textaufbau
- Referenz auf die Personen und Objekte
- Darstellung von Ort, Zeit und Situation
- Wertung des Dargestellten in der Darstellung.

Dies zeigt auch, daß bei einer Beschleunigung des Aspektiven Lesens von vornherein die Wahl der Aspekte schon unter den Zweck der Lektüre gestellt werden kann, um für diesen Zweck weniger relevante Aspekte auszuklamern (vgl. dazu die Bemerkung in 2.4).

Anwendung der Methode RELATEX (vgl. Abb. 2):

VORGABE: siehe vor ASPECTEX

1. Schritt: Lexikon-Eintrag und Klassifikation

Das Lexikon für die ersten Zeilen lautet:

Argumente: Franzosen-Billigkeit, der frz. General, ein Bürger, eine Anzahl Stämme.

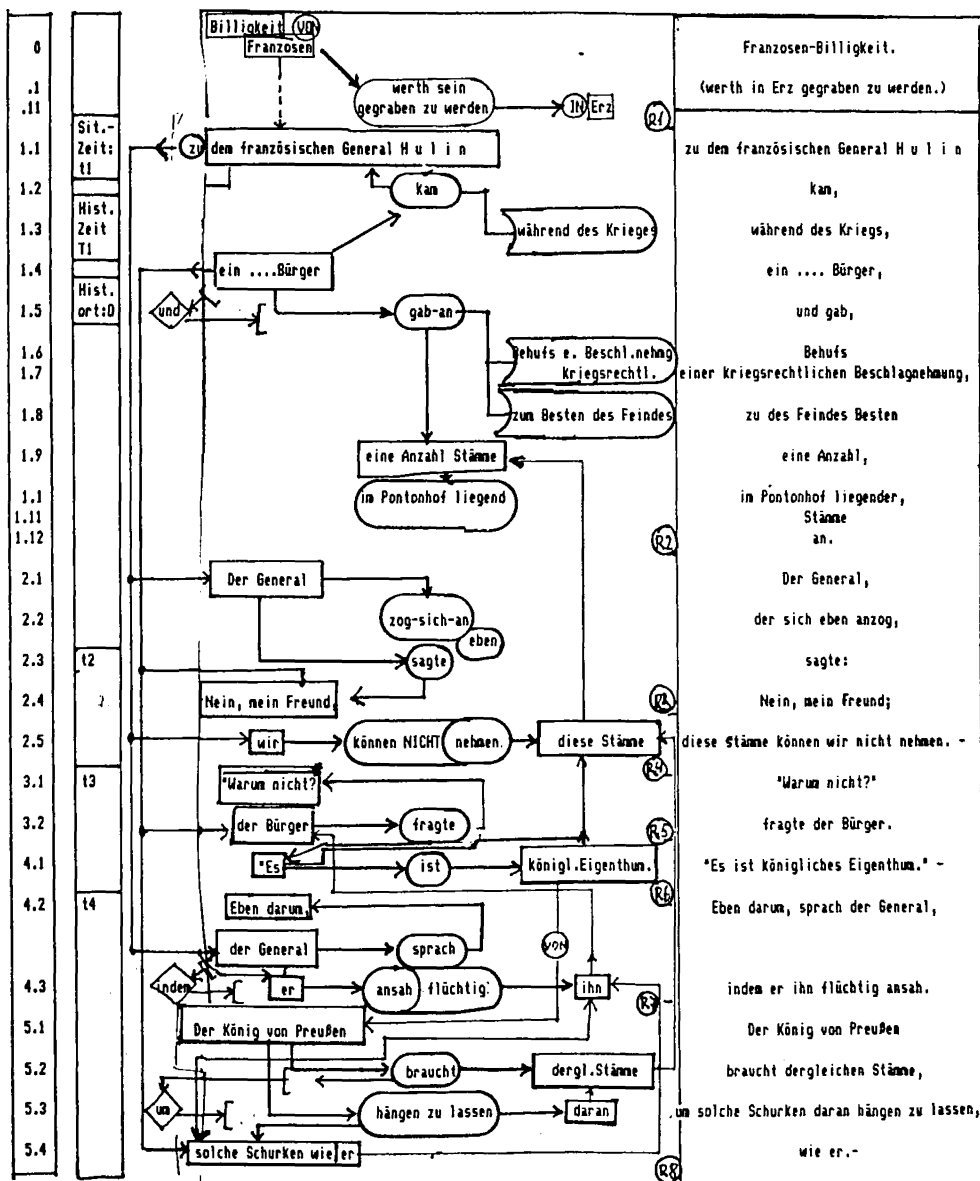
Bei den Relatoren werden die im Text vorkommenden Argumentstellen durch das entsprechende Fragepronomen bzw. die Präposition markiert, damit deutlich wird, welche Stellen wie besetzt werden können.

Relatoren: "werth sein in Erz gegraben zu werden" (dieser Ausdruck wird hier als ein Relatorausdruck gewählt, da er als idiomatisch angesehen werden kann und da im weiteren Text keine Teile dieses Ausdrucks isoliert vorkommen), kam (WER, ZU), gab-an (WER, WAS, WOZU), im Pontonhof liegend (WER).

Relationsmodifikatoren: während des Kriegs, Behufs einer..., zum Besten des Feindes,

Relationsverknüpfers: und

Abb. 2: Kleists Anekdote "Franzosen-Billigkeit.", analysiert mit der Methode RELATEX.



2. Schritt: Erstellen der Relationen

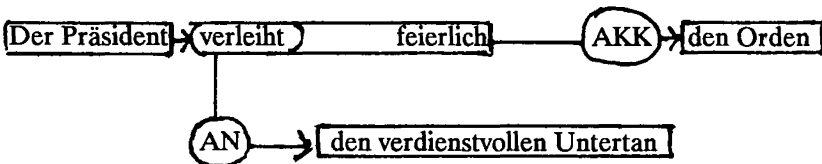
Hierzu muß zunächst der graphische Aufbau der Abbildung 2 kommentiert werden:

Auf der rechten Seite ist der Text der Anekdote dargestellt. In der linken Spalte eine Nummerierung, die für alle drei Methoden einheitlich gewählt wurde, um Querverbindungen ziehen zu können. Die Spalte rechts daneben enthält die sich aus dem Text ergebenden Angaben über die historische Zeit (T1) bzw. den historischen Ort (O). Damit ist die Einordnung in eine geschichtliche Epoche bzw. der historische Ort des Kriegsgeschehens gemeint. Da die Angaben dazu unbestimmt sind, werden hier nur die Konstanten T1 und O verwendet. Sie ändern sich während der Erzählung nicht. Davon ist der zeitliche und örtliche Verlauf des Dargestellten zu unterscheiden. Der zeitliche Verlauf ist angegeben durch die Markierung der aufeinander folgenden Zeitintervalle t1 bis t4. Über die Dauer eines Intervalls wird nichts ausgesagt. Als Kriterium für die Abgrenzung der Intervalle bietet sich hier der Redewechsel an.

Der verbleibende mittlere Bereich enthält die Darstellung der semantischen Relationen, die sich aus dem Text ergeben. Dazu wurden folgende graphische Symbole gewählt, die hier an dem Satz

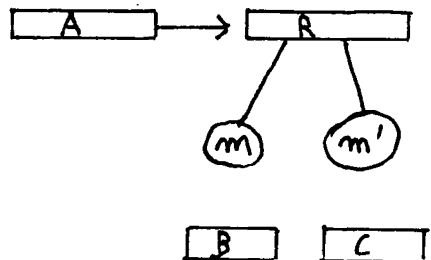
"Der Präsident verleihete feierlich den Orden an den verdienstvollen Untertan"

dargestellt werden soll:



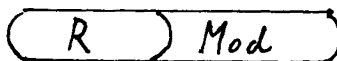
D.h. im allgemeinen wird der Relator als Oval dargestellt, die Argumente als Kästchen. Die einzelnen Argumente werden voneinander dadurch unterschieden:

- der Subjektausdruck zeigt mit dem Pfeil auf das Verb
- vom Verb ausgehend zeigen die Pfeile auf weitere Argumente. Wenn mehr als ein Argument anschließt, werden die Pfeile mit Marken (Kreise mit Eintrag des



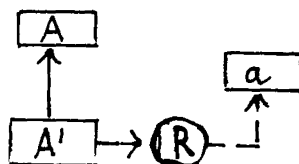
Kasus oder der Präposition)
versehen.

- Verbmodifikatoren ("feierlich") werden durch Halbovale an das Verb angehängt
- Konjunktionen (im Text "und", "um", "indem") werden in Rauten geschrieben, die Sätze miteinander verklammern.
- Pfeile, die Kästchen direkt miteinander verbinden (also ohne über ein Verb-Oval zu laufen), geben die sich aus dem Text ergebenden Rück- oder Vorverweise auf die Bezugsausdrücke an.



Relation 1

Relation 2



Gestrichelte Linien sind Hypothesen des Lesers, die er beim Textverstehen benutzt. Solange es um die Auflösung von Pronomen etc. geht, sind dies meist, aber nicht immer relativ triviale Hypothesen. Alle Hypothesen des Lesers müssen hinreichend unterschieden sein gegenüber den aus dem Text gewonnenen Relationen.

Jetzt kann der 2. Schritt erklärt werden: Zunächst wird zum Titel (Textstelle 0) die Relation (R1) formuliert. Sie besteht aus einem komplexen Begriff und einem komplexen Relator.

Die Textstelle 1. 1 - 1.2 besteht aus zwei durch "und" verknüpfte Relationen, zu denen in der Spalte links die historische Zeit, der historische Ort und die aktuelle Zeit t_1 angegeben sind. Bei der Darstellung der Relationen wurde die Wortfolge im Text nach Möglichkeit beachtet (außer bei der Zusammenziehung abtrennbarer Verben wie "gab-an"). Die Darstellung der Relationen selbst dürfte jetzt verständlich sein anhand des oben gebildeten Lexikons und der Vorschriften zur graphischen Relationsdarstellung.

3. Schritt: Integration der Relationen in ein Netz

Titel: Wir sehen, daß zwischen Titel und 1. Textstelle keine direkte Wiederaufnahme eines Begriffs stattfindet. Diese beiden Relationen sind also voneinander getrennt. Überhaupt bleibt der Titel auch vom ganzen späteren Text getrennt. Wenn der Leser sich also fragt, was der Titel mit dem Text zu tun

hat, dann muß er selbst Hypothesen bilden. Hier ist eine einfache eingetragen: die gestrichelte Linie zwischen "Franzosen" und "französisch" deutet auf die (hier nicht weiter spezifizierte) semantische Beziehung hin. Die Anknüpfung von "Billigkeit" an den Text erfordert komplizierte Hypothesen, die natürlich vom Leser zu leisten sind, wenn er den Titel verstanden hat. Hier soll sie nur skizziert werden. Es würde sich (in Anlehnung an die holistische Interpretation in Abb. 3) eine Kette von Hypothesen ergeben, die wie folgt darzustellen wäre: aus dem ersten Satz ergibt sich die Beziehung: dies ist eine Aktion, die als Hochverrat zu bezeichnen ist. Auf Hochverrat steht in Preußen und in Frankreich die Todesstrafe. Andererseits könnte aber ein Franzose, dem der Hochverrat eines gegnerischen Bürgers Vorteile verschaffen könnte, seine Kenntnis der Tat und der Strafe "abschalten" und Nutzen aus dieser Tat ziehen. Dies wäre zwar nicht nach dem Prinzip: gleiches Recht (und auch gleiche Strafe) für alle, aber in Kriegssituationen verständlich. Daß der Franzose sich trotzdem seinen unbestechlichen Gerechtigkeitsinn bewahrt und den Bürger einen Schurken nennt, dies macht die "Billigkeit" des Franzosen aus.

1. Satz: "ein....Bürger" ist Argument (Subjekt) der beiden miteinander verbundenen Sätze. Die Verknüpfung mit dem 2. Satz des Textes läuft über den Begriff "der General", der auf das 1. Argument des 1. Satzes "...General Hulin" Bezug nimmt. Man sieht hieran: die Bezugnahme muß nicht wörtlich sein, sondern nach dem Kriterium, daß sich beide Ausdrücke auf dasselbe Objekt beziehen, herstellbar sein. Hiernach ergibt sich (nach der Bemerkung zur Thema-Rhema-Gliederung in 2. G4.), daß "Der General" Thema des zweiten Satzes ist, der Rest rhematisch.

Beim weiteren Verlauf des Netzaufbaus sehen wir, daß sich von drei Argumenten aus Linien durch den ganzen Text ziehen lassen (General, Bürger, Stämme). Damit haben wir im Sinne der Bestimmung von 2. G4. drei Isotopielinien festgestellt, auf die noch im nächsten Punkt eingegangen werden soll.

4. Schritt: Zweckorientierte Teilnetzbildung

Unter dem angegebenen Zweck der Exemplifizierung sollen hier als spezielle Isotopielinien die Referenzlinien zu "General Hulin", "ein Bürger" und "eine Anzahl Stämme" herausgegriffen werden.

Faßt man alle Referenzausdrücke graphisch zusammen, die sich auf den General Hulin beziehen, so erhält man eine Linie, die ausgeht vom General

an der Textstelle 1.1 und von dort links nach unten führt bis zum Ausdruck in 4.2. bzw. "er" in 4.3. Fünf Referenzausdrücke werden dadurch miteinander verbunden: zu dem französischen General Hulin, der General, wir, der General, er.-

Eine weitere Linie führt von dem Kasten mit "ein Bürger" (1.4) bis zur letzten Zeile. Ebenfalls 5 Referenzausdrücke werden dadurch verbunden: ein Bürger, mein Freund, der Bürger, ihn, solche Schurken wie er.

Abgesehen davon, daß beide Personen durch die Anzahl der Referenzausdrücke "gleichgewichtig" auftreten, fällt aber die Konstanz der Charakterisierung des Generals (dreimal so benannt) auf gegenüber einer Entwicklung in der Charakterisierung des Bürgers: von der unsicheren Referenz im ersten Ausdruck über eine kurze positive Anrede bis hin zur Geringschätzung am Schluß.

Ein weiteres Referenzobjekt, das mehrmals erwähnt wird, ist das Objekt, um das es in der Anekdote geht: die Stämme. Sie sind (zu meinem eigenen Erstaunen) sogar mit 6 Referenzausdrücken vertreten: eine Anzahl im Pontonhof liegender Stämme, diese Stämme, es, dergleichen Stämme, daran. Dabei wandelt sich zwar der Verwendungszweck der Stämme - vom Brückenmaterial zum Galgen - , nicht aber ihre Wichtigkeit im Krieg (Der König von Preußen BRAUCHT dergleichen Stämme,) -

4. Schritt: Zur Gewichtung

Die Argumentation im 3. Schritt zeigt, welche Gewichtung sich aus dem Zweck ergeben hat: zur Veranschaulichung der Methode habe ich besonders die graphische Darstellung, die Thema-Rhema-Frage und die Entwicklung der Isotopielinien betont.

Anwendung der Methode HOLONTEX (vgl. Abb.3.):

VORGABE: siehe vor ASPECTEX

1. Schritt:

Bei der ersten Lektüre ergibt sich hinsichtlich der Textgestalt der Eindruck, daß der Text nicht besonders gestaltet ist, daß er sehr gedrängt formuliert ist. Die direkten Reden hätten durchaus zeilenweise gegeneinander abgesetzt werden können, um den Dialog deutlicher zu machen. Hinsichtlich des Inhalts ergeben sich die Systeme, die in Verbindung mit dem vorgezogenen 4. Schritt (cf. 2.4) schon auf die hier relevanten eingeschränkt wurden (siehe nächster Schritt).

0		Franzosen-Billigkeit. (werth in Erz gegraben zu werden.)	metaphorisch für: sich dauerhaft erinnern, als Beispiel vor Augen haben
.1	Militärische Hierarchie		
.1.1	Oberster Kriegsherr	zu dem französischen General H u l i n	
1.1	General	← kam,	SYSTEM: KRIEG
1.2		Zeit Ort WELCHEN GEGNERN
1.3	Soziale Hierarchie:	während des Kriegs,	(1806/7) (Frankreich-Preußen)
1.4	Adlige	ein Bürger,	(Sieger: Frankreich)
1.5	Bürger	und gab,	Heerführer Sieger
1.6	(*Ruhe ist nun die erste Bürgerpflicht!*)	Behufs	ben ötig hat das Recht
1.7	(*Bürger in Uniform*)	einer kriegsrechtlichen Beschlagnahmung,	einer kriegswichtigen Material bei Flußüberquerung
1.8	Bürger wie Adliger unter einer allg. Wehrpflicht denn:	zu des Feindes Besten	
1.9	ab 3.8.1808 ist auch Nichtadligen der Zugang zum Offiziersberuf geöffnet.	eine Anzahl,	Ponton= Kahn als Teil einer schwimmenden Brücke (Währig) also: wichtig bei Flußüberquerung
1.1	im Pontonhof liegender,	Stämme sind geeignet als Pontons also: diese Stämme sind kriegswichtig
1.11	Stämme	
1.12	an.	angeben" bezeichnet hier euphemistisch den Akt des Hochverrats (zu des Feindes Besten)
2.1	Eine "Führungskraft" zeigt zunächst einmal Souveränität und Unbestechlichkeit, indem sie den Alltagsgeschäften ohne Unterbrechung nachgeht	Der General,	übliche Erwartung: der Feind zeigt sich erkenntlich für solche Hinweise und anerkennt den Überläufer als Freund.
2.2	der sich eben anzog,	Bestätigung: der General bezeichnet ihn als seinen Freund!
2.3	sagte:	Höfliche Bedenken: zuviel der Sympathiebeziehung(?)
2.4	Der General geht mit der freundlichen Anrede sofort auf den Paktierwunsch positiv ein - so scheint es.	Nein, mein Freund;	
2.5	diese Stämme können wir nicht nehmen. -	
3.1	"Warum nicht?"	Der Bürger zerstreut die Bedenken des Generals
3.2	fragte der Bürger.	Logik: Eigentum des Gegners ist Freibeute also: nach Kriegsrecht bedenkenlos konfiszierbar
4.1	Der flüchtige Blick paßt (noch) nicht zu dem "Wohlwollen" des Generals, denn: flüchtig bedeutet geringschätzend, verächtlich	"Es ist königliches Eigenthum." -	Gegenlogik des Generals: sieben darum nicht konfiszierbar, denn:
4.2	Eben darum, sprach der General,	im Klartext: ein solches Verhalten lehnt der franz. General auch in Kriegszeiten ab (Dieses Verhalten ist nicht zu billigen, auch Gegner sollen fair miteinander umgehen, damit
4.3	indem er ihn flüchtig ansah.	beschämt Hulín den preuß. "Bürger": "ein Schurke!"
5.1	Auch der Feind wird noch in seiner Rolle als König respektiert (und damit als respektabler Kriegsgegner)	Der König von Preußen	außerdem zeigt Hulín, daß er nicht auf diesen Verrat angewiesen ist, um den Krieg weiterzuführen
5.2	braucht dergleichen Stämme,	zusätzlich drückt er dem Bürger gegenüber aus, daß auf eine solche Handlung nach der Logik des eigenen Landes die Todesstrafe steht und Hulín drückt aus, daß dieser Mensch mit Recht diese Strafe verdient
5.3	Umschwung (Pointe der Anekdote) in mehreren Hinsichten:	um solche Schurken daran hängen zu lassen wie er.	
5.4	1. die Haltung des Generals wird plötzlich mißbilligend: er hat keine Verwendung für solche "Freunde", 2. der Verwendungsweg der Stämme wird plötzlich umgedeutet: statt Brücken Galgen 3. für den Leser: der (jetzt nicht mehr so apostrophierbare "Bürger" erhält eine Abfuhr, an die der Leser nach Kenntnis der Kriegslage (und der Franzosen?) nicht mehr glaubt hat.		
	Moralischer Effekt der Anekdote: durch die Pointe wird der gegnerischer Feldherr schließlich zum Vorbild für ein faires Verhalten. Dies muß man den preuß. Bürgern vielleicht erst klarmachen (offensichtlich, da Solches vorgekommen ist.)		

Abb.3: Kleists Anekdote "Franzosen-Billigkeit.": analysiert nach der Methode HOLONTEX.



2. Schritt und 4. Schritt:

In der Abb. 3 sind die präferierten Systeme links und rechts vom Text aufgeführt. Sie sind folgendermaßen bezeichnet:

- Militärische Hierarchie,
- Soziale Hierarchie,
- System Krieg,
- - Subsystem: Verrat
- Anekdotenform: Pointe und moralischer Effekt.

3. Schritt:

Das Lesen unter den verschiedenen Systemen führt zu den Textsystemen. Die Dreiteilung zwischen System, Text und Text-System konnte aus Platzgründen nicht immer so eingehalten werden, wie es in dem formalen Beispiel am Ende von 2.3 dargestellt wurde.

Vom Text gehen Pfeile aus, die auf ein geeignetes System zeigen. Wenn man den Pfeil mit Ausgangspunkt im Text und Endpunkt im System zusammennimmt, erhält man die jeweilige Textsystem-Einheit. Z.B. das System: "militärische Hierarchie" enthält u.a. die System-Positionen: Oberster Kriegsherr, General. Auf "General" deutet ein Pfeil, der von der Text-Position "General Hulin" im Text ausgeht. Der andere Pfeil müßte von "König von Preußen" zu "Oberster Kriegsherr" gehen (hier weggelassen).

Zusammengenommen ergeben sich als Textsystem-Positionen:

Oberster Kriegsherr: <preuß.: König von Preußen>

General: <franz.: Hulin>.

Teilweise sind die Systeme ohne Einbeziehung des Textes dargestellt, teilweise ist die Verbindung zwischen Hintergrund-System und Auffüllung durch Textinformationen oder Hypothesen schon hergestellt:

z.B. System: KRIEG: Zeit: <1806/7>.

Insgesamt ergibt sich eine Gesamtsicht, die zeigt, wie die einzelnen (Text-)Systeme ineinandergreifen.

4. Schritt: Die Gewichtung der Systemteile

Dieser Schritt ist hier zur Exemplifizierung von 2.4 schon in den 2. Schritt integriert worden.

Lieber Leser, wenn Sie zu diesen Beispielanalysen Kritisches oder Verbesserungsvorschläge anbringen möchten, dann liegt dies in meinem Interesse. Es würde zeigen, daß die Methoden und das Beispiel zu einer transparenten Diskussion über unterschiedliche, klarer lokalisierbare Positionen einladen.

Literatur

- Bàtori, I., Krause, J., Lutz, H.D. (Hrsg.) (1982): Linguistische Datenverarbeitung. Versuch einer Standortsbestimmung im Umfeld von Informationslinguistik und Künstlicher Intelligenz. Tübingen.
- Beaugrande, R. de und Dressler, W. (1981): Einführung in die Textlinguistik. Tübingen.
- Berelson, B (1952): Content Analysis in Communication Research. New York.
- Collins, A und Quillian, M (1969): "Retrieval time from semantic memory." in: Journal of Verbal Learning and Verbal Behavior 8. S. 240-247.
- Fühlau, I. (1982): Die Sprachlosigkeit der Inhaltsanalyse: linguistische Bemerkungen zu einer sozialwissenschaftlichen Methode. Tübingen.
- Gottschalk, L.A. und Gleser, G.C. (1969): The Measurement of Psychological States Through the Content Analysis of Verbal Behavior. Los Angeles.
- Koch, H.D. und Mudersbach, K. (1988): ASPECTRA - Programm zur Methode des aspektiven Lesens und berechenbaren Übersetzung. Programmbeschreibung und Erläuterung des linguistischen Modells. Dargestellt in Mudersbach, K. und Koch, H.D. (1988): "Eine Methode des wissenschaftlichen Übersetzens (mit Computerunterstützung)" in: B. Spillner (Hrsg.): Angewandte Linguistik und Computer. Kongreßbeiträge der 18. Jahrestagung der Gesellschaft für Angewandte Linguistik GAL. Tübingen. S. 67-68.
- Koch, U. und Schöfer, G. (Hrsg.) (1986): Sprachinhaltsanalyse in der psychiatrischen und psychosomatischen Forschung. Weinheim.
- Lenders, W. et al. (Hrsg.) (1982): Maschinelle Auswertung sprachhistorischer Quellen. Tübingen.
- Merten, K. (1983): Inhaltsanalyse: Einführung in Theorie, Methode und Praxis. Opladen.
- Minsky, M. (1975): "A framework for representing knowledge." in: Winston (Hrsg.): The psychology of computer vision. New York, S. 211-277.
- Mudersbach, K. (1979): "The method of Formal Interpretation of Philosophical Texts." in: Berghel, H., Köhler, E. (eds.): Wittgenstein, the Vienna

- Circle and Critical Rationalism. Proceedings of the 3rd International Wittgenstein Symposium, Kirchberg am Wechsel/Austria 1978. Wien. S. 174-178.
- Mudersbach, K. (1983a): "Hol-Atomismus als Vereinheitlichung von Holismus und Atomismus." in: Weingartner, P., Czermak H. (eds.): Epistemology and Philosophy of Science. Proceedings of the 7th International Wittgenstein Symposium, Kirchberg am Wechsel/Austria 1982. Wien. S. 347-349.
- Mudersbach, K. (1983b): "Leksemantik - eine hol-atomistische Bedeutungstheorie." in: *Conceptus* XVII, Nr. 40/41. S. 139-151.
- Mudersbach, K. (1987): "Kommunizieren als Übersetzungsproblem. Über Mißverständnisse und deren Verhinderung." in: Liedtke, F und Keller, R. (Hrsg.): *Kommunikation und Kooperation*. Tübingen. S. 37-69.
- Mudersbach, K. (1988): "Die Methode der Gesetzesanalyse als Beitrag der Individual-Linguistik zur Erfassung der Patienten-Wirklichkeit." in: *Zeitschrift für Literaturwissenschaft und Linguistik* 69. S. 84-110.
- Mudersbach, K. und Gerzymisch-Arbogast, H. (1989): "Isotopy and Translation." in: Krawutschke, P.W. (Hrsg.): *Translator and Interpretator Training and Foreign Language Pedagogy*. American Translators Association Scholarly Monograph Series. Vol. III. State University of New York at Binghamton. S. 147-170.
- Neuhaus, V. (1988): "Computer-assisted Interpretation of Detective stories. A Research Project." in: Best, H. et al. (Hrsg.) (1988): *Cologne Computer Conference. Volume of Abstracts*. Köln. C. 4-9.
- Rollinger, C.-R. (Hrsg.) (1984): *Probleme des (Text-)Verstehens. Ansätze der künstlichen Intelligenz*. Tübingen.
- Saussure, F. de (1916): *Cours de linguistique générale*. Lausanne, Paris. Dt. übers. Berlin 1931, 2. Aufl. 1967.
- Schank, R. (1975): *Conceptual Information Processing*. Amsterdam.
- Schank, R. und Abelson, R. (1977): *Scripts, plans, goals and understanding*. New Jersey.
- Schönherr, L. und Mudersbach, K. (1988): RELATAN: Ein Programm zur Auswertung von semantischen Netzen. Programm-Beschreibung und Erläuterung des linguistischen Modells. Dargestellt in Mudersbach, K., Jacob, W., Schönherr, L. (1988): "Das Netz subjektiver Sprach- und Sachgesetze eines Individuums, ermittelt aus seinen Äußerungen." in: B. Spillner (Hrsg.): *Angewandte Linguistik und Computer*. Kongreßbeiträge der 18. Jahrestagung der Gesellschaft für Angewandte Linguistik GAL. Tübingen. S. 55-57.

- Sembdner, H. (Hrsg.) (1982): Berliner Abendblätter. Herausgegeben von H.v. Kleist. Darmstadt.
- Sembdner, H. (Hrsg.) (1983): H.v. Kleist, Sämtliche Werke und Briefe. 7. Auflage. München.
- Stutterheim, C. von (1988): "Sprachanalyse als medizinisches Diagnoseinstrument. Eine exemplarische Diskussion des Gottschalk/Gleser Verfahrens." in: Zeitschrift für Literaturwissenschaft und Linguistik Nr. 69. S. 54-83.

Historia ex machina

Der EDV-Einsatz in der Geschichtswissenschaft und eine Erinnerung an seine Voraussetzungen

von **Hans-Christoph Hobohm**

0. Unübersehbar hat nun der Computer auch Einzug in die Geisteswissenschaften gehalten. Aufgrund von mehr oder weniger staatlichen Modernisierungsprogrammen ist er mittlerweile - vor allem in Form von "Mikrocomputern" und ihren Extensionen - in unmittelbarer Reichweite einer jeden Disziplin. Dennoch hat die Faszination der Maschine kaum nachgelassen. Im Gegenteil, dadurch, daß die Geräte jetzt materiell, psychologisch und ergonomisch immer zugänglicher werden, fallen einige der kritischen Barrieren aus der Anfangszeit der Datenverarbeitung anscheinend weg. Die ansprechende Benutzeroberfläche des Desktop-Systems täuscht darüber hinweg, daß die alten Probleme der Formalisierbarkeit geisteswissenschaftlicher Daten noch lange nicht gelöst sind. Noch hat sich wenig daran geändert, daß das Funktionsprinzip des Computers lediglich binär ist. Im Folgenden soll ein gewiß unvollständiger Überblick¹ über einige der verbreitetsten Anwendungen des neuen EDV-Instrumentariums im Bereich der Geschichte gegeben und gleichzeitig versucht werden, deren jeweilige Prämissen zu diskutieren.

Entsprechend der beiden Hauptcharakteristika des Computers: als Archiv und als Algorithmus, lassen sich zwei Komplexe maschineller Informationsverarbeitung nicht nur im Bereich der Geisteswissenschaften ausmachen. Der eine verweist auf die Fähigkeit des Computers, für den Menschen nicht mehr überschaubar große Mengen von Information zu speichern und wiederzugewinnen, und der andere ist die Möglichkeit, mit ihm extrem komplexe Bedingungsgefüge zu berechnen und neu, d.h. heuristisch reduziert, darzustellen. Nach einer längeren Phase der Betonung des Archivaspekts und nach dessen Konsolidierung, z.B. in diversen Datenbanksystemen, gewinnt zur Zeit der zweite Aspekt immer mehr an Bedeutung. Die Debatte der mit dem ersten grundlegend verbundenen Probleme der Quantifizierung in den Geisteswissenschaften hat an Schärfe verloren, und man hat den Eindruck,

¹ Es sei hier speziell auf das *Humanities Computing Yearbook*, Oxford: Clarendon Press, 1988ff verwiesen. Vgl. zum Methodischen allg. auch Mario B o r i l l o: *Informatique pour les sciences de l'homme. Limites de la formalisation du raisonnement*. Bruxelles/Liège: Mardaga, 1984.

daß der Computer als Arbeitsinstrument, in Form des Zettelkastens oder als "Denkkasten" allgemein akzeptiert ist.

1. Das maschinenlesbare historische Archiv befindet sich nunmehr buchstäblich "on your desk", von dem aus entweder via Telematik die großen kommerziellen Datenbanken abgefragt werden können, oder auf dem sich ein Abspielgerät für eins jener neuen technischen Medien mit bisher unvorstellbaren Speicherkapazitäten wie CD-ROM oder Videodisk befindet. Wenn man z.B. die Gesamtbestände der großen Nationalbibliotheken oder die Gesamtheit der bekannten Texte der Antike nunmehr auf einer kleinen westentaschengroßen Diskette zur Verfügung hat, so ermißt man leicht den Unterschied in Dimension und Potential der neuen Zettelkästen. Die Faszination der quantitativen Differenz ließ aber oft die qualitativen Möglichkeiten des neuen Mediums aus dem Blick geraten und führt manchmal zu eher abschreckenden Ergebnissen. Wenn eine Serie von Bildplatten² mit Material für den Geschichtsunterricht sich letztlich nur durch die Druckknopfsteuerung der Fernbedienung und den Monitor vom Medium Buch unterscheidet, so wird der Geschichtslehrer (sofern er überhaupt "Medien" einsetzt) sich wahrscheinlich überlegen, doch bei Overheadprojektor oder Epidiaskop zu bleiben. Daß es auch ganz anders geht, zeigt das vielzitierte Beispiel der interaktiven Bildplatten des britischen Domesday Projektes, auf denen dem Benutzer (zugegebenermaßen noch nicht in wirklich historischer Perspektive) die Möglichkeit gegeben wird, sich quasi nach dem Modus der verbreiteten adventure games in einer simulierten Realität zu bewegen oder eine unvorstellbare Menge von Daten über das England der frühen achtziger Jahre auf dem Bildschirm zu bearbeiten und ganz den eigenen Vorstellungen entsprechend darzustellen.

Neben den jetzt neu zu stellenden eher didaktischen Fragen der Präsentation und Aufbereitung historischer Informationen bleiben die alten Probleme der Quantifizierungsdebatte und der histoire sérielle natürlich bestehen³. Auch die im Rahmen des Historical Workstation Projektes von Manfred Thaller⁴ mögliche Verwendung neuerer mathematischer und EDV-techni-

² Ich nenne absichtlich keine Beispiele.

³ Vgl. z.B. Jean-Louis R o b e r t: *L'histoire quantitative, réflexions épistémologiques*. *Historical Social Research / Historische Sozialforschung*, 25 (1983), S. 24-33. Jerzy T o p o l s k i: *Die Wissenschaftlichkeit der Geschichtsschreibung und ihre Grenzen*. in: Rossi, P. (Hg) *Theorie der modernen Geschichtsschreibung*. Frankfurt/M.: Suhrkamp, 1987, S. 191-218.

⁴ Manfred T h a l l e r: *Data Bases v. Critical Editions*. *Historical Social Research /*

scher Konzepte (Stichworte: fuzzy sets und hypertext) ändern nur wenig an den grundlegenden Problemen der Quellenkritik und des meist aus forschungspragmatischen Überlegungen notwendigen Reduktionismus der Datenaufnahme. Sehr viel wichtiger ist dagegen die von Jean-Philippe Genet aufgeworfene Frage nach der Art der durch die neuen Dimensionen der Speicherkapazitäten möglich gewordenen Datenbasen⁵. Für den informatisierten Arbeitsplatz des Geisteswissenschaftlers sind sicher nicht nur ganz spezifische, aus einzelnen Forschungsprojekten gewonnene Datenbanken notwendig, sondern sehr vielfältige, allgemeine "Wissensbasen" wie Enzyklopädien oder biographische Lexika⁶. Eine nicht oft genug zu betonende Chance und Gefahr für die Geschichte bietet schließlich eine ganz besondere Art von Quellen. Nicht nur die Historiker bedienen sich mehr und mehr des Computers: auch die Geschichte produzierenden Instanzen selbst, die politischen Institutionen und auch die wissensproduzierenden Verlage haben mittlerweile maschinenlesbare Archive. Dies bedeutet eine Gefahr für die Geschichte deshalb, weil ebenso wie die Möglichkeiten der Wiedergewinnung von Informationen auch deren Vernichtung - und sei es versehentlich - sehr viel bequemer geworden sind: Säurefreie Papierarchive mit Informationen "schwarz auf weiß" sind immer noch sicherer als die relativ wartungsintensiven Magnetbandarchive. So ist oft darüber spekuliert worden, daß zukünftige Historiker des 20. Jahrhunderts auf große Dokumentationslücken treffen werden, weil viele Archive entweder dem Säurefraß des Papiers oder dem mangelnden archivarischen Bewußtsein früherer Magnetbandoperateure zum Opfer fielen: eine Chance und Herausforderung der neuen Technologien der Datenspeicherung.

Doch nicht nur regelrechte Archive sind mittlerweile binarisiert, kodiert und maschinell transportabel. Fast jegliche Textproduktion läuft zu dem einen oder anderen Zeitpunkt über das "Medium" Computer. Und alle Texte, die noch nicht maschinell vorliegen, lassen sich in Windeseile per OCR scanner

Historische Sozialforschung, 13 (1988) 129-139; sowie ders.: Kleio 3.1.1 : ein Datenbanksystem. St. Katharinen: Scripta Mercaturae, 1989.

⁵ Jean Philippe Genet: Pour l'informatisation des dictionnaires biographiques, une expérience. Histoire & Mesure, 1 (1986) 85-98; vgl. auch allgemein: ders. (Hg): Standardisation et échange de bases de données historiques. Paris: Editions du CNRS, 1988.

⁶ Vgl. Hans-Christoph Hobohm: Using Databases for Everyday Work in Literary History. Exchange and Standardization Problems. in: Standardisation et échange de bases de données historiques, hg. v. J.P. Genet. Paris: Editions du CNRS, 1988, S. 323-328.

einlesen. Hier liegt vielleicht eine der wirklich relevanten Innovationen für den Historiker, und zwar vor allem für den "qualitativ" arbeitenden. Die Geschichte als Quellen- und Textwissenschaft wird in naher Zukunft ein Instrumentarium an die Hand bekommen, von dem vorhergehende Historikergenerationen nur träumten: die elektronische Bibliothek⁷. Textkategorisierungsprogramme werden den Leuchtstift in der Hand des Lesers ablösen, Textsysteme die handschriftliche Exzerptensammlungen und Textdatenbanken mit direktem Zugriff auf einzelne Wörter oder bestimmte Konzepte eines Autors die "Gesammelten Werke" ersetzen⁸. Wesentliche Voraussetzung allerdings für eine befriedigende Nutzbarkeit der immensen Textmengen in Computerspeichern ist eine nicht unerhebliche Weiterentwicklung der Grundlagenforschung im Bereich der computergestützten Inhaltsanalyse. Obschon diese auf eine beachtliche Tradition zurückblickt, muß sie vielfach immer noch die starken Beschränkungen ihrer Leistungsfähigkeit bekennen⁹. Neben der Speicherung von sprachlich-textuellem Material sollte der graphisch-visuelle Bereich in seiner Bedeutung für die historischen Wissenschaften nicht unerwähnt bleiben, obwohl auch hier noch sehr viel Vorarbeit geleistet werden muß. Hier bietet sich zum Beispiel konkret die Möglichkeit der graphischen Visualisierung historischer Prozesse, etwa auf chronologisch-thematischen Karten an¹⁰. Es scheint letztlich, daß der Menge

⁷ Welche Konsequenzen dies allerdings auch für die Textproduktion selber hat, zeigt Wolfgang H a g e n: Die verlorene Schrift. Skizzen zu einer Theorie der Computer. in: *Arsenale der Seele: Literatur- und Medienanalyse seit 1870*, hg.v. F.A. Kittler u. G.C. Tholen, München: Fink, 1989, S. 211-229.

⁸ Vgl. z.B. das recht einfache, aber nützliche Programm: TDCII (Textual Data Categorisation) von Wayne S t a a t s und Linda B r i g g s zur konzeptuellen Markierung von Texten, vertrieben über das amerikanische National Collegiate Software Clearinghouse oder das von Joachim N e u b e r t vorgeschlagene geisteswissenschaftliche Karteisystem: *Der elektronische Arbeitsplatz in den Geisteswissenschaften. Nachrichten für Dokumentation*, 39 (1988) 291-296, 357-364. Rainer K u h l e n u.a.: *Hypertext - Grundlagen und Funktionen der Entlinearisierung von Text. Nachrichten für Dokumentation*, 40 (1989) 295-307.

⁹ Vgl. Peter Ph. M o h l e r: *Computergestützte Inhaltsanalyse: Überblick über die linguistischen Leistungen*. in: *Computational Linguistics. An International Handbook on Computer Oriented Language Research and Applications*, hg. v. I.S. Bàtori u.a., Berlin u.a.: W. de Gruyter, 1989, S. 580-585. Das ganze Handbuch vermittelt im übrigen den Eindruck eines noch enormen Forschungsbedarfs. Vgl. zu dem angeschnittenen Themenkomplex den hier abgedruckten Artikel von Klaus M u d e r s b a c h.

¹⁰ James B.M. S c h i c k: *Historians and Computing, circa 1987. Social Science Microcomputer Review*, 5 (1987) 514-528.

wie auch der Art der gespeicherten Information kaum Grenzen gesetzt sind. Nur sollte dabei nicht übersehen werden, daß am Ende der Gutenberg-Galaxis nicht nur ein einfacher Transformationsprozess vollzogen, sondern auch gleichzeitig das Lesen und Schreiben in dem neuen Medium neu erfunden werden muß.

2. Der Algorithmus-Aspekt des Computers drückt sich schon innerhalb der Archivdimensionen der quantifizierenden Historiographie in der immer weitergehenden Anwendung spezifischer Statistikprozeduren aus. Entsprechend ihrer Meßniveau-Probleme und anderer Charakteristika ihrer Daten (z.B. der Verteilung in Zeitreihen) haben sich in den nicht naturwissenschaftlichen Statistikanwendungen mittlerweile eine Vielzahl mehr oder weniger komplexer multivariater Analysemethoden herausgebildet, die mit der jeweiligen Software auf Tastendruck aus den archivierten Daten spezielle Maßzahlen oder Graphiken extrahieren¹¹. Als besondere Methoden der Datenanalyse lassen sich die unter den Stichworten "Künstliche Intelligenz" und "Expertensystem" firmierenden Verfahren bezeichnen¹². Im Gegensatz zu dem eher formelhaften Algorithmus der Statistik besteht dieser in dem Versuch einer Automatisierung der Argumentation des interpretierenden oder analysierenden Forschers. Um zum Beispiel einzelne Personen mit Hilfe eines heterogenen Korpus von Sekundärinformationen zu identifizieren, lassen sich Regeln beschreiben, die befolgt werden müssen, um Informationen einem Namen zuordnen zu können. Die infragekommende Person muß einem bestimmten Profil entsprechen (Lebensdaten und Aufenthaltsorte, sozioökonomischer Status), um ein weiteres mit einer gleichlautenden Namensform in Verbindung zu bringendem Informationselement ihr zuordnen zu können - und umgekehrt. In der Begrifflichkeit von Expertensystemen: ein Satz von Regeln (z.B.: "X muß vor dem Todesdatum bzw. in einem bestimmten Lebensalter von Y, der Person, stattgefunden haben") wird über Inferenzoperatoren in Verbindung gebracht mit den Fakten einer Datenbasis (z.B. den aufbereiteten Auszügen einer Quelle). Was bei diesem Automat die Bezeichnung "Intelligenz" verdient, ist die Tatsache, daß die Regeln derart programmiert werden können, daß sie sich selber beeinflussen, also lernen können. Es handelt sich also hierbei um Simulationen der "Expertise" des

¹¹ Vgl. hierzu auch den Artikel von Johannes G o r d e s c h in diesem Band.

¹² Vgl. z.B. Richard E n n a l s: *Artificial Intelligence: Applications to Logical Reasoning and Historical Research*. Chichester: Ellis Horwood, 1985. kritisch: Hubert L. D r e y f u s und Stuart E. D r e y f u s: *Mind over Machine: the Power of Human Intuition and Expertise in the Era of the Computer*. Oxford: Blackwell, 1989.

Historikers, die den Vorteil haben, stringenter und reliabler zu sein als menschliche Argumentation - sofern die zugrundeliegenden Regeln und die Tatsachenbeschreibungen korrekt sind ¹³.

Die bisher entwickelten, in diesem Sinn analytischen "Simulationen" historischen Wissens können, da sie immer nur auf konkrete Erkenntnisprobleme angewandt werden ¹⁴, kaum Repräsentativität beanspruchen, bzw. eignen sich nicht dazu, die Grundprobleme des algorithmischen Aspekts der EDV-Anwendung in der Geschichte zu illustrieren. Es sei deshalb lediglich ein weniger komplexes Beispiel näher erläutert. Die Konstruktion eines regelrechten Expertensystems ist ein sehr aufwendiges Unterfangen vor allem wegen des Aufbaus der zugrundegelegten Wissensbasis. Die für notwendig erachteten Wissens Elemente müssen adäquat eingegrenzt und regelgerecht beschrieben werden. Der Archivaspekt des Computers wird in diesem Zusammenhang weit mehr als bei einer Kodierung im Hinblick auf statistische Auswertungen zu einem Problem der Semiotik ¹⁵. Dies läßt sich an dem kleinen, beeindruckenden Programm ETHNO ¹⁶ zur Analyse von Handlungsgrammatiken und Argumentationsstrukturen gut nachvollziehen. Der Inferenzmotor des Systems ist in diesem Fall festgelegt auf die metonymische Beziehung der logischen Abfolge von Handlungsabläufen der Art: Textstruktur-, bzw. Handlungselement A: "Rotkäppchen geht in den Wald" ist notwendig für Element B, "trifft auf den Wolf", deren Vollendung das Element C, "wird gefressen", ist. Je nachdem welche Handlungselemente nun für welche Handlungsstränge als relevant ausgewählt und in die Wissensbasis eingetragen werden, zeichnet das Programm unterschiedliche Diagramme zur Heuristik der analysierten Textstrukturen. Schnell erkennt man, wie wichtig

¹³ Vgl. die Rekonstruktion geisteswissenschaftlicher Argumentationen bei Jean Claude G a r d i n u.a.: *Systèmes experts et sciences humaines*. Paris: Eyrolles, 1987.

¹⁴ Z.B. Marie-Danielle D e m é l a s und Monique R e n a u d: *L'historien peut-il être "artificiel"? Un prototype de système expert, appliqué à l'histoire politique*. Vortrag auf der Tagung: "L'outil ordinateur et le métier historien." Aix-en-Provence, 1.-2.10.1987. Helmut K a p h e n g s t, Volker J o k s c h, Gerd G r a ß h o f f: *Methodik wissenschaftlicher Heuristiken am Beispiel der Interpretation eines astrologischen Gedichtes*. Hamburger Philosophische Schriften, 3 (1989) Heft nr. 2

¹⁵ Speziell der Sigmatik: der semantischen Beziehung zwischen dem Zeichen und dem von ihm beschriebenen realen Objekt.

¹⁶ David R. H e i s e: *ETHNO. A System to Construct Action Grammars to Interpret Texts*. Version 2.1, 1988, ebenfalls vertrieben vom: National Collegiate Software Clearinghouse. Vgl. auch ders.: *Computer Analysis of Cultural Structures*. *Social Science Computer Review*, 6 (1988) 184-196.

die konzeptionelle Analyse der behandelten Strukturen und wie gering der intelligente "Mehrwert" des Automaten ist, wenn die Segmentierung und Beschreibung der Elemente unzureichend war (Ist "Gefressenwerden" wirklich die Beendigung des Handlungsstranges "In-den-Waldgehen"?). Die Maschine kann also nur kognitives oder heuristisches Instrument sein zur Unterstützung von Analyse und Beschreibung komplexer, hier: narrativer Strukturen.

Noch einleuchtender sind die Grenzen der Modellierung komplexer Strukturen im Zusammenhang mit den eher synthetischen denn analytischen Anwendungen der Simulation historischer Prozesse zum Zwecke mehr oder weniger didaktischer Darstellung. Grob verkürzt kann man sagen, daß es zwei Arten von Simulationen gibt. Zum einen solche, die zum Testen und Darstellen von Hypothesen aus Theorien über soziohistorische oder ökonomische Dynamiken verwandt werden, und zum anderen solche, die die historischen Situationen und Bedingungen einzelner Individuen (Akteure) durchspielen¹⁷. Simulationen der ersten Art sind z.B. bekannt geworden im Zusammenhang der Zukunftsprognosen von Jay F o r r e s t e r und des Club of Rome¹⁸. Durch Veränderung zugrundegelegter Parameter - entsprechend der jeweils angelegten Theorie - kann man den je unterschiedlichen Ablauf eines Algorithmus beobachten, der bestimmte Kurven oder andere Graphiken auf dem Bildschirm zeigt zur Veranschaulichung historischer Entwicklungen. Ein berühmtes Modell dieser Art ist das von William S. B a i n b r i d g e: "REVOLT: the contending parties of Weimar Germany"¹⁹, mit dem beginnenden Soziologiestudenten gezeigt werden soll, daß bestimmte soziohistorische Bedingungen, wie Dichte und Art der sozialen Netzwerke, Stabilität der öffentlichen Meinung, Klassenstrukturen, ökonomische Bedingungen etc., die Machtübernahme der Nationalsozialisten mehr oder weniger favori-

¹⁷ Immer noch nützliche Einführung: Steffen H a r b o r d t: Computersimulation in den Sozialwissenschaften. 2 Bde., Reinbeck: Rowohlt, 1974. vgl. auch: Stephen W. B a s k e r v i l l e: Pasts and Metapasts - Beware of Simulations, Computing and History Today - Newsletter, 2 (1987) 29-33. Pierre C o r b e i l: Rethinking History with Simulations, History Microcomputer Review, 4:1 (1988) 15-20. James E. S a r g e n t und Stephen P. H u e s t o n: Simulating History: Programming the American Revolution, History Microcomputer Review, 3:2 (1987) 17-22.

¹⁸ Vgl. Hans R a u c h: Modelle der Wirklichkeit. Simulationen dynamischer Systeme mit dem Mikrocomputer. Hannover: Heise, 1985.

¹⁹ William Sims B a i n b r i d g e: Sociology Laboratory. Computer Simulations for Learning Sociology. Belmont, CA: Wadsworth, 1987.

sieren konnten. Bei Wahl der historisch "richtigen" Parameter zeigt das Modell, daß es unvermeidlich war, daß die Nationalsozialistische Partei im Laufe der Verschlechterung der ökonomischen Verhältnisse großen Zulauf bekam. Der Algorithmus des Modells übersetzt die gewählten Ausgangsparameter in eine - im "Jahresrhythmus" - wiederholte Berechnung der Stimmenanteile der einzelnen Parteien in den unterschiedlichen, ebenfalls vorher definierten sozialen Gruppen. Zur Illustration der Entwicklung werden am Bildschirm die imaginären Bewohner von "Deutschstadt, 1918 ->" in der jeweiligen Farbe ihrer politischen Meinung dargestellt. Die eingangs gewählte Art der sozialen Interaktionen und die jährlichen Änderungen der wirtschaftlichen Lage bestimmen das Bild der Stimmenverteilung des darauffolgenden Jahres, und wenn der Bildschirm voller brauner Deutschstädter ist, hat man offensichtlich die "richtigen" Parameter zur Erklärung der Machtübernahme der Nazis gewählt. Bei anderen Simulationen dieser Art ist die Wahl der zugrundegelegten Parameter weniger dem Benutzer überlassen. Vielfach wird der Fortgang der Simulationen von automatisch generierten Zufallszahlen bestimmt, die dann die vom Autor des Programms vertretene Einstellung zu den relevanten "historischen Kräften" ²⁰ veranschaulichen. Wie man sieht, unterscheidet sich diese Art der Beschreibung und Erklärung historischer Phänomene nur durch die Wahl der Ausdrucksmittel von der narrativen Geschichtsschreibung. Doch anders als die natürliche Sprache verlangt ein Algorithmus absolute Explizierung der zugrundegelegten Prämissen.

Neben der Simulation der Großen Historischen Erzählung (Lyotard) bietet der Computer aber auch die Möglichkeit, konkrete Lebenssituationen einzelner Individuen zu simulieren, also Mikrogeschichte zu betreiben. Ein Beispiel für eine Anwendung der Simulationstechniken in diesem Bereich, das Furore gemacht hat, ist die von Carolyn L o u g e e geschaffene Geschichte einer Familie im Frankreich des siebzehnten Jahrhunderts: "The Would-Be-Gentleman" ²¹. Hier wählt man keine Parameter in bezug auf historische oder soziologische Theorien, sondern muß sich konkret in die Lage eines historischen Individuums versetzen und seine Rolle im Jahreswechsel des Alltagslebens übernehmen: Denis Martin, 1638 30 Jahre alt, Bürger von Rouen und Auditeur de la Chambre des Comptes dieser Stadt, Erbe von 42 Hektar Land und 5054 Livres Bargeld. Vom Beginn der Fronde bis zum Tode

²⁰ "to illustrate historical forces" ist z.B. die erklärte Konzeption des Modells "French Wars of Religion" von Stephen H u e s t o n, vgl. Sargent / Hueston, art. cit.

²¹ Vgl.: Carolyn Chappell L o u g e e "The Would-be-Gentleman": A Historical Simulation of the France of Louis XIV, History Microcomputer Review, 4:1 (1988) 7-14. Das Programm wird vertrieben von Kinko's Academic Courseware Exchange.

des Sonnenkönigs muß Denis Martin (bzw. später sein Sohn, wenn er es schafft, einen Erben zu zeugen) jedes Frühjahr und jeden Herbst wichtige Entscheidungen bezüglich der Organisation seines sozialen und ökonomischen Familienlebens treffen. So muß also derjenige, der am Computer seine Rolle übernimmt, nicht nur schwierige Entscheidungen über Kapitalanlagen unter den Bedingungen des Ancien Régime treffen (z.B. Zeitpunkt des Verkaufs der Ernte, Spekulation mit Textilien, sozial angemessener Kauf oder Verkauf von Titeln, Ämtern und Pachten etc.), sondern auch das familiäre und persönliche Schicksal von Denis Martin in die Hand nehmen. So muß man im Verlauf seines Lebens die Protektion eines hochgestellten Gönners gewinnen, muß angemessen heiraten und entsprechend den gesellschaftlichen Normen seine "Familienplanung" betreiben und erlebt so "am eigenen Leibe", was es heißt, unter den Bedingungen des Absolutismus zu leben. Königliche Willkür, Adelstitel oder Beziehungen zum Hof wirken sich nicht nur auf das in Bargeld und Besitz gemessene Vermögen, sondern auf das anhand einer "Prestigeskala" dargestellte soziale Ansehen aus. Macht man zum Beispiel einer Tochter aus zu hoch gestelltem Hause den Hof oder wirbt man um die Gunst eines weit über seinem eigenen Stand stehenden Großen, so riskiert man bei einer Ablehnung neben dem Verlust seiner Investitionen in Form von Präsenten auch noch starke Prestigeeinbußen. Es gilt, einen seinem Stand angemessenen Lebenswandel zu führen und darauf zu achten, die bienséances zu wahren. Diese Art der Simulation, besser: des Rollenspiels, ist natürlich auf einer sehr viel konkreteren historischen Ebene angesiedelt als die auf eher globale Konjekturen angewiesenen dynamischen Modelle. Aber weitaus mehr als in einem Film hat man hier durch seine aktive Beteiligung die Möglichkeit, sich mit der historischen Person Denis Martin in seiner Lebenssituation zu identifizieren und historische Alterität zu erleben. Hier kann das andere Medium mit seinen potentiell ganz anderen Kommunikationsmitteln wirklich grundsätzlich neue Formen der Vermittlung historischen Wissens entwickeln.

3. Die hier kurz dargestellten, neuen technischen Möglichkeiten für die historischen Wissenschaften weisen eine ganze Anzahl von Problemen auf. Nicht zuletzt sind hier natürlich die Probleme der Standardisierung und der Kompatibilität zu nennen. Schon zwischen der MacIntosh- und der IBM-Welt scheint nicht nur technisch oft kaum ein Austausch möglich, und die weitreichenden didaktischen Erfahrungen, die jenseits des Ärmelkanals gesammelt wurden, sind ebenfalls vielfach aufgrund der insulären Inkompatibilitä-

ten der in den englischen Schulen benutzten Maschinen nur schwer vermittelbar. Die Wiederverwendung einmal erfaßter historischer Daten scheitert oft genug an mangelnder Dokumentation, divergierenden Datenaufnahmestandards oder zu spezifischen Zielsetzungen der jeweiligen Primärstudie ²².

Doch dies sind banale Probleme im Vergleich zu den nicht oft genug zu betonenden grundsätzlichen Schwierigkeiten epistemologischer Art. Das Hauptproblem ist so bekannt, wie es immer wieder verdrängt wird: der Computer verarbeitet nur präzise Informationen. Eine Person lebt oder lebt nicht zu einem bestimmten Zeitpunkt. Das "Vielleicht" historischer Überlegungen läßt sich im Computer - normalerweise - nicht abbilden, denn er versteht ja bekanntlich nur das on und off elektrischer Zustände. Jeder Historiker kennt aber die Probleme der Ungenauigkeiten der aus den Quellen zu entnehmenden Informationen - und sei es unter lediglich paläographischen Gesichtspunkten. Wie schon Michel de C e r t e a u konstatierte, ist das Fehlen einer "unscharfen Logik", das Haupthindernis für eine adäquate Anwendung des Computers in den Geisteswissenschaften ²³. Ohne hier in Einzelheiten gehen zu wollen, kann man feststellen, daß sich in neueren Strömungen der Mathematik gewisse Lösungen dieses Problems abzeichnen, die zum Zeitpunkt von de Certeaus Äußerung noch kaum wahrnehmbar waren: Lofti Z a d e h s Theorie der unscharfen Mengen, René T h o m s Katastrophentheorie zur Beschreibung komplexer Morphologien oder auch Benoit M a n d e l b r o t s Theorie der fraktalen Objekte könnten in naher Zukunft dem Historiker Instrumente an die Hand geben, mit denen sich die notwendige reduktionistische Formalisierung historischer Daten für den EDV-Einsatz mildern ließe ²⁴.

Ein anderes, von de Certeau erwähntes Problem ist der Drang des Historikers zur Exhaustivität ²⁵. Die unermesslichen Speicherkapazitäten des

²² Vgl. J.P. G e n e t, op.cit.. Es sei hier darauf hingewiesen, daß zur Beschreibung von Texten mittlerweile ein internationaler Standard existiert (Standard Generalized Markup Language: SGML), an den sich aber leider kaum jemand hält.

²³ Michel de C e r t e a u: Science et fiction: l'histoire, de l'informatique à l'anthropologie. Nouvelles littéraires, 9 janv. 1977, 17-21. (wieder in: Philosophy of History and Contemporary Historiography, hg.v. D. Carr, W. Dray, T. Geraets. University of Ottawa Press, 1982, S. 19-39).

²⁴ Vgl. auch den Artikel von J. G o r d e s c h im vorliegenden Band.

²⁵ Michel de C e r t e a u: L'écriture de l'histoire. Paris: Gallimard, 1975, S.85ff.

Computers verleiten ihn, nach Möglichkeit die Totalität der über eine Epoche, ein Ereignis oder eine Person verfügbaren Informationen zu erfassen. Er übersieht dabei in seiner "passion de l'exhaustif", daß die uns überlieferte Information keineswegs ein abstraktes und neutrales Sediment historischer Realität ist, sondern nur durch - teilweise divergierende - Filter von Konzeptionen des Überlieferungswerten erhalten blieb. Es ist also nicht die Totalität aller auffindbaren Daten über die Vergangenheit, die die Grundlage bilden kann für die historische Analyse, sondern "un ensemble réduit et fini de données extraites de l'ensemble infini qu'est le 'réel historique'"²⁶. Die Konstruktion einer Metaquelle ist die wesentliche Aktivität des Historikers und nicht die Datenerfassung ad libitum²⁷. Es ist bezeichnend, daß es gerade das Potential des Computers ist, das uns hier wieder in die Schranken verweist. Dennoch bleibt die Frage, ob die Konstruktion eines Expertensystems die (gerade wieder entdeckte) historische Erzählung ersetzen kann. Jean Claude G a r d i n, der zwar gerne die angenehme Seite der nicht formalisierten historiographischen Prosa hervorhebt²⁸, deutet selbst, wenn er die Projekte automatischer Geschichtengeneratoren mit denen historischer Expertensysteme vergleicht²⁹, die Möglichkeit an, im Computeralgorithmus das narrative Wissen der Geschichte aufgehoben zu finden.

Diese beiden von de Certeau thematisierten Probleme, nämlich das Fehlen einer unscharfen Logik und die Notwendigkeit der Konstruktion von Metaquellen, verweisen letztlich auf die beiden eingangs erwähnten Aspekte des Computers als Archiv und Algorithmus, und damit epistemologisch auf Semiotik und Kybernetik. Der erste Problembereich entspricht der Frage nach der Repräsentation von Wissen, die deshalb so kompliziert ist, weil sie für den Menschen so ungewöhnlich ist. Man ist weder gewöhnt, sich sein Wissen oder seine Äußerungen in einem stringenten Symbolsystem vorzustellen, noch ist man in der Lage, es sich in "kleinste bedeutungstragende Einheiten" zu zerlegen, scheidet man ja sogar manchmal daran, es überhaupt in natürlicher Sprache zu formulieren. Auch der zweite Aspekt, der Versuch, die Beziehungen dieser Elemente untereinander zu formalisieren und deren

²⁶ "eine begrenzte und endliche Menge von Daten aus der unendlichen Menge 'historischer Realität'", Jean Philippe G e n e t: Histoire, Informatique, Mesure. Histoire & Mesure, 1 (1986) 7-18; S.IO.

²⁷ Zu dem "Paradigma" des Konstruktivismus vgl.: Der Diskurs des Radikalen Konstruktivismus, hg. v. Siegfried J. S c h m i d t. Frankfurt/M.: Suhrkamp, 1988.

²⁸ a.a.O. S.256f.

²⁹ a.a.O. S. 247.

Inbezugsetzung zu automatisieren, widerstrebt dem natürlichsprachlichen Empfinden. Beide haben aber den entscheidenden Vorteil, den Geisteswissenschaftler zu einer anderen Art von Wissenschaftlichkeit zu zwingen, und die Hermeneutik auf die Füße von Explizitheit und Intersubjektivität zu stellen. Zu allererst erweist sich also der Computer als Katalysator im Hinblick auf die Entwicklung neuer wissenschaftlicher Methodologien.

Die Rigorosität der Maschine legt es den Geisteswissenschaften nahe, in mancher Hinsicht naturwissenschaftliche Vorgehensweisen an den Tag zu legen. Gerade die erst dabei offensichtlich werdenden Differenzen können sicher auch zu einer wesentlichen Standpunktklärung innerhalb der Wissenschaftspraxis dienen. So findet man z.B. in einem naturwissenschaftlichen Forschungsbericht³⁰ normalerweise nach einem Exposé von Forschungsziel und Beobachtungsgebiet eine Beschreibung der zu analysierenden Tatsachen mithilfe einer der Teildisziplin spezifischen Notation, d.h: einem symbolischen oder semiotischen Wissensrepräsentationssystem. Dem folgt die Erläuterung der Thesen als Produkt einer Kette von Operationen. Außerhalb des wissenschaftlichen Textes greifen dann schließlich Mechanismen der Validierung der Thesen. Dieses Modell des wissenschaftlichen Prozedere (Beobachtung - Beschreibung - Argumentation - Validierung) ist in den Geisteswissenschaften kaum anzutreffen. Allein schon die allgemeinen Forschungsziele und Forschungskontexte sind hier grundsätzlich andere: es geht nicht um Experiment und Beweis, sondern um Philosophie und Charisma. Das Ziel des geisteswissenschaftlichen Forschers ist nicht zu überzeugen, sondern zu bekehren. Sein Wissensrepräsentationssystem wählt er nahe, wenn nicht sogar identisch mit der natürlichen Sprache mit dem Argument, von möglichst vielen verstanden zu werden. Seine Argumentation selbst ist ebenfalls einer natürlichen Logik entlehnt, so daß sich oft weder auf der Beschreibungs- noch auf der Argumentationsseite nachvollziehbare Verifikationsansätze bieten. Solche Freiheiten bieten natürlich Anlaß zu einem Überfluß geisteswissenschaftlicher Literatur mit fatalen Folgen für das in den Naturwissenschaften wesentliche Moment der Forschung: der Validierung der Thesen. Die Vielfältigkeit geisteswissenschaftlicher Repräsentationen und Konstruktionen hat allerdings nicht den gleichen Status wie der fruchtbare Theorienwettstreit in den Naturwissenschaften. Die Mannigfaltigkeit der Erklärungsmöglichkeiten menschlicher Phänomene erscheint bestenfalls unvermeidlich - normalerweise fällt sie ganz aus dem Blickfeld, denn Validierung ist ja nicht das Ziel.

³⁰ Vgl. Gardin, op.cit. S.4-10.

Hier könnte der Einsatz neuer, auf den Computer gestützter Darstellungs- und Analysemethoden Mentor sein auf dem Weg zu einer eigenständigen Wissenschaftlichkeit der Geisteswissenschaften. Ob ihnen die Erstellung besonders großer Wissensbasen jedoch allein schon den Status von kumulativen Wissenschaften verleiht, sei dahingestellt. Die neuen Instrumentarien machen aber zumindest immer dringlicher darauf aufmerksam, daß auch die Geisteswissenschaften mit harten, verifizierbaren Methoden arbeiten können. Auf beiden Seiten des epistemologischen Systems der Zeichenverarbeitung (und dies auf der Objekt- wie auch auf der Metaebene des Wissenschaftsdiskurses) gibt der Computer Hilfestellung und schafft Notwendigkeiten des Umdenkens.