ZA-Information 30

Mai 1992

Dieses Dokument wurde von einer gedruckten Vorlage eingescannt und mit einer OCR-Software weitgehend automatisch bearbeitet. Kleinere Übertragungsfehler sind daher möglich.

ZA-Information 30

Mai 1992

Herausgeber:
Zentralarchiv für empirische Sozialforschung
Universität zu Köln
Das Zentralarchiv ist Mitglied der GESIS

Direktor: Prof. Dr. E. K. Scheuch
Geschäftsführer: E. Mochmann

Anschrift:
Bachemer Straße 40
5000 Köln 41

Telefon:

Zentrale	0221/4 76 94- 0
Telefax	- 44
Redaktion	- 50

Redaktion:
Franz Bauske

ISSN: 0723-5607
© Zentralarchiv

Die ZA-INFORMATION erscheint jeweils im Mai und November eines Jahres.
Sie wird kostenlos an Interessenten und Benutzer des Zentralarchivs abgegeben.

inhalt

ZA-Information 30 **Mai 1992**

Mitteilungen der Redaktion 5

Berichte aus dem Archiv

Informationsquellen über die Datenbestände des Zentralarchivs 6

Erweiterungen im Datenangebot des Zentralarchivs 7

Daten zum Ost-West-Vergleich in der Bundesrepublik:
Die Basisumfrage 1991, eine Sondererhebung aus der ALLBUS-Reihe . . 9

Die Datensätze des Forschungsinstituts der Konrad-Adenauer-Stiftung im Zentralarchiv 12

Bericht über das Frühjahrsseminar 1992 20

Zur Analyse von Paneldaten mit SPSS/PC: Die EGLS-Schätzung des Fehlerkomponentenmodells
von ***Steffen Kühnel*** 23

RZOO: Ein Hilfsmittel zur Verwaltung sozialwissenschaftlicher Daten
von ***Götz Rohwer*** 43

Eine geschlossene Kollektion empirischer Studien der Hochschule für Architektur und Bauwesen Weimar im Zentralarchiv
von ***Evelyn Brislinger*** und ***Brigitte Hausstein*** 46

Erste Kontakte zum spanischen Zentrum für Soziologische Forschung CIS 58

Forschungsnotizen

Zur aktuellen Situation von Glauben und Kirche im vereinigten Deutschland: Eine Analyse der Basisumfrage 1991
von ***Michael Terwey*** 59

Regressionsanalyse mit Zeitreihendaten: Zur Interpretation dynamischer Modelle
von ***Helmut Thome*** 80

Clusteranalyse und Korrespondenzanalyse in der Lebensstilforschung: Ein Vergleich am Beispiel der Wohnungseinrichtung
von ***Jörg Blasius*** und **Werner *Georg*** 112

Berichte, Ankündigungen und Mitteilungen

Workshop: Einführung in die Korrespondenzanalyse 134

Erleichterung computergestützter Inhaltsanalysen durch verbesserte automatische Texterkennung
von ***Peter Hauptmanns*** und ***Rainer Schnell*** 135

Politischer Umbruch und Kriminalitätsentwicklung:
Kongreß in Budapest 139

Buchhinweise:
Johannes Gogolok, Rudolf Schuemer und ***Gerhard Ströhlein:***
Datenverarbeitung und statistische Auswertung mit SAS 140

Reinhard Wittenberg:
Handbuch für computerunterstützte Datenanalyse:
Band 1: Grundlagen computerunterstützter Datenanalyse 142

Reinhard Wittenberg und ***Hans Cramer:***
Band 2: Datenanalyse mit SPSS 142

Dieter Jaufmann, Ernst Kistler, Klaus Meier und ***Karl-Heinz Strech:***
Empirische Sozialforschung im vereinten Deutschland:
Bestandsaufnahme und Perspektiven 143

Hans-Jürgen Andreß:
Einführung in die Verlaufsdatenanalyse 145

Tagungsankündigungen:
ZHSF-Herbstseminare 1992 147

Informationen zum 26. Deutschen Soziologentag:
Lebensverhältnisse und soziale Konflikte im neuen Europa 150

Bei Beiträgen, die **nicht** von Mitarbeitern des Zentralarchivs verfaßt wurden, ist die Anschrift des Autors beim jeweiligen Artikel angegeben.
Die Inhalte der Beiträge entsprechen der Meinung der Autoren und geben nicht unbedingt in jedem Punkt die Ansicht der Redaktion wieder.

Mitteilungen der Redaktion

Auch in dieser Ausgabe bieten wir wieder Informationen zum Ost-West-Vergleich in der Bundesrepublik. So wollen wir auf die Sondererhebung der ALLBUS-Reihe, die Basisumfrage 1991, hinweisen. Neben einem Gesamtüberblick über die Befragungsinhalte präsentiert ***Michael Terwey*** eine erste Auswertung eines Teilaspektes dieser Befragung. In seiner Analyse geht er auf die aktuelle Situation von Glauben und Kirche im vereinigten Deutschland ein und stellt fest, daß es gravierende Unterschiede in den beiden Landesteilen gibt. Einen Ost-West-Vergleich erlauben auch die jüngst archivierten Studien vom Forschungsinstitut der Konrad-Adenauer-Stiftung. Bei dieser Gelegenheit stellen wir die Gesamtheit der Studien vor, die wir im Laufe der letzten Jahre von dieser Stiftung erhalten haben. ***Evelyn Brislinger*** und ***Brigitte Hausstein*** berichten über eine Tagung in Weimar, die sich auch mit der Sicherung der Datenbestände im Ostteil der Republik beschäftigte. Sie stellen auch die Neuzugänge aus der Hochschule für Architektur und Bauwesen in Weimar vor.

Einen weiteren Schwerpunkt dieser Ausgabe bilden Beiträge, die im Kontext zum letzten Frühjahrsseminar zu sehen sind. Neben einem Bericht über das Seminar stellt ***Steffen Kühnel*** ein Verfahren zur Analyse von Paneldaten vor, das mit SPSS/PC berechnet werden kann. ***Helmut Thome*** geht in seinem Beitrag auf die Interpretation dynamischer Modelle mittels Regressionsanalyse von Zeitreihendaten ein. Schließlich stellt ***Götz Rohwer*** ein Programm zur effizienten Verwaltung sozialwissenschaftlicher Daten vor, über das er im Frühjahrsseminar berichtet hat.

Nachdem TexterkennungsSoftware und Scanner preiswerter geworden sind, gibt es neue Möglichkeiten für die computergestützte Inhaltsanalyse. ***Peter Hauptmanns*** und ***Rainer Schnell*** haben einige Tests durchgeführt, die etwas über den Stand dieser Technik aussagen.

Franz Bauske

Informationsquellen über die Datenbestände des Zentralarchivs

Das Zentralarchiv hält für seine Benutzer zwei gedruckte Informationsquellen bereit, die Auskunft über Art und Umfang der Bestände des Archivs vermitteln.

Datenbestandskatalog

Der Datenbestandskatalog bietet die umfangreichste Übersicht über die archivierten Datensätze, die bis Mitte des Jahres 1990 akquiriert und aufbereitet worden sind. Gegenüber der ersten Auflage wurde der Katalog vor allem durch ein Namensregister der Primärforscher, eine Auflistung der Erhebungsinstitute sowie die inzwischen erfolgten Neuzugänge ergänzt. Darüber hinaus enthält er auch das Datenangebot des Zentrums für Historische Sozialforschung (ZHSF), das seit einiger Zeit Teil des Zentralarchivs ist.

Der Katalog bietet Interessenten die Möglichkeit einer ersten Auswahl von relevanten Datensätzen, ohne daß eine direkte Suche in den Beständen des Archivs notwendig ist.

Der Datenbestandskatalog wird vom Campus-Verlag in der Reihe des Zentralarchivs "Beiträge zur Empirischen Sozialforschung" verlegt und über den Buchhandel in einer Hard-Cover-Version vertrieben. Unseren Benutzern können wir den über 800seitigen Katalog in einer Soft-Cover-Version zum Sonderpreis von DM 48.- anbieten.

Datenbestandsverzeichnis

Die aktuellste Auflistung der Bestände des Archivs befindet sich jeweils im Datenbestandsverzeichnis, das im etwa ein- bis zweijährlichen Turnus erscheint. Es enthält lediglich die Auflistung der Archivtitel. Das 88-seitige Verzeichnis enthält auch die Benutzungs- und Gebührenordnung des Zentralarchivs sowie eine Übersicht über das Dienstleistungsangebot des Instituts. Das kostenlose Verzeichnis erhält man durch einen Anruf (0221/47 69 4-0) oder durch ein Telefax (0221/47 69 4-44).

Erweiterungen im Datenangebot des Zentralarchivs

Die Neuzugänge der letzten Monate sind in dieser Liste unter Angabe der Archivnummer, Studientitel, Primärforscher bzw. Erhebungsinstitute sowie des Erhebungsjahres aufgeführt. Weitere Details zu den einzelnen Datensätzen sind auf Anfrage in Form von *Studienbeschreibungen* erhältlich.

1990 ALLBUS Basisumfrage 1991 in Gesamtdeutschland (siehe Beschreibung auf Seite 9)

2001 Junge Generation und europäische Einigung
Institut für Politikwissenschaft, Universität Mainz;
Infratest, München; 1987

2002 Umfragen in Deutschland
CONTEXT, Düsseldorf;
FORSA, Dortmund; 1991

2005 AIDS im öffentlichen Bewußtsein der Bundesrepublik (1990)
Bundeszentrale für gesundheitliche Aufklärung, Köln;
FORSA, Dortmund

2009 Media-Analyse (MA 91, Funkmedien)
2010 Media-Analyse (MA 91, Pressemedien)
Media-Micro-Census, Frankfurt (Main)

2031 Eurobarometer 35.0
2032 Eurobarometer 35.1
2081 Eurobarometer 36
Kommission der Europäischen Gemeinschaften, Brüssel

2040 Berichtssystem Weiterbildung '88
Bundesministerium für Bildung und Wissenschaft, Bonn;
Infratest, München; 1988

2043 Lagebericht Mittelstand 1989/90
2044 Lagebericht Mittelstand im Sommer 1990
2045 Lagebericht Mittelstand 1990/91
2046 Lagebericht Mittelstand im Sommer 1991
Westdeutsche Genossenschafts-Zentralbank, Düsseldorf

2050 Direktinvestitionen mittelständischer Unternehmen in Schwellenländern (Persönliche Befragung)
2051 Direktinvestitionen mittelständischer Unternehmen in Schwellenländern (Telefoninterview)
Forschungsstelle für empirische Sozialökonomik, Köln, 1986

2082 Typologie der Wünsche 1991
Burda, Offenburg;
Basis Research, Frankfurt (Main); GFM-Getas, Hamburg

2091 Politbarometer West (Januar 1991)
2092 Politbarometer West (Februar 1991)
2093 Politbarometer West (März 1991)
2094 Politbarometer West (April 1991)
2095 Politbarometer West (Mai 1991)
2096 Politbarometer West (Juni 1991)
2097 Politbarometer West (August 1991)
2098 Politbarometer West (September 1991)
2099 Politbarometer West (Oktober 1991)
2100 Politbarometer West (November 1991)
2101 Politbarometer West (Dezember 1991)
2103 Politbarometer Ost (Januar 1991)
2104 Politbarometer Ost (Februar 1991)
2105 Politbarometer Ost (März 1991)
2106 Politbarometer Ost (April 1991)
2107 Politbarometer Ost (Mai 1991)
2108 Politbarometer Ost (Juni 1991)
2109 Politbarometer Ost (August 1991)
2110 Politbarometer Ost (September 1991)
2111 Politbarometer Ost (Oktober 1991)
2112 Politbarometer Ost (November 1991)
2113 Politbarometer Ost (Dezember 1991)
Forschungsgruppe Wahlen, Mannheim

2115 Landtagswahl in Hessen 1991 (Vor Beginn des Golfkrieges)
2116 Landtagswahl in Hessen 1991 (Nach Beginn des Golfkrieges)
2117 Landtagswahl in Rheinland-Pfalz 1991
2118 Bürgerschaftswahl in Hamburg 1991
2119 Bürgerschaftswahl in Bremen 1991
Forschungsgruppe Wahlen, Mannheim

2120 Einstellungen zu aktuellen Fragen der Innenpolitik (1991) - West
2121 Einstellungen zu aktuellen Fragen der Innenpolitik (1991) - Ost
2122 Umweltschutz - West
2123 Umweltschutz - Ost
Institut für praxisorientierte Sozialforschung, Mannheim; 1991

2132 Frühjahrsstudie 1990
2133 Herbststudie 1990
2134 Einstellung zur Vereinigung der beiden deutschen Staaten (März 1990)
2135 Einstellung zur Vereinigung der beiden deutschen Staaten (August 1990)
2136 Vor der Bundestagswahl 1990
Forschungsinsitut der Konrad-Adenauer-Stiftung, St. Augustin
(siehe dazu auch die Beschreibung auf den Seiten 12 bis 19)

Daten zum Ost-West-Vergleich in der Bundesrepublik: Die Basisumfrage 1991, eine Sondererhebung aus der ALLBUS-Reihe.

Die Basisumfrage 1991 (Gesamtdeutschland) steht für Sekundäranalysen zur Verfügung. Die folgende Studienbeschreibung gibt einen Überblick über die Gesamtheit des Inhalts dieser Studie. ***Michael Terwey*** analysiert diese Daten in einem Beitrag auf den Seiten 59 bis 79 dieses Heftes im Hinblick auf die aktuelle Situation von Glauben und Kirche im vereinigten Deutschland.

Basisumfrage 1991

Erhebungszeitraum:
Mai 1991 bis Juli 1991

Primärforscher:
P. Ph. Mohler, M. Braun, M. Häder (ZUMA, Mannheim)
E. K. Scheuch (Universität zu Köln)

Datenerhebung:
Infratest, München

Inhalt:
Trenderhebung zur Erfassung des sozialen Wandels in der Bundesrepublik. Die Schwerpunkte dieser Untersuchung sind:
1.) Einstellungen zu Politik und Wirtschaft
2.) Einstellungen zum Vereinigungsprozeß
3.) Arbeit und Beruf
4.) Einstellungen zur sozialen Ungleichheit
5.) Einstellungen zu Ehe und Familie
6.) Sonstiges
7.) Schriftliche Befragung im Rahmen des ISSP: Einstellung zu religiösen Verhaltensweisen
8.) Standarddemographie

Themen: 1.) Einschätzung des Sozialismus als Idee; Zufriedenheit mit der Demokratie in der Bundesrepublik; Beurteilung des Funktionierens unseres politischen Systems; Zufriedenheit mit den Leistungen der Bundesregierung und der Landesregierung; Wahlberechtigung **und** Wahlbeteiligung bei der letzten Bundestagswahl; Wahlrückerinnerungsfrage; Wahlabsicht;

Nationalstolz und Stolz auf deutsche Einrichtungen und deutsche Leistungen; wichtigste Kritikpunkte an der ehemaligen DDR; politisches Interesse; Postmaterialismus; politische Partizipation; Präferenz für Steuersenkung oder erhöhte soziale Leistungen; Einstellung zum Zuzug von Aussiedlern, Asylsuchenden, EG-Arbeitnehmern und Nicht-EG-Arbeitnehmern; Selbsteinstufung auf einem Links-Rechts-Kontinuum; Beurteilung der derzeitigen und zukünftigen Wirtschaftslage in der Bundesrepublik und im eigenen Bundesland; Beurteilung der eigenen derzeitigen und zukünftigen wirtschaftlichen Situation; Einschätzung der Wirtschaftsentwicklung in den alten und den neuen Bundesländern.

2.) Beurteilung der Lastenverteilung zwischen alten und neuen Bundesländern nach der Vereinigung; Vor- und Nachteile der Wiedervereinigung; wechselseitige Fremdheit der Bürger; Leistungsdruck in den neuen Bundesländern; Einstellung zum Umgang mit der Stasi-Vergangenheit des Einzelnen.

3.) Arbeitsorientierung und wichtigste Charakteristika des Berufs; Interesse, sich künftig beruflich selbständig zu machen; eigene berufliche Selbständigkeit in der Vergangenheit; Einstellung zur Berufstätigkeit von Frauen und Müttern.

4.) Gerechter Anteil am Lebensstandard; wichtigste Voraussetzungen für Erfolg in der Gesellschaft; Einstellung zum Sozialstaat und zu sozialen Unterschieden.

5.) Einstellung zur Familie und zur Heirat; Einstellung zur Berufstätigkeit von Frauen; Taufe der Kinder; regionale Herkunft des Lebenspartners bzw. Ehepartners.

6.) Erziehungsziele; allgemeines Vertrauen zu Mitmenschen und zu Politikern; Optimismus und Zukunftsorientierung (Anomie); Ortsverbundenheit und Verbundenheit mit dem Bundesland, mit der alten Bundesrepublik bzw. DDR, mit Gesamtdeutschland sowie mit der EG; Wichtigkeit von Lebensbereichen.

7.) In Form einer schriftlichen Zusatzbefragung für das "International Social Survey Programme" (ISSP) wurde befragt: persönliche Glückseinschätzung; Einschätzung der sozialen Verantwortung des Staates bezüglich der Arbeitsbeschaffung und der Einkommensnivellierung; härtere Strafen und Todesstrafe als Maßnahmen zur Verbrechensbekämpfung; Einstellung zum vorehelichen Geschlechtsverkehr und zum Seitensprung in der Ehe; Einstellung zur Homosexualität und zur Abtreibung; Beurteilung der Rollenverteilung in der Ehe und Einstellung zu berufstätigen Frauen; Steuerehrlichkeit und Einstellung zur Ehrlichkeit des Bürgers gegenüber dem Staat; Vertrauen in Institutionen wie Bundestag, Handel, Industrie, Behörden, Kirchen, Gerichte und Schulen; Einstellung zu gottesungläubigen Politikern und Amtsträgern; Beeinflussung der Wähler sowie der Regierung durch Kirchenführer; Beurteilung der Macht von Kirchen und religiösen Organisationen; Zweifel oder fester Gottesglaube; empfundene Nähe zu Gott; Entwicklung des persönlichen Glaubens an Gott; Glauben an ein Leben nach dem Tod; Glaube an den Teufel, den Himmel, die Hölle und an Wunder; Überzeugung gegenüber der Bibel; Fatalismus; Sinn des Lebens und christliche Lebensdeutung; Verbindung mit Verstorbenen; religiöse Bindung durch einen Wendepunkt im Leben; Religionszugehörigkeit des Vaters und der Mutter; eigene Religionszugehörigkeit; Kirchgangshäufigkeit

des Vaters, der Mutter und des Partners; eigene Glaubensrichtung und Kirchgangshäufigkeit im Jugendalter; Häufigkeit des Betens und der Teilnahme an religiösen Aktivitäten; Selbsteinstufung eigener Religiosität; Einstellung zum Schulgebet; eigenes Gewissen, gesellschaftliche Regeln oder Gottes Gesetze als Entscheidungsbasis für richtig oder falsch; Einstellungen zum Verbot von Religionskritik in Literatur und Filmen; Aufgaben der Kirche; Aberglaube; Glaube an Glücksbringer, Wahrsager, Wunderheiler, Sternzeichen und Horoskope.
8.) ALLBUS-Standarddemographie.
9.) Zusätzlich vercodet wurde: Anwesenheit des Befragten in den letzten 4 Tagen vor dem mündlichen Interview; Interviewdauer.
10.) Abgeleitete Indizes: Inglehart-Index; Familientypologie; Haushaltsklassifikationen; Klassenlagenschema nach Goldthorpe und verschiedene Einordnungsberufsvariablen.

Grundgesamtheit und Auswahl:
Untersuchungsgebiet: BRD
Mehrstufig geschichtete Zufallsauswahl in den alten und neuen Bundesländern aus allen Personen, die in Deutschland in Privathaushalten wohnen und die bis zum Befragungstag das 18. Lebensjahr vollendet hatten. Grundlage für die Auswahl war in Westdeutschland das von Infratest modifizierte ADM-Design, in Ostdeutschland das Infratest-Mastersample Ost mit jeweils anschließendem Random-Route.

Erhebungsverfahren:
Mündliche Befragung mit standardisiertem Fragebogen, ergänzt um eine schriftliche Zusatzbefragung.

Datensatz:
Anzahl der Einheiten: 3058
Anzahl der Variablen: 518

Weitere Hinweise:
Die übrigen ALLBUS-Erhebungen seit 1980 sind unter den ZA-Studien-Nrn. 1000, 1160, 1340, 1500, 1601, 1669, 1670 und 1800 archiviert. Darüber hinaus gibt es kumulierte Datensätze dieser Befragungen, die unter den ZA-Studien-Nrn. 1335 und 1795 zu finden sind.
Bei einer Reihe von Fragen handelt es sich um Replikationen aus früheren ALLBUS- und ISSP-Untersuchungen.
Bei einem Teil der Fragen aus der Standarddemographie waren die Antwortvorgaben an die jeweiligen Besonderheiten in Ost- und Westdeutschland angepaßt. Die Frage nach Kritikpunkten an der früheren DDR wurde nur in Ostdeutschland gestellt.
Der Stichprobenansatz war so gestaltet, daß die Zahl der auswertbaren Interviews ca. 3000 betrug, davon jeweils ca. 1500 aus West- und Ostdeutschland.

Die Datensätze des Forschungsinstituts der Konrad-Adenauer-Stiftung im Zentralarchiv

Das Forschungsinstitut der Konrad-Adenauer-Stiftung führt regelmäßig Studien zur politischen Kultur in der Bundesrepublik durch. Seit rund zwei Jahrzehnten werden diese Datensätze an das Zentralarchiv weitergegeben. Im Laufe der Zeit hat sich die Stiftung zu einem der bedeutendsten Datengeber des Zentralarchivs entwickelt. Die jüngst zum Thema der deutschen Vereinigung ins Zentralarchiv eingebrachten Datensätze sind Anlaß für die folgende Zusammenstellung der Gesamtheit der eingebrachten Bestände und anschließende Präsentation der aktuellen Datensätze, die wir von diesem Institut erhalten haben. Zuvor jedoch eine kurze Information über das Forschungsinstitut.

Das 1967 gegründete Forschungsinstitut der Konrad-Adenauer-Stiftung, in St. Augustin bei Bonn, ist in vielen Feldern der Politik- und Sozialwissenschaft tätig. Es erarbeitet wissenschaftliche Grundlagen und Expertisen für politisches Handeln auf nationaler sowie internationaler Ebene. Defizite der politik- und sozialwissenschaftlichen Forschung werden aufgearbeitet und aktuelle politische Probleme wissenschaftlich analysiert sowie künftige Entwicklungen aufgezeigt. Im Institut sind 60 Mitarbeiter, etwa die Hälfte als Wissenschaftler, beschäftigt. Ihm ist ein eigenes Rechenzentrum für die Umfrageforschung angeschlossen. Das Forschungsinstitut gliedert sich in folgende Forschungsbereiche:

- Empirische Sozialforschung mit politischer Meinungs-, Wahl-, Werte- und Verhaltensforschung (Austauschpartner für das Zentralarchiv);
- Vergleichende Institutionenforschung, insbesondere von Parteien und Gewerkschaften auf europäischer Ebene;
- Außen-, Sicherheits- und Entwicklungspolitik mit den Schwerpunkten Atlantisches Bündnis, Sowjetunion, Osteuropa und globalstrategische Sicherheitsfragen;
- Europa- und Deutschlandpolitik, wobei Fragen des europäischen Einigungsprozesses sowie Entwicklungen und Probleme der Deutschlandpolitik im Vordergrund stehen.

Die Arbeit des Forschungsinstituts findet ihren Niederschlag in wissenschaftlichen Studien, in Veröffentlichungen größerer Monographien (Institutsreihe: Studien zur Politik), in Sammelbänden, in Forschungsberichten sowie in Beiträgen für Fachzeitschriften. Darüber hinaus führt das Institut Konferenzen zu aktuellen politischen Fragen durch, die dem Austausch von Wissenschaft und Politik dienen.

Der Forschungsbereich Empirische Sozialforschung analysiert Entwicklungen der politischen Kultur in der Bundesrepublik und ihre Auswirkungen auf die Legitimationsstruktur und das Parteiensystem. Zur Erfassung der Veränderungen im politischen Klima der Bundesrepublik läßt das Institut bundesweit Repräsentativuntersuchungen durchführen. Auf der Basis dieser empirischen Daten - die Trendreihen reichen bis in das Jahr 1971 zurück - ist es möglich, aktuelle politische Analysen wie auch theoretische Beiträge zur Fortentwicklung der empirischen Sozialforschung zu erstellen.

Das Forschungsinstitut bezeichnet seine halbjährlich durchgeführten Umfragen zur politischen Situation in der Bundesrepublik als Frühjahrs- bzw. Herbststudien. Unter diesem Titel sind die Datensätze dann auch vom Zentralarchiv übernommen worden, obwohl eine Bezeichnung wie etwa Vor- bzw. Nachwahluntersuchung aufschlußreicher wäre. Einige Studien sind auch als mehrwellige Paneluntersuchungen angelegt. Neben die traditionelle Langzeitbeobachtung von Entwicklungen der politischen Kultur in den alten Bundesländern trat in jüngster Zeit die Analyse von Einstellungen und Einstellungsveränderungen zum Fortgang des Vereinigungsprozesses. Allein 1990 wurden insgesamt 10 größere Meinungsumfragen sowohl im Westen als auch im Osten der Bundesrepublik durchgeführt.

Über die Forschung zu gesamtdeutschen Fragen hinaus werden Analysen zu Landtagswahlen und zu besonderen politischen Fragen durchgeführt. So wurden beispielsweise Spezial-Studien zur Situation der Studenten in der Bundesrepublik, der Jugend, der Frauenarbeitslosigkeit und weiteren aktuellen Themen durchgeführt. Die aktuellsten Untersuchungen befassen sich mit Fragen der Vereinigung, schwerpunktmäßig mit der Analyse der Beurteilung des Einigungsprozesses und dessen Folgen durch die Bürger in Ost und West. Die detaillierte Darstellung der Frageninhalte veranschaulicht die umfassende Untersuchung dieses Themas. Zunächst jedoch die Liste der Studien, die von der Konrad-Adenauer-Stiftung seit den 70er Jahren in das Zentralarchiv eingebracht wurden, sowie Anmerkungen über nicht durchgeführte Erhebungen, die den halbjährlichen Rhythmus unterbrechen. Die vorangehende Ziffer ist die Archivnummer im Zentralarchiv.

1501 Frühjahrsstudie 1972
Keine Herbststudie 1972
1502 Frühjahrsstudie 1973
1322 Länderstudie Herbst 1973
1236 Länderstudie Frühjahr 1974
1126 Länderstudie Herbst 1974
1127 Länderstudie Frühjahr 1975
Keine Herbststudie 1975
1503 Frühjahrsstudie 1976
Keine Herbststudie 1976
1504 Frühjahrsstudie 1977
1505 Herbststudie 1977
1506 Frühjahrsstudie 1978
Keine Herbststudie 1978
1507 Frühjahrsstudie 1979
1508 Herbststudie 1979
1450 Frühjahrsstudie 1980
Keine Herbststudie 1980
1451 Frühjahrsbus 1981

1452 Herbststudie 1981
1453 Frühjahrsstudie 1982
1454 Herbststudie 1982
1455 Frühjahrsstudie 1983
1456 Herbststudie 1983
1807 Frühjahrsstudie 1984
1808 Herbststudie 1984
Keine Frühjahrs Studie 1985
1809 Herbststudie 1985
1810 Frühjahrsstudie 1986
1811 Herbststudie 1986
1953 Frühjahrsstudie 1987
Keine Herbststudie 1987
Keine Frühjahrsstudie 1988
1954 Herbststudie 1988
1955 Frühjahrsstudie 1989
1956 Herbststudie 1989
2132 Frühjahrsstudie 1990
2133 Herbststudie 1990

0839 Bundestagswahl 1972 (Panel: 1. Welle, April - Mai 1971)
0840 Bundestagswahl 1972 (Panel: 2. Welle, Oktober 1971 - Januar 1972)
0841 Bundestagswahl 1972 (Panel: 3. Welle, September - November 1972)
0842 Bundestagswahl 1972 (Panel: 4. Welle, September 1972 - Januar 1973)

1133 Bundestagswahl 1976 (Panel: 1. Welle, Oktober 1973)
1134 Bundestagswahl 1976 (Panel: 2. Welle, November 1975)
1135 Bundestagswahl 1976 (Panel: 3. Welle, November 1976)

1323 Bundestagswahl 1980 (Panel: 1. Welle, November 1978)
1324 Bundestagswahl 1980 (Panel: 2. Welle, Oktober 1979)
1325 Bundestagswahl 1980 (Panel: 3. Welle, September 1980)

1949 Bundestagswahl 1983 (Panel: 1. Welle, September 1980)
Bundestagswahl 1983 (Panel: 2. Welle, Februar 1982)
Bundestagswahl 1983 (Panel: 3. Welle, Februar 1984)
Kein Wahl-Panel 1987
Kein Wahl-Panel 1990

0773 Politische Einstellungen (Vorwahluntersuchung, Oktober 1968)
0774 Politische Einstellungen (Vorwahluntersuchung, November 1968)
0775 Politische Einstellungen (Vorwahluntersuchung, Dezember 1968)
0776 Politische Einstellungen (Vorwahluntersuchung, Januar 1969)

0777 Politische Einstellungen (Vorwahluntersuchung, Februar 1969)
0778 Politische Einstellungen (Vorwahluntersuchung, März 1969)
0779 Politische Einstellungen (Vorwahluntersuchung, April 1969)
0780 Politische Einstellungen (Vorwahluntersuchung, Mai 1969)
0781 Politische Einstellungen (Vorwahluntersuchung, Juni 1969)
0782 Politische Einstellungen (Vorwahluntersuchung, Juli 1969)
0783 Politische Einstellungen (Vorwahluntersuchung, August 1969)
0784 Politische Einstellungen (Vorwahluntersuchung, September 1969)
0785 Politische Einstellungen (Vorwahluntersuchung, Oktober 1969)
0786 Politische Einstellungen (Vorwahluntersuchung, November 1969)
07 87 Politische Einstellungen (1967)

0788 Die politische Stimmung nach der Bundestagswahl 1969
0789 Resonanz der Bevölkerung auf den Wahlkampf und das Wahlergebnis der Bundestagswahl 1969
0790 Politische Motiv- und Meinungsverteilung in Schleswig-Holstein 1970
0791 Die politische Stimmung in Bayern 1970
0792 Voraussetzungen politischer Meinungsbildung in der Bundesrepublik 1964 (4 Befragungen)
0796 Westdeutsche Führungsschicht 1972
0820 Die Situation der Studenten in der Bundesrepublik 1975
1086 Auswirkungen der Arbeitslosigkeit bei Frauen (Panel) 1978
1168 Großstadt und Umland (Regionalstudie Köln/Bonn) 1972
1449 Jugendstudie (1983)
1957 Bundestagswahl 1987 (Nachwahluntersuchung)
1958 Einstellungen zur Europäischen Gemeinschaft 1988
2134 Einstellung zur Vereinigung der beiden deutschen Staaten (März 1990)
2135 Einstellung zur Vereinigung der beiden deutschen Staaten (August 1990)
2136 Vor der Bundestagswahl 1990

Es folgt die detaillierte Beschreibung der jüngsten Studien.

Frühjahrsstudie 1990 (ZA-Nr. 2132)
Erhebungszeitraum: März 1990
Datenerhebung: EMNID, Bielefeld
Inhalt
Beurteilung von Parteien und Politikern. Politische Einstellungen.
Themen: Beurteilung der allgemeinen sowie persönlichen wirtschaftlichen Lage sowie Einschätzung der weiteren Entwicklung in einem Jahr; Politikinteresse; geschätzter Einfluß der Wähler auf die Politik; größte politische Probleme; wichtigste gesellschaftliche Ziele (Skala);

eigene Wechselwählerschaft; Wahlverhalten bei der letzten Bundestagswahl 1987; Zufriedenheit mit der gewählten Partei; Parteipräferenz (Erst- und Zweitstimme, Stimmzettelverfahren); Einstellung zu den Republikanern im Bundestag und Akzeptanz der Republikaner; Einstellung zur Nato und zu einem neutralen Europa; Nationalstolz; Demokratiezufriedenheit; Selbsteinschätzung und Einstufung der Bundesparteien auf einem Links-Rechts-Kontinuum; Parteineigung; Beurteilung des Gerechtigkeitsstrebens; empfundene Benachteiligung in der eigenen gesellschaftlichen Position und erwartete zukünftige Veränderungen; Einstellung zur Aufnahme von Übersiedlern aus der DDR, von deutschstämmigen Aussiedlern aus den Ostblockstaaten und Gastarbeitern sowie Asylsuchenden; detaillierte demografische Angaben über die häufigsten Gesprächspartner und Verhältnis der Gesprächspartner untereinander; Parteipräferenz der Hauptkontaktpersonen; Kirchennähe; Furcht vor Arbeitslosigkeit oder Kurzarbeit; Selbsteinschätzung der Schichtzugehörigkeit; Gewerkschaftsmitgliedschaft und Verbundenheit mit den Gewerkschaften; Tätigkeit im öffentlichen Dienst; hauptsächliche oder teilweise Erledigung von Arbeiten im Haushalt.

Übersiedler wurden zusätzlich gefragt: Zeitpunkt der Übersiedlung in die BRD.

Grundgesamtheit und Auswahl

Untersuchungsgebiet: BRD

Mündliche Befragung mit standardisiertem Fragebogen

Herbststudie 1990 (ZA-Nr. 2133)

Erhebungszeitraum: August bis September 1990

Datenerhebung: GFM-GETAS, Hamburg

Inhalt

Beurteilung von Parteien und Politikern. Politische Einstellungen. Einstellungen zur Vereinigung der beiden deutschen Staaten.

Themen: Zukunftszuversicht; Beurteilung der wirtschaftlichen Lage in der Bundesrepublik und in der DDR; Einschätzung der weiteren wirtschaftlichen Entwicklung im Ost- bzw. Westteil Deutschlands; eigene finanzielle Lage und erwartete Entwicklung; Einschätzung der Entwicklung der Arbeitslosigkeit im Lande; Politikinteresse; wichtigste politische Probleme und Lösungskompetenz der Parteien; Parteineigung und Parteiverbundenheit; eigene Wechselwählerschaft und Wahlverhalten bei der letzten Bundestagswahl bzw. Volkskammerwahl; Zufriedenheit mit den gewählten Parteien; Parteipräferenz (Stimmzettelverfahren, Erst- und Zweitstimme); Bedeutung von ausgewählten Wertvorstellungen und Wichtigkeit gesellschaftlicher Ziele (Skala); Issue-Relevanz; Issue-Kompetenz der Parteien; Einstellung zu einem Einzug der Republikaner, der PDS bzw. der Grünen in den Bundestag; Beurteilung der Chancen dieser Parteien bei einer gesamtdeutschen Wahl sowie der Bereitschaft, diese Parteien zu wählen; Charakterisierung der CDU/CSU sowie der SPD und Beurteilung ihres Erscheinungsbildes (Skala); wichtigste und vertrauenswürdigste Informationsquellen bezüglich politischer Fragen; Einstellung zur Nato; eigene Meinungsführerschaft bzw. Meinungsgefolgschaft; Sympathie-Skalometer für die Parteien und die Kanzlerkandidaten; Selbsteinschät-

zung auf einem Links-Rechts-Kontinuum; Beurteilung der Vereinigungsgeschwindigkeit; Einstellung zur Vereinigung der beiden deutschen Staaten; Präferenz für ein marktwirtschaftliches oder ein sozialistisches System in der Bundesrepublik; Vorrang der Lösung von Umweltfragen vor allen übrigen Problemen; Auswirkung der Vereinigung auf die Bedeutung der europäischen Einigung, auf die Geldwertstabilität und den Lebensstandard im Ost- und Westteil Deutschlands; Nationalstolz; vermutete Beliebtheit der Deutschen im Ausland; Religiosität; Verbundenheit mit der Kirche; Beschäftigung im öffentlichen Dienst; Mitgliedschaft in einem Verein, in einer Bürgerinitiative, einer Partei oder einer Berufsvereinigung sowie übernommene Ämter; Selbsteinschätzung der Schichtzugehörigkeit und Einstufung der Schichtzugehörigkeit der Eltern; Gewerkschaftsmitgliedschaft; Gewerkschaftsamt; Gewerkschaftsverbundenheit; Telefonbesitz; Angst vor Arbeitslosigkeit bzw. Kurzarbeit; Beurteilung der Fähigkeit der Bundesrepublik zum Aufbau der DDR-Wirtschaft; Kompetenz der Bundesregierung bzw. der SPD in Fragen der Vereinigung; Beurteilung des Zeitpunkts der Wirtschafts- und Währungsunion sowie erwartete Auswirkungen für die DDR-Bürger; erwartete Dauer bis zu einer wirtschaftlichen Besserung in der DDR; Ausrichtung der zukünftigen Bundesrepublik an den Grundsätzen der DDR oder der BRD; erwartete kurzfristige und langfristige Folgen der Vereinigung.

In der Bundesrepublik wurde zusätzlich gefragt: Präferenz für eine Begrenzung der öffentlichen Gelder für den wirtschaftlichen Aufbau der DDR; Aussiedler oder Übersiedler bzw. Flüchtling aus der DDR; Jahr der Übersiedlung.

Interviewerrating: Anwesenheit anderer Personen beim Interview und deren Verwandtschaft zum Befragten; Eingriff Dritter in das Interview; Kooperationsbereitschaft des Befragten und Zuverlässigkeit der Antwortgebung.

Zusätzlich verkodet wurden: Interviewdauer; Interviewdatum; Interviewergeschlecht.

Grundgesamtheit und Auswahl

Untersuchungsgebiet: BRD und DDR

Mündliche Befragung mit standardisiertem Fragebogen

Einstellung zur Vereinigung der beiden deutschen Staaten (ZA-Nr. 2134)

Erhebungszeitraum: März bis April 1990

Datenerhebung: EMNID, Bielefeld

Inhalt

Einstellung zur Vereinigung der beiden deutschen Staaten.

Themen: Zukunftszuversicht; Beurteilung der eigenen wirtschaftlichen Lage zum derzeitigen Zeitpunkt sowie in einem Jahr; präferierte Art der Zusammenarbeit von BRD und USA; Nationalstolz; Einstellung zur Aufnahme von Übersiedlern aus der DDR, deutschstämmigen Aussiedlern aus den Ostblockstaaten, von Gastarbeitern und Asylsuchenden; Einschätzung der daraus entstehenden Belastungen für die Bundesrepublik; Einstellung zur Vereinigung der beiden deutschen Staaten; kurzfristig und langfristig erwartete Folgen der Vereinigung (zwei vergleichbare Skalen); erwartete Vorteile der Wiedervereinigung für Unternehmen oder

Arbeitnehmer; Bereitschaft zu gesellschaftlichen und politischen Veränderungen im Rahmen der Vereinigung; Präferenz für eine Souveränität der DDR und vermuteter Standpunkt der Ost- wie auch der Westbevölkerung zu dieser Frage; gewünschter Zeitpunkt für eine Vereinigung; geschätzte Zeitdauer für das Erreichen westlichen Lebensstandards in der DDR; vermuteter Einfluß der DDR-Wahlen auf den Übersiedlerstrom; Kompetenz der Parteien in Fragen der Vereinigung; Auswirkungen der Währungsunion auf die Geldwertstabilität; Fähigkeit der BRD, eine neue DDR-Wirtschaft aufzubauen; Parteipräferenz; Wahlverhalten bei der letzten Bundestagswahl 1987; Zufriedenheit mit der gewählten Partei; Religiosität; Interesse an einem Umzug in die DDR.
Grundgesamtheit und Auswahl:
Untersuchungsgebiet: BRD
Auswahl von Wahlberechtigten. Telefonische Befragung mit standardisiertem Fragebogen

Einstellung zur Vereinigung der beiden deutschen Staaten (ZA-Nr. 2135)
Erhebungszeitraum: August 1990
Datenerhebung: GFM-GETAS, Hamburg
Inhalt
Beurteilung von Parteien und Politikern sowie politische Einstellungen. Beurteilung der deutschen Vereinigung.
Themen: Zukunftszuversicht; Beurteilung der allgemeinen wie auch der eigenen derzeitigen wirtschaftlichen Lage sowie der wirtschaftlichen Lage in einem Jahr in der Bundesrepublik wie auch in der DDR; Einschätzung der Entwicklung der Arbeitslosigkeit in der Bundesrepublik und in der DDR; wichtigste Probleme und Lösungskompetenz der Parteien; Parteineigung; Beurteilung der Vereinigungsgeschwindigkeit und der Vereinigung der beiden deutschen Staaten; Kompetenz der Bundesregierung sowie der SPD in Fragen der Vereinigung; Parteipräferenz (Zweitstimme, Rangordnungsverfahren); Sicherheit der eigenen Wahlentscheidung; Beurteilung der Chancen der FDP, der Grünen, der Republikaner, der PDS bei einer gesamtdeutschen Wahl; Issue-Relevanz; Issue-Kompetenz der Parteien; vermutete Auswirkungen der Vereinigung auf die Geldwertstabilität und den Lebensstandard in Ost und West; Beurteilung des Zeitpunktes der Wirtschafts- und Währungsunion; geschätzter Zeitpunkt einer wirtschaftlichen Besserung in der DDR; Einschätzung der Fähigkeit der BRD, die Wirtschaft in der DDR aufzubauen (BRD); Bereitschaft der BRD zum Aufbau der DDR-Wirtschaft (in der DDR gefragt); Wahlverhalten bei der letzten Bundestagswahl; Befürchtung von Arbeitslosigkeit; Verbundenheit mit der Kirche; Verbundenheit mit den Gewerkschaften; Gewerkschaftsmitgliedschaft; BRD- bzw. DDR-Besuch nach der Maueröffnung.
Zusätzlich verkodet wurde: Interviewdatum und Interviewdauer; Intervieweridentifikation.
Untersuchungsgebiet: BRD und DDR
Mündliche Befragung mit standardisiertem Fragebogen in der DDR. In der BRD wurde die Befragung telefonisch durchgeführt.

Vor der Bundestagswahl 1990 (ZA-Nr. 2136)
Erhebungszeitraum: Oktober bis November 1990
Datenerhebung: Basis-Research, Frankfurt und Dresden
Inhalt
Beurteilung von Parteien und Politikern. Politische Einstellungen.
Themen: Zukunftszuversicht; wichtigste politische Probleme; Beurteilung der wirtschaftlichen Lage in der alten Bundesrepublik und in den neuen Bundesländern sowie Einschätzung der weiteren Entwicklung in einem Jahr; Beurteilung der derzeitigen und zukünftigen eigenen finanziellen Lage; Einschätzung der Entwicklung der Arbeitslosigkeit in der alten Bundesrepublik und in den neuen Bundesländern; Beurteilung der Weiterentwicklung im vereinten Deutschland; Wahlbeteiligungsabsicht; Parteipräferenz (Rangordnungsverfahren); Sicherheit der eigenen Wahlentscheidung; beste Partei im Wahlkampf; vermuteter Wahlsieger; vermutete Wahlchancen der FDP, der Grünen, der PDS und der Republikaner bei der kommenden Bundestagswahl; Issue-Relevanz; Issue-Kompetenz der Parteien; erwartete Entwicklung der DM-Stabilität nach der Vereinigung; Hoffnungen und Befürchtungen im Zusammenhang mit der Vereinigung der beiden deutschen Staaten; geschätzter Zeitpunkt der wirtschaftlichen Besserung in der ehemaligen DDR; Bereitschaft, Steuererhöhungen zugunsten der ehemaligen DDR in Kauf zu nehmen; kompetenteste Partei zur Verbesserung der wirtschaftlichen Lage und zur Lösung vereinigungsbedingter Probleme; Wahlverhalten bei der letzten Bundestagswahl bzw. Volkskammerwahl; Zufriedenheit mit der gewählten Partei; Bundeskanzlerpräferenz; Kohl oder Lafontaine als kompetentester Kanzlerkandidat; Verbundenheit mit der Kirche; Gewerkschaftsmitgliedschaft und Verbundenheit mit den Gewerkschaften.
Zusätzlich verkodet wurde: Interviewdauer.
In den neuen Bundesländern wurde zusätzlich gefragt: Berufsausbildung und Vorgesetztenfunktion; befürchtete Arbeitslosigkeit bzw. Kurzarbeit eines Haushaltsmitglieds; Erreichbarkeit zu Hause, nach Tagen und Uhrzeit differenziert.
Untersuchungsgebiet: BRD (alte und neue Bundesländer)
Mündliche Befragung mit standardisiertem Fragebogen in den neuen Bundesländern. In den alten Bundesländern wurde die Befragung telefonisch durchgeführt.

Bericht über das Frühjahrsseminar 1992

Vom 9. bis 27. März fand das Frühjahrsseminar 1992 über die Analyse zeitbezogener Daten statt. In Vorlesungen, Arbeitsgruppen und Computerübungen wurden Methoden der Panelanalyse, der multivariaten Zeitreihenanalyse und der Ereignisanalyse vermittelt. Für die Vorlesungen hatten wir Experten zum jeweiligen Thema eingeladen. Die Arbeitsgruppen und Computerübungen am Nachmittag wurden von den ZA-Mitarbeitern ***Dr. Jörg Blasius, Dr. Petra Hartmann, Harald Rohlinger*** und mir betreut.

Dozent der ersten Woche des Frühjahrsseminars zum Thema Panelanalyse war ***Dr. U. Rendtel***, der als Statistiker im Deutschen Institut für Wirtschaftsforschung in Berlin in der Projektgruppe des sozio-ökonomischen Panels (SOEP) tätig ist. Ein Schwerpunkt seiner Vorlesungen lag in der Berücksichtigung von Fehlerkorrelationen in linearen Regressionsmodellen, wie sie typischerweise in Paneldaten vorkommen. Im Dummy-Variablen- oder Fixed-Effekt-Modell wird diesem Umstand dadurch Rechnung getragen, daß für jede Untersuchungseinheit über die Spezifikation einer Dummy-Variable ein eigener Effekt geschätzt wird. Im Error-Component- oder Random-Effekt-Modell wird dagegen angenommen, daß die individuellen Effekte Realisationen einer Zufallsvariable sind, deren Varianz geschätzt wird. Neben der ausführlichen Herleitung der beiden Modelle und geeigneten Verfahren zur Parameterschätzung wurden weitere Panelmodelle vorgestellt, darunter auch nichtlineare Modelle für binäre Daten. Diskutiert wurde schließlich auch noch die Aufdeckung und die Konsequenzen von nicht zufälligen Panelausfällen.

Für die Arbeitsgruppen und Computerübungen hatte *Dr.* ***Rendtel*** SAS- und GAUSS-Programme geschrieben und zur Verfügung gestellt[1]. Mit einigen dieser Programme können auch Daten simuliert werden. Über solche simulierte Daten lassen sich die Auswirkungen von Fehlspezifikationen oder Ausfällen leicht sichtbar machen. Für Analysen mit empirischen Daten stand eine Teilmenge der Paneldaten aus der "Kölner Begleituntersuchung zur Volkszählung 1987" zur Verfügung. Von der ursprünglich geplanten Verwendung von Daten aus dem sozio-ökonomischen Panel wurde abgesehen, da die Daten nach der Auffassung der Datenschutzbeauftragten des DIW nur zu Forschungszwecken verwendet werden dürfen.

Die Vorlesungen der zweiten Woche zur multivariaten Zeitreihenanalyse hielt ***Prof. Dr. W. Stier*** von der Hochschule St. Gallen. Grundlage war ein umfangreiches Skript des Dozenten. Nach einer Einführung in die Begrifflichkeit wurden zunächst univariate ARIMA-Modelle besprochen. Darauf aufbauend wurden Transferfunktionen-Modelle und anschließend vektorielle ARIMA-Modelle vorgestellt. In Transferfunktionen-Modellen wird der Einfluß einer

1 Die Vorgehensweise bei der Schätzung eines dieser Modelle mit SPSS/PC habe ich in dieser ZA-Information auf Seite 23 ff. beschrieben.

Zeitreihe auf eine zweite untersucht, in vektoriellen ARIMA-Modellen die multivariate Regression zwischen verschiedenen Zeitreihen. In den Computerübungen wurden die Programme Mesosaur, Autobox-Plus und MTS verwendet. Im Unterschied zu anderen Programmen können Autobox-Plus und MTS univariate bzw. vektorielle ARIMA-Modelle automatisch identifizieren und schätzen. Die Arbeit mit den Programmen wurde von *Dr. Rainer Metz* vom Zentrum für historische Sozialforschung des Zentralarchivs vermittelt. Neben klassischen Zeitreihen wurden in den Übungen multivariate Jahreszeitreihen zur Höhe der Gewerkschaftsmitgliedschaft und Indikatoren wirtschaftlicher Entwicklung in Großbritannien und den USA analysiert.

In der dritten Woche wurden Modelle zur Ereignisanalyse diskutiert. Die Vorlesungen hielt *Prof. Dr. H.-P. Blossfeld* vom Europäischen Hochschulinstitut in Florenz. Vorgestellt wurde die klassische Sterbetafelanalyse, Kaplan-Meier Schätzer, das Exponentialmodell, die Berücksichtigung zeitveränderlicher Kovariaten im Exponentialmodell über Episodensplitting, parametrische Modelle mit Verweildauerabhängigkeit, das Proportional-Hazard-Modell von Cox und Modelle, in denen unbeobachtete Heterogenität berücksichtigt werden kann. Am Ende der Woche wurde auch kurz auf diskrete Hazardratenmodelle eingegangen. Empirische Anwendungen der Programme wurden mit dem Programm TDA berechnet, das von *Dr. G. Rohwer* vom Europäischen Hochschulinstitut in Florenz entwickelt wurde und den Teilnehmern zur Verfügung gestellt wurde. Neben der Einführung in die Arbeit mit TDA durch den Verfasser wurde für die Arbeitsgruppen und Übungen ein umfangreiches Skript von *H.-P. Blossfeld und G. Rohwer* bereitgestellt, in dem die Ein- und Ausgabe einer Vielzahl von Anwendungsbeispielen von TDA dokumentiert ist. Die in den Übungen verwendeten Daten sind der Lebensverlaufsstudie des Berliner Max-Planck-Instituts für Bildungsforschung entnommen.

Über den eigentlichen Stoff des Frühjahrsseminars hinaus wurden in Gastvorträgen weitere Informationen angeboten. *J. Frick* und *Dr. M. Riebschläger* berichteten in der ersten Woche über die Datenstruktur und die Analysemöglichkeiten des sozio-ökonomischen Panels (SOEP). *PD Dr. H. Thome* diskutierte in der zweiten Woche Probleme der Interpretation dynamischer Regressionsmodelle. In der dritten Woche informierte *G. Rohwer* in einer Zusatzveranstaltung über sein Programm RZoo zur effizienten Verwaltung sozialwissenschaftlicher Daten[2]. Außerdem informierten *Ekkehard Mochmann* über die GESIS-Institute IZ, ZA sowie ZUMA und *Erwin Rose* über die Dienstleistungsangebote des Zentralarchivs.

Das Interesse am Frühjahrsseminar war in diesem Jahr deutlich größer als in den Vorjahren. Dies ist sicherlich auch darauf zurückzuführen, daß für die Teilnehmer die Möglichkeit bestand, nur an einer oder zwei Wochen des dreiwöchigen Seminars teilzunehmen. Um eine

2 Eine kurze Besehreibung des Programms gibt *G.* ***Rohwer*** in dieser Ausgabe auf den Seiten 43 ff.

möglichst große Anzahl von Interessierten berücksichtigen zu können, hatten wir uns entschlossen, in diesem Jahr Arbeitsgruppen und Computerübungen zeitversetzt zu organisieren, so daß eine Hälfte der Teilnehmer im PC-Pool am Computer arbeitete während die andere Hälfte im Arbeitsgruppentreffen diskutierte. Insgesamt haben auf diese Weise 98 Teilnehmer das Frühjahrsseminar 1992 besuchen können, wobei sich 50 Teilnehmer über Panelanalyse, 38 über multivariate Zeitreihenanalyse und 41 über Ereignisanalyse informierten. 9 Teilnehmer kamen aus den neuen Bundesländern. Da wegen der Reorganisation der Universitäten und Forschungsinstitute die Situation für Sozialwissenschaftler aus den neuen Bundesländern zur Zeit besonders schwierig ist, hat die ***Thyssen-Stiftung*** dankenswerterweise die Teilnehmer aus diesem Gebiet durch Stipendien unterstützt.

Frühjahrsseminar 1993 über lineare Strukturgleichungsmodelle

Thema des nächsten Frühjahrsseminar werden lineare Strukturgleichungsmodelle mit beobachteten und latenten Variablen sein. Neben der konfirmatorischen Faktorenanalyse und der Kausalanalyse von Modellen mit metrischen Indikatoren werden auch neuere Entwicklungen zur Analyse von Modellen mit ordinalen Indikatoren behandelt. Das Frühjahrsseminar 1993 wird vom 1. bis zum 26. März 1993 stattfinden. Es wird wieder die Möglichkeit bestehen, nur an Teilen des Seminars teilzunehmen. Interessenten für das Frühjahrsseminars 1993 bitten wir schon jetzt mit dem Vordruck in diesem Heft der ZA-Information die Teilnahmebereitschaft zu bekunden. Wir werden dann bereits vor dem Erscheinen der nächsten Ausgabe der ZA-Information genauere Informationen über das Seminar und das Anmeldeformular versenden.

Steffen M. Kühnel

Zur Analyse von Paneldaten mit SPSS/PC: Die EGLS-Schätzung des Fehlerkomponentenmodells

Ein Nachtrag zum Frühjahrsseminar 1992

von Steffen M. Kühnel

Bei Paneldaten werden interessierende Merkmale von Untersuchungseinheiten im Zeitverlauf mehrfach erhoben. Bei der Analyse solcher Daten mit Regressionsmodellen stellt sich das Problem, daß als Folge der Meßwiederholungen mit korrelierten Residuen zu rechnen ist. Statistische Modelle zur Panelanalyse berücksichtigen dies. Ein Modell, das u.a. im Frühjahrsseminar 1992 zur Analyse zeitbezogener Daten vorgestellt wurde, ist das Fehlerkomponentenmodell. ***U. Rendtel*** (1992) hat in seiner Vorlesung zur Panelanalyse gezeigt, wie dieses Modell in den Programmsystemen SAS oder GAUSS über mehrfache einfache Kleinstquadratschätzungen berechnet werden kann. Ich möchte in diesem Beitrag auf die Berechnung des Modells mit SPSS/PC eingehen. Zuvor werde ich kurz auf das statistische Modell und die Idee der verwendeten Schätzmethode eingehen. Dabei werde ich auf formale Darstellungen in Matrizenschreibweise sowie auf mathematisch-statistische Ableitungen verzichten. Diese sind etwa bei ***Hübler*** (1990) und bei ***Rendtel*** (1992) zu finden.

1. Das Fehlerkomponentenmodell zur Analyse von Paneldaten

In der quantitativen Sozialforschung erfolgt die Analyse des Zusammenhangs zwischen einer abhängigen Variable Y und erklärenden Variablen $X_1, X_2, ..., X_k$ oft über ein lineares Regressionsmodell. Das Modell postuliert, daß für alle Fälle i=1, 2, ..., n die Beziehung zwischen den Werten y_i der abhängigen Variable und den Werten x_{ji} der Prädiktoren als lineare Gleichung dargestellt werden kann:

$$(1) \qquad \begin{aligned} y_i &= \alpha + x_{1i}\cdot\beta_1 + x_{2i}\cdot\beta_2 + \ldots + x_{ki}\cdot\beta_k + u_i \\ &= \alpha + \sum_{j=1}^{k} x_{ji}\cdot\beta_j + u_i \end{aligned}$$

Das statistische Modell geht in der Regel davon aus, daß die Residuen u_i in Gleichung (1) voneinander unabhängige identisch verteilte Zufallsvariablen sind, deren Erwartungswerte $\mu(u_i)$ stets Null sind. Ziel der statistischen Analyse ist es dann u.a., die Regressionskoeffizienten $\alpha, \beta_1, \beta_2, ..., \beta_k$ möglichst optimal zu schätzen.

Bei Panelanalysen liegen für jeden Fall mehrere zeitversetzte Beobachtungen vor. Formal läßt sich dies so darstellen, daß die Variablen in Gleichung (1) um Zeitindizes t ergänzt werden, die den Meßzeitpunkt t der Beobachtung identifizieren. Die Gleichung hat dann folgende Form:

(2) $$y_{it} = \alpha + \sum_{j=1}^{k} x_{jit} \cdot \beta_j + u_{it}$$

Bei der Berücksichtigung des Meßzeitpunktes ist es sinnvoll, zwischen zeitveränderlichen und zeitkonstanten erklärenden Variablen zu unterscheiden. Zeitveränderliche Variablen können zu jedem Meßzeitpunkt einen anderen Wert aufweisen, zeitkonstante Variablen variieren dagegen nicht über die Zeit. Typische Beispiele für zeitveränderliche abhängige Variablen sind in der Umfrageforschung die Antworten auf Einstellungsfragen. Das Geschlecht eines Befragten ist dagegen eine zeitkonstante Variable: unabhängig vom Meßzeitpunkt ist ein Befragter immer weiblich oder männlich. Um zwischen zeitveränderlichen und zeitkonstanten erklärenden Variablen unterscheiden zu können, wird im folgenden die Gesamtheit der erklärenden Variablen in die Teilmenge $\{X_1, X_2, ..., X_p\}$ der p zeitveränderlichen Prädiktoren und in die Teilmenge $\{Z_1, Z_2, ..., Z_k\}$ der zeitkonstanten Prädiktoren zerlegt. Die Regressionskoeffizienten der zeitveränderlichen Variablen werden durch das Zeichen "β" symbolisiert, die Regressionskoeffizieten der zeitkonstanten Variablen dagegen durch das Symbol "γ". Da bei zeitkonstanten Variablen der Zeitindex t ausgelassen werden kann, wird Gleichung (2) zu:

(3) $$y_{it} = \alpha + \sum_{j=1}^{p} x_{jit} \cdot \beta_j + \sum_{l=1}^{k} z_{li} \cdot \gamma_l + u_{it}$$

Versucht man nun, mit Hilfe der einfachen Kleinstquadratmethode die Regressionskoeffizienten aus Gleichung (3) zu schätzen, so ist mit ineffizienten Ergebnissen zu rechnen, wenn die Residuen u_{it} einer Person i nicht unabhängig voneinander sind. So ist es etwa denkbar, daß ein Befragter relativ gesehen zu den übrigen Befragten, unabhängig von den Werten der Prädiktoren, die Tendenz haben kann, bei der abhängigen Variable immer etwas höhere Werte anzugeben. Wegen der daraus folgenden Verletzung der Homoskedastizitätsannahme werden bei einer einfachen Kleinstquadratschätzung (OLS-Regression) der Regressionskoeffizienten die Standardfehler verzerrt geschätzt. Infolgedessen sind auch die inferenzstatistischen Tests ungültig.

Um diese Fehlerquelle zu beseitigen, ist es nötig, die Autokorrelationen zwischen den Meßwiederholungen zu berücksichtigen. Dies geschieht dadurch, daß der nichterklärte Rest u_{it} der Regression der abhängigen Variable in zwei Komponenten zerlegt wird:

(4) $$u_{it} = \delta_i + \varepsilon_{it}$$

In der Gleichung steht δ_i für eine Residualkomponente, die über die Zeit konstant ist, also z.B. für die Tendenz, stets einen etwas höheren Wert bei der abhängigen Variable anzugeben. Bei der zweiten Komponente ε_{it} wird dagegen angenommen, daß sie zu jedem Meßzeitpunkt einen unterschiedlichen Wert aufweisen kann. Aufgrund der Aufteilung der Residuen in zwei Komponenten wird dieses Modell in der Literatur als **Fehlerkomponentenmodell** bezeichnet.[1] Zur Abkürzung verwende ich im folgenden die Bezeichnung **EC-Modell,** wobei 'EC' für 'error component' steht.

Da es sich bei beiden Komponenten um Residuen handelt, wird unterstellt, daß jede Komponente einen Erwartungswert von Null aufweist und daß jede Komponente auch von den erklärenden Variablen des Modells unabhängig ist. Weiter wird davon ausgegangen, daß die einzelnen Fälle unabhängig sind. Aufgrund der Unabhängigkeit der zeitvariablen Komponente ε_{it} von δ_i gilt dann für die Grundgesamtheitsmittelwerte (Erwartungswerte) $\mu(u_{it})$, Varianzen $\sigma^2(u_{it})$ und Kovarianzen $\sigma(u_{it},u_{i't'})$ der ursprünglichen Residuen u_{it}:

(5)
$$\begin{aligned}
\mu(u_{it}) &= \mu(\delta_i) = \mu(\varepsilon_{it}) = 0 \\
\sigma^2(u_{it}) &= \sigma_\delta^2 + \sigma_\varepsilon^2 \\
\sigma(u_{it},u_{it'}) &= \sigma_\delta^2 \qquad t \neq t' \\
\sigma(u_{it},u_{i't'}) &= 0 \qquad i \neq i'
\end{aligned}$$

Wenn die Varianzen σ_δ^2 und σ_ε^2 der Fehlerkomponenten bekannt wären, ließen sich die Regressionskoeffizienten mit Hilfe des verallgemeinerten Kleinstquadratschätzers (GLS-Schätzer) unverzerrt und effizient schätzen. Tatsächlich kann die Schätzung in zwei Schritten erfolgen. Im ersten Schritt werden die Varianzen der beiden Fehlerkomponenten geschätzt. In einem zweiten Schritt werden diese Schätzwerte dann in der verallgemeinerten Kleinstquadratschätzung der Regressionskoeffizienten eingesetzt. Diese Schätzmethode wird als **EGLS-Schätzung** bezeichnet, wobei der Ausdruck "EGLS" für "estimated gene-

[1] Es gibt auch komplexere Modelle, in denen zusätzlich für die Meßzeitpunkte eine Fehlerkomponente modelliert wird (vgl. *Hübler*, 1990: 70).

ralized least squares" steht. Der Vorteil dieser Schätzmethode besteht darin, daß sie durch mehrfache Anwendung gewöhnlicher Regressionsrechnungen bewerkstelligt werden kann.

2. EGLS-Schätzung über mehrfache einfache Regressionen

Im ersten Schritt der EGLS-Schätzung müssen die Varianzen der Fehlerkomponenten geschätzt werden. Dies erfolgt über zwei einfache Regressionen. Zunächst wird für jeden Fall i der Stichprobe der Mittelwert der Variablen über die Zeit gebildet. Aus dem EC-Modell folgt dann für die Mittelwerte:

(6)
$$\bar{y}_i = \frac{1}{T_i}\sum_{t=1}^{T_i} y_{it} = \alpha + \sum_{j=1}^{p} \bar{x}_{ji}\cdot\beta_j + \sum_{l=1}^{K} z_{li}\cdot\gamma_l + \bar{u}_i$$
$$\text{mit:} \quad \bar{u}_i = \delta_i + \bar{\varepsilon}_i$$

In Gleichung (6) steht T_i für die Anzahl der Beobachtungen eines Falles i über die Zeit. Es ist nicht notwendig, daß alle Fälle gleich oft beobachtet werden. Zieht man die Mittelwerte über die Zeit von den ursprünglichen Werten ab, kürzen sich die zeitkonstanten Variablen und Koeffizienten heraus. Das Modell reduziert sich zum sogenannten **Within-Modell:**

(7)
$$(y_{it} - \bar{y}_i) = \sum_{j=1}^{p} (x_{jit} - \bar{x}_{ji})\cdot\beta_j + (\varepsilon_{it} - \bar{\varepsilon}_i)$$
$$\tilde{y}_i = \sum_{j=1}^{p} \tilde{x}_{ji}\cdot\beta_j + \tilde{\varepsilon}_i$$

Zu beachten ist, daß Gleichung (7) keine Regressionskonstante enthält. Es läßt sich nun zeigen, daß die Regression von $\tilde{y}$ auf die Prädiktoren $\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_p$ nach der gewöhnlichen Kleinstquadratmethode (ohne Interzept) zu erwartungstreuen und konsistenten Schätzern $\beta_{(w)1}, \beta_{(w)2}, \ldots, \beta_{(w)p}$ der Regressionskoeffizienten $\beta_1, \beta_2, \ldots, \beta_p$ führt. Darüber hinaus ergibt sich als erwartungstreuer Schätzer für die Varianz σ_ε^2:[2]

[2] Der Abzug der zeitbezogenen Mittelwerte von den Ausgangsdaten läßt sich in Matrizenschreibweise als eine Multiplikation von links mit einer blockdiagonalen Matrix darstellen, wobei jeder Block eine idempotente-symmetrische Submatrix ist. Obwohl die Residuen der Regression der transformierten Daten nicht homoskedastisch sind, ist der Erwartungswert der Summe der quadrierten Residuen der Kleinstquadratregression hier proportional zur Varianz σ_ε^2.

(8)
$$\hat{\sigma}_\varepsilon^2 = \frac{\sum_{i=1}^{n} \left(\bar{y}_i - \sum_j \bar{x}_{ji} \cdot \beta_{(W)j}\right)^2}{\sum_{i=1}^{n} T_i - n - p}$$

Über das Within-Modell kann also die Varianz einer der beiden Fehlerkomponenten geschätzt werden. Die Vermutung liegt nahe, daß die Varianz der anderen Fehlerkomponente über die Regression der Mittelwerte aus Gleichung (6) ermittelt werden kann. Dieses Regressionsmodell über die Mittelwerte wird als **Between-Modell** bezeichnet. Betrachten wir dazu die Residuen dieser Gleichung, so gilt:

(9)
$$\sigma^2(\bar{u}_i) = \sigma_\delta^2 + \frac{1}{T_i} \cdot \sigma_\varepsilon^2$$

Bei über die Fälle variierenden T_i sind die Residuen heteroskedastisch: Je nach der Anzahl der Beobachtungen eines Falles über die Zeit ist die Varianz $\sigma^2(\bar{u}_i)$ unterschiedlich groß. Faßt man dagegen Fälle mit gleicher Anzahl von Beobachtungen zusammen, so haben in dieser Teilgruppe die Residuen der Regression über die Mittelwerte die gleiche Varianz. Schätzt man für eine solche Teilgruppe, bei denen jeweils T Beobachtungen vorliegen, die Between-Regressionskoeffizienten $\hat{\alpha}_{(B)}$, $\hat{\beta}_{(B)1}$, $\hat{\beta}_{(B)2}$, ..., $\hat{\beta}_{(B)p}$, $\hat{\gamma}_{(B)1}$, $\hat{\gamma}_{(B)2}$, ..., $\hat{\gamma}_{(B)k}$ aus Gleichung (6) mit der üblichen Kleinstquadratmethode, gilt[3] für die beobachteten Residuen dieser Regression:

(10)
$$\frac{\mu\left(\sum_{i=1}^{n}\left(\bar{y}_i - \hat{\alpha}_{(B)} - \sum_j \bar{x}_{ji} \cdot \hat{\beta}_{(B)j} - \sum_l z_{li} \cdot \hat{\gamma}_{(B)l}\right)^2\right)}{n-p-k-1} = \frac{\mu\left(\sum_{i=1}^{n} \bar{e}^2\right)}{n-p-k-1} = \sigma_\delta^2 + \frac{1}{T}\sigma_\varepsilon^2$$

Ersetzt man in Gleichung (10) den Erwartungswert der Summe der beobachteten quadrierten Residuen durch die Summe selber und setzt für σ_ε^2 den Schätzwert aus Gleichung (8) ein, ergibt sich ein konsistenter Schätzer für die zweite Fehlervarianzkomponente:

---

[3] In Gleichung (10) und (11) ist zu beachten, daß sich die Fallzahl n hier nur auf die Fälle bezieht, für die genau T Beobachtungen vorliegen.

(11)
$$\hat{\sigma}_\delta^2 = \frac{\sum_{i=1}^{n} \bar{e}^2}{n-p-k-1} - \frac{1}{T}\hat{\sigma}_\varepsilon^2$$

Damit ist der erste Schritt der EGLS-Schätzung abgeschlossen. Im zweiten Schritt werden die geschätzten Fehlervarianzkomponenten für die verallgemeinerte Kleinstquadratschätzung genutzt. Anstelle einer direkten verallgemeinerten Kleinstquadrat-Schätzung werden die Ursprungsdaten so transformiert, daß die Residuen homoskedastisch werden. Die übliche Kleinstquadratschätzung über die so transformierten Daten ergibt die gleichen Ergebnisse wie eine direkte verallgemeinerte Kleinstquadratschätzung.

Für die Transformation sind zunächst die Werte ϑ_i einer Skalierungsvariable zu berechnen:

(12)
$$\vartheta_i = 1 - \sqrt{\frac{\hat{\sigma}_\varepsilon^2}{\hat{\sigma}_\varepsilon^2 + T_i \cdot \hat{\sigma}_\delta^2}}$$

Anschließend wird in den Ausgangsdaten von allen Variablen, einschließlich der Konstante "Eins" zur Schätzung des Interzepts, der jeweilige mit ϑ_i multiplizierte Mittelwert über die Zeit abgezogen:

(13)
$$\begin{aligned} q_{it} &= y_{it} - \vartheta_i \cdot \bar{y}_i \\ c_i &= 1 - \vartheta_i \\ v_{jit} &= x_{jit} - \vartheta_i \cdot \bar{x}_{ji} \qquad j = 1,2,\ldots,p \\ w_{li} &= z_{li} - \vartheta_i \cdot z_{li} \qquad l = 1,2,\ldots,k \end{aligned}$$

Die Berechnung der gewöhnlichen Kleinstquadratlösung (ohne Interzept) der Regressionsgleichung:

(14)
$$q_{it} = c_i \cdot \hat{\alpha} + \sum_{j=1}^{p} v_{jit} \cdot \hat{\beta}_j + \sum_{l=1}^{p} w_{li} \cdot \hat{\gamma}_l + e_{it}$$

ergibt schließlich die gesuchten EGLS-Schätzer der ursprünglichen Regressionsgleichung (3). Asymptotisch korrekt sind auch die dabei mitberechneten Standardfehler und T-Werte der Koeffizienten.

3. Die Berechnung der EGLS-Lösung mit SPSS/PC

Die Darstellung der EGLS-Schätzmethode im vorherigen Abschnitt zeigt, daß alle Berechnungen mit einfachen Kleinstquadratschätzungen multipler Regressionsgleichungen durchgeführt werden können. Das EC-Modell kann somit mit beliebigen Statistikprogrammen geschätzt werden. Im folgenden möchte ich darstellen, wie die Berechnung in SPSS/PC erfolgt.

Betrachtet man die Ausgangsgleichung (3) des EC-Modells, so fällt auf, daß Messungen einer Größe zu verschiedenen Zeitpunkten als *eine* doppelt indizierte Variable aufgefaßt werden. Oft sind die Paneldaten in einer Datei allerdings so angeordnet, daß die mehrfachen Messungen einer Größe zu verschiedenen Zeitpunkten als verschiedene Variablen abgespeichert sind. Wenn dies der Fall ist, müssen die Variablen zunächst rearrangiert werden. Dazu werden die relevanten Variablen jeder Panelwelle in temporäre Systemdateien gespeichert und anschließend mit dem SPSS/PC-Kommando "JOIN ADD" so zusammengefügt, daß die Meßwiederholungen unter einem Variablennamen gespeichert werden. Zeitkonstante Variablen werden dabei mehrfach herausgeschrieben. Zur späteren Identifikation werden außerdem die Fallnummern und eine Variable, die den Meßzeitpunkt enthält, in die temporären Dateien herausgeschrieben. Außerdem empfiehlt es sich, auf diese Weise gleich ungültige Fälle auszuschließen.

Wenn "idnr" der Variablenname zur Identifikation der Fallnummer, "yl" die abhängige Variable zum ersten Meßzeitpunkt sei, "xll" und "x21" zwei zeitveränderliche erklärende Variablen zu diesem Meßzeitpunkt und "z" eine zeitkostante erklärende Variable, würden die SPSS-Anweisungen zum Herausschreiben der Daten des ersten Meßzeitpunkts folgendermaßen aussehen:

```
COUNT nmissing=yl xll x21 z (MISSING).
COMPUTE zeit=l.
PROCESS IF (nmissing EQ 0).
SAVE OUT="templ.sys"
    /KEEP=idnr zeit yl xll x21 z /RENAME (yl xll x21 z=y xl x2 z).
```

Analog würden auch die Daten der übrigen Meßzeitpunkte in temporäre Dateien herausgeschrieben. Wichtig ist, daß die herausgeschriebenen Daten, die sich auf eine Größe beziehen, den selben Variablennamen erhalten. Dies wird durch die Spezifikation "RENAME .." im SAVE-Kommando erreicht. Die temporären Dateien werden dann mit dem Kommando "JOIN ADD" so zu einer Datei zusammengesetzt, daß alle Meßwerte einer Größe unabhängig vom Meßzeitpunkt in einer Variable stehen:

```
JOIN ADD /FILE="templ.sys" /FILE="ternp2.sys" /FILE="temp3.sys".
SORT CASES BY idnr zeit.
```

Im Beispiel werden die Daten eines dreiwelligen Panels zusammengefügt. Das Kommando "SORT CASES" sortiert anschließend die Daten nach Fallnummer und Meßzeitpunkt.

Für die Berechnung der Between- und Within-Regressionen müssen die Mittelwerte über die Zeit gebildet und in einer Datei abgespeichert werden. Die Mittelwertsbildung erfolgt mit dem Kommando "AGGREGATE":

```
AGGREGATE OUT=*
          /PRES  /BREAK=idnr
          /yq=MEAN(y)
          /xlq=MEAN(xl)
          /x2q=MEAN(x2)
          /z=MEAN(z)
          /ti=NU(y).
SAVE OUT="tempm.sys"  /COMP.
```

Das Kommando "AGGREGATE" faßt über den Unterbefehl "BREAK=idnr" die Werte eines Falles zusammen. Die Berechnung der Mittelwerte der abhängigen Variable über die Meßzeitpunkte wird durch den Unterbefehl "yq=MEAN(y)" erreicht. Analog werden die Mittelwerte der anderen Variablen berechnet. Für die Within-Regression sind die Ausgangsdaten um diese Mittelwerte zu bereinigen. Dazu müssen die Mittelwerte zunächst fallweise mit den Ausgangsdaten verknüpft werden. Dies geschieht mit dem "JOIN MATCH"-Kommando:

```
JOIN MATCH  TABLE="tempm.sys"
            /FILE="templ.sys"
            /BY idnr.
COMPUTE yw=y-yq.
COMPUTE xlw=xl-xlq.
COMPUTE x2w=x2-x2q.
```

Die Mittelwerte (Aggregatdaten) sind in der Datei "tempm.sys" gespeichert, die Ausgangsdaten (Individualdaten) in der Datei "templ.sys". Durch die Spezifikation "TABLE=..." wird dem SPSS-Programm mitgeteilt, daß eine hierarchische Verknüpfung der Daten erfolgt, so daß ein Fall in der Aggregatdatendatei mehreren Fällen in der Individualdatendatei zugeordnet werden kann. Die Verknüpfung erfolgt über die Fallnummer, was dem Programm durch den Unterbefehl "BY idnr" mitgeteilt wird. Nachdem die Mittelwerte mit den Ursprungsdaten verknüpft sind, werden in den nachfolgenden COMPUTE-Anweisungen die Mittelwerte von den Ausgangsdaten abgezogen. Die OLS-Regression über diese Daten ergibt die WITHIN-Schätzung:

```
REGRESSION VAR yw xlw x2w
           /ORIGIN /DEP=yw /ENT
           /SAVE=RESID(wresid).
COMPUTE wres2=wresid*wresid.
```

Bei der Regression ist zu beachten, daß keine Regressionskonstante geschätzt wird. In SPSS wird dies über die Spezifikation "ORIGIN" sichergestellt. Mit dem Unterbefehl "SAVE=RESID(wresid)" werden die Residuen der Within-Regression als zusätzliche Variable gespeichert. Die nachfolgende COMPUTE-Anweisung quadriert diese Residuen. Die Summe der Variable "wres2" über alle Beobachtungen ist der Zähler auf der rechten Seite von Gleichung (8). Die Division durch die Gesamtzahl der Beobachtungen minus der Fallzahl und der Anzahl der zeitvariablen Prädiktoren ergibt die Varianz σ_ε^2.

Die Between-Regression erfolgt über die aggregierten Mittelwerte. Durch das Kommando "SELECT IF (ti=t)" vor der Regression wird sichergestellt, daß nur Fälle mit gleicher Beobachtungszahl t in die Rechnung eingehen:[4]

```
SELECT IF (ti = t) .
REGRESSION VAR yq xlq x2q zq
           /DEP yq
           /ENT
           /SAVE=RESID(bresid).
COMPUTE bres2=bresid*bresid.
```

Wie bei der Within-Regression werden auch hier wieder die Residuen gespeichert und in der nachfolgenden COMPUTE-Anweisung quadriert. Teilt man die Summe der Werte dieser Variable durch die in der Regression eingehenden Fälle mit Beobachtungszahl t minus der Anzahl der in der Between-Regression geschätzten Regressionskoeffizienten und zieht davon die durch t geteilte Varianz σ_ε^2 ab, ergibt sich die Schätzung für die zweite Varianzkomponente σ_δ^2. Der Wert dieser Varianzen sei in den Variablen "sigma2e" und "sigma2d" gespeichert. Mit einer COMPUTE-Anweisung läßt sich dann die Skalierungsvariable ϑ nach Gleichung (12) berechnen. Anschließend erfolgt die Transformation der Ausgangsdaten entsprechend Gleichung (13) und schließlich die Regression zwischen den transformierten Variablen:

```
COMPUTE theta=l-SQRT(sigma2e/(sigma2e+sigma2d*ti)).
COMPUTE constt=l-theta.
COMPUTE yt=y-theta*yq.
COMPUTE xlt=xl-theta*xlq.
COMPUTE x2t=x2-theta*x2q.
COMPUTE zt=z-theta*z.
REGRESSION VAR=yt constt xlt x2t zt
           /ORIGIN /DEP=yt /ENT.
```

Das Ergebnis dieser letzten Regression ist die gesuchte EGLS-Lösung.

---

[4] Die Größe "t" muß natürlich vorher festgelegt und mit einer COMPUTE-Anweisung generiert worden sein.

4. Ein Anwendungsbeispiel

Im Anwendungsbeispiel soll die Haltung zur Volkszählung 1987 durch die Beurteilung der Notwendigkeit der Zählung, das politische Interesse und das Geschlecht des Befragten erklärt werden. Die Daten stammen aus dem dreiwelligen Panel der Kölner Begleituntersuchung zur Volkszählung.[5] Die Haltung zur Volkszählung wurde auf einer siebenstufigen Antwortskala erhoben, das politische Interesse auf einer fünfstufigen Skala. Die Notwendigkeit der Volkszählung ist dichotomisiert. Die Variablennamen sind der Notation dieses Beitrages entsprechend angepaßt. So bezeichnet "Yl" die Haltung zur Volkszählung in der ersten Welle, "Y2" die Haltung in der zweiten Welle und "Y3" die Haltung in der dritten Welle. "XU" enthält die Antworten für das politische Interesse zum ersten Meßzeitpunkt, "X12" und "X13" die Messungen in den Folgewellen. "X21", "X22" und "X23" stehen für die Beurteilung der Notwendigkeit der Zählung zum ersten, zweiten bzw. dritten Meßzeitpunkt. Das Geschlecht wird als zeitkonstante Variable "Z" genannt. Neben den Werten dieser Variablen enthält die SPSS-Systemdatei "ECBEISP.SYS" noch die Fallnummer (Variablenname: "IDNR").

Die nachfolgenden Tabellen zeigen, daß aus der ersten Welle 2514 Beobachtungen der zu analysierenden Variablen vorliegen, aus der zweiten Welle 1380 Beobachtungen und aus der dritten Welle 1254 Beobachtungen. Der Tabelle mit den Häufigkeit der Meßwiederholungen ist zu entnehmen, daß 1056 Personen nur einmalig befragt wurden, 312 Personen zweimal befragt wurden und 1167 Personen in allen drei Wellen befragt wurden.[6]

Beobachtungen pro Erhebungszeitpunkt	
1. Welle	2514
2. Welle	1380
3. Welle	1254
insgesamt:	5148

Anzahl Meßwiederholungen	
keine	1056
eine	312
zwei	1167

Die Berechnung der EGLS-Lösung erfolgt mit dem im Anhang dokumentierten SPSS-Befehlen. Im ersten Schritt werden die Daten so rearrangiert, daß die Messungen einer

[5] Der Datensatz ist unter der Studiennummer ZA1592 im Zentralarchiv archiviert. Zur Beschreibung der Studie vgl. *Scheuch u.a.,* 1989.

[6] Von den 312 Personen, die einmal wiederbefragt wurden, sind 225 Personen in der 1. und 2. Welle befragt worden und 87 Personen in der 1. und 3. Welle.

Größe über die Meßzeitpunkte hinweg in einer Variable stehen. Anschließend werden in den Variablen "P" und "K" die Anzahl der zeitveränderlichen und zeitkonstanten Prädiktoren festgehalten. Der Wert 3 in der Variable "TMAX" legt fest, daß später bei der Between-Regression nur Fälle berücksichtigt werden, bei denen drei Beobachtungen vorliegen. Ab dieser Stelle erfolgt die gesamte Berechnung automatisch. Die im Laufe der Berechnung notwenigen Hilfsgrößen zur Berechnung der Varianzkomponenten werden während der Schätzung ermittelt und zwischengespeichert.

```
Page  27   EGLS-Schätzung des 'Error-Component-Models' (EC-Modell)      5/7/92

Berechnung 'between'-Regression' über vollständige Fälle

        * * * *  M U L T I P L E  R E G R E S S I O N  * * * *

Equation Number 1    Dependent Variable..   YQ

Multiple R            .69961
R Square              .48945
Adjusted R Square     .48814
Standard Error       1.13089

Analysis of Variance
                 DF        Sum of Squares        Mean Square
Regression        3          1425.92970          475.30990
Residual       1163          1487.37736            1.27891

F =     371.65109       Signif F =  .0000

        * * * *  M U L T I P L E  R E G R E S S I O N  * * * *

Equation Number 1    Dependent Variable..   YQ

------------------ Variables in the Equation ------------------

Variable             B         SE B       Beta          T  Sig T

ZQ             .169773      .070895    .053699      2.395  .0168
X2Q           2.737096      .083530    .689449     32.768  .0000
X1Q            .135327      .041336    .073687      3.274  .0011
(Constant)    2.902456      .142225                20.407  .0000

Equation Number 1    Dependent Variable..   YQ
From Equation   1:   1 new variables have been created.

  Name        Contents
  ----        --------
  BRESID       Residual
```

Ergebnisse der Between-Regression

Nach der Neuanordnung und anschließender Aggregierung wird zunächst die Between-Regression über die Mittelwerte der Fälle gerechnet, für die alle drei Beobachtungen vorliegen. Die Ausgabe von SPSS/PC ist als Abbildung wiedergegeben. Die eigentlichen Schätzergebnisse sind für unsere Zwecke uninteressant. Von Bedeutung ist nur, daß die Residuen in der Variable "BRESID" gespeichert werden. Diese Residuen werden quadriert

und deren Summe (Variable: SSEB) zusammen mit der zugrundeliegenden Fallzahl (Variable: NBETW) in der Datei "tempb.sys" festgehalten.

Anschließend werden die Mittelwerte mit den Ursprungsdaten zusammengeführt. Nach dem Abziehen der Mittelwerte von den Ausgangsdaten wird die Within-Regression über die mittelwertsbereinigten Daten berechnet. Auch für diese Regression ist die Ausgabe von SPSS/PC wiedergegeben.

```
Page  35   EGLS-Schätzung des 'Error-Component-Models' (EC-Modell)      5/7/92

within-Regression (ohne Konstante!)

        * * * *  MULTIPLE REGRESSION THROUGH THE ORIGIN  * * * *

Equation Number 1    Dependent Variable..   YW

Block Number  1.  Method:  Enter
Variable(s) Entered on Step Number
   1..     X2W
   2..     X1W

Analysis of Variance
                    DF       Sum of Squares      Mean Square
Regression           2           354.93665        177.46832
Residual          5146          4261.89669           .82820

F =     214.28300       Signif F =  .0000

------------------ Variables in the Equation ------------------

Variable             B         SE B       Beta          T  Sig T

X2W              .986420     .047717    .277419    20.672  .0000
X1W              .005376     .028149    .002563      .191  .8485


From Equation   1:   1 new variables have been created.

  Name       Contents
  ----       --------
  WRESID      Residual
```

Ergebnisse der Within-Regression

Erklärende Variablen der Within-Regression sind allein die zeitvariablen Prädiktoren, im Beispiel also das politische Interesse und die Beurteilung der Notwendigkeit der Zählung. Da keine Regressionskonstante geschätzt wird, ist der von SPSS ausgegebenen Determinationskoeffizient nicht aussagekräftig (und in der Abbildung deswegen nicht wiedergegeben). Das Programm weist darauf auch durch einen kurzen Text hin.

Die geschätzten Regressionskoeffizienten sind dagegen brauchbare Schätzungen des Effektes der zeitvariablen Prädiktoren. Die ausgegebenen Standardfehler, T- und F-Test sind allerdings vor einer Interpretation zu korrigieren. Dies liegt daran, daß SPSS bei der Schätzung der Residualvarianz der Within-Regression die Summe der quadrierten Residu-

en durch die Anzahl der Beobachtungen minus der Anzahl der geschätzten Koeffizienten teilt. Korrekt wäre hier dagegen die Berechnung nach Gleichung (8). Korrigiert man die Werte der Standardschätzfehler um einen Korrekturfaktor q:

(15)
$$q = \sqrt{\frac{\sum_i T_i - p}{\sum_i T_i - n - p}}$$

erhält man korrekte Ergebnisse. Im Beispiel würde sich q nach

$$q = \sqrt{\frac{5148-2}{5148-2514-2}} = \sqrt{1.955} = 1.398$$

berechnen, da insgesamt 5148 Beobachtungen von 2514 verschiedenen Fällen vorliegen und zwei Regressionskoeffizienten in der Within-Regression geschätzt werden. Durch Multiplikation des Korrekturfaktors mit dem von SPSS ausgegebenen Wert erhält man die korrekten Standardfehler, im Beispiel den Wert 0.067 für den Standardfehler des Effekts der Notwendigkeit der Volkszählung (Variable: X2W) und 0.039 für den Standardfehler des Effekts des politischen Interesses (Variable: X1W). Die korrekten T-Werte erhält man, wenn die ausgegebenen Werte durch die Korrekturgröße geteilt werden. Im Beispiel sind die korrekten Werte 14.8 für die Notwendigkeit der Zählung und 0.14 für das politische Interesse. Auch der F-Test des Gesamteinflusses der Prädiktoren läßt sich korrigieren. Der von SPSS ausgegebene Wert (hier: 214.28) muß durch das Quadrat von q geteilt werden. Für die Beispieldaten erhält man einen Wert von 109.6. Die Freiheitsgrade sind 2 und 2632 (=5148-2514-2).

Für die eigentliche EGLS-Schätzung ist allein die Summe der quadrierten Residuen der Within-Regression von Bedeutung. Zur Berechnung werden die Residuen unter dem Variablennamen "WRESID" abgespeichert, quadriert und aufsummiert (Variable: SSEW). Das Ergebnis wird zusammen mit der Gesamtbeobachtungszahl (hier 5148) in der Datei "tempw.sys" festgehalten. Für die Berechnung der Varianzen der Fehlerkomponenten wird diese Datei mit den Dateien "tempb.sys", die die Ergebnisse der Between-Regression enthält, und "temp2.sys", die die Fallzahl enthält, zusammengefügt. Anschließend werden nach Gleichung (9) und (11) die Fehlervarianzen berechnet und ausgegeben.

```
Page  47   EGLS-Schätzung des 'Error-Component-Models' (EC-Modell)      5/7/92
***** Varianzen der Fehlerkomponenten:
  NFALL    NGES   NBETW      SIGMA2E      SIGMA2D   NEGVAR
   2514    5148    1167      1.61926       .73916        .
```

Für die zeitkonstante Komponente σ_δ^2 ergibt sich bei den Daten ein Schätzwert von 0.739, für die zeitveränderliche Komponente σ_ε^2 ein Wert von 1.619. Ausgegeben werden zusätzlich die Gesamtbeobachtungszahl (NGES $\hat{=} \Sigma T_i$), die Fallzahl (NFALL $\hat{=} n$), die Anzahl der Fälle der Between-Regression (NBETW) und eine Indikatorvariable NEGVAR, die den Wert 1 annehmen würde, wenn für σ_δ^2 ein negativer Wert geschätzt worden wäre.[7]

Die eigentliche Begründung für das EC-Modell liegt in der Gefahr der Autokorrelationen der Residuen aufgrund der Meßwiederholungen über die gleichen Untersuchungseinheiten. Aus den beiden Varianzkomponenten läßt sich diese Autokorrelation schätzen:

(16)
$$\hat{\rho} = \frac{\hat{\sigma}_\delta^2}{\hat{\sigma}_\delta^2 + \hat{\sigma}_\varepsilon^2}$$

Für das Beispiel ergibt sich ein Wert von 0.313. Da dieser Wert deutlich von Null abweicht, kann die Autokorrelation nicht einfach ignoriert werden. Die Schätzung der Regressionskoeffizienten nach Gleichung (3) über eine einfache OLS-Regression würde tatsächlich zu verzerrten Standardfehlern und Testergebnissen führen.

Nachdem die Varianzen der Fehlerkomponenten bestimmt sind, wird die eigentliche EGLS-Lösung berechnet. Dazu wird zunächst die Korrekturvariable $\hat{\vartheta}$ berechnet, dann die Ausgangsdaten (einschließlich der Konstante des Regressionsmodells) transformiert und schließlich die Kleinstquadratlösung der transformierten Daten berechnet und ausgegeben. Die vom Programm berechneten Regressionskoeffizienten, Standardfehler, T- und F-Werte sind interpretierbar. Der Ausschnitt aus der SPSS/PC-Ausgabe zeigt, daß die Beurteilung der Notwendigkeit einen deutlichen Einfluß auf die Haltung zur Volkszählung hat.

---

[7] Falls die Differenz aus Gleichung (11) negativ wird, wird der Schätzwert auf Null umkodiert, um das Programm fehlerfrei weiterrechnen zu lassen.

```
Page  51   EGLS-Schätzung des 'Error-Component-Models' (EC-Modell)      5/7/92

EGLS-Regress. über OLS der transform. Daten

         * * * *  MULTIPLE REGRESSION THROUGH THE ORIGIN  * * * *

Equation Number 1    Dependent Variable..   YT

Variable(s) Entered on Step Number
   1..    ZT
   2..    X2T
   3..    X1T
   4..    CONSTT

Analysis of Variance
                    DF       Sum of Squares       Mean Square
Regression           4          52425.39716       13106.34929
Residual          5144           9353.02203           1.81824

F =    7208.26494       Signif F =  .0000

------------------ Variables in the Equation ------------------

Variable              B        SE B       Beta          T  Sig T

ZT              .070629     .055949    .009795      1.262  .2069
X2T            1.990164     .046804    .295345     42.521  .0000
X1T             .088032     .024252    .056697      3.630  .0003
CONSTT         3.212612     .090192    .647564     35.620  .0000
```

Ergebnisse der EGLS-Schätzung

Befragte, die die Zählung für notwendig halten, weisen im Durchschnitt eine um 1.99 Einheiten positivere Haltung zur Volkszählung auf als Personen, die die Zählung nicht für notwendig halten. Der Effekt des politischen Interesses ist dagegen recht gering. Da ein steigender Wert der Variable für geringeres politisches Interesse steht, bedeutet das Ergebnis, daß sinkendes politisches Interesse mit einer geringfügig positiveren Bewertung der Volkszählung einhergeht. Das Geschlecht des Befragten hat schließlich keinen signifikanten Effekt auf die Haltung zur Zählung.

Von Interesse ist oft der Gesamteffekt der erklärenden Variablen. Berechnet man den Determinationskoeffizienten nach der Standardformel:

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y}_i)^2} = 1 - \frac{\sum_i e_i^2}{\sum_i (y_i - \bar{y})^2}$$

stellt sich die Frage, welche Residuen e; in die Formel eingehen sollen. Es liegt nahe, die Schätzungen von u_{it} aus Gleichung (4) zu verwenden. Diese liegen jedoch zunächst nicht vor, da die EGLS-Lösung auf der Basis der nach Gleichung (13) transformierten Daten berechnet wurde. Sie lassen sich jedoch berechnen, indem man von den Ausgangswerten der abhängigen Variable die Vorhersagewerte abzieht, die sich aufgrund der mit EGLS geschätzten Regressionskoeffizienten ergeben:

(17) $$\hat{u}_{it} = y_{it} - \hat{\alpha} - \sum_{j=1}^{p} x_{jit} \cdot \hat{\beta}_j - \sum_{l=1}^{k} z_{li} \cdot \hat{\gamma}_l$$

Etwas einfacher ist die Berechnung, wenn man die Summe der geschätzten Varianzen der beiden Fehlerkomponenten geteilt durch die geschätzte Varianz der abhängigen Variable von 1 abzieht:

(18) $$R^2_{adj} = 1 - \frac{\hat{\sigma}^2_\delta + \hat{\sigma}^2_\varepsilon}{\hat{\sigma}^2_y}$$

Da bei der Berechnung nach Gleichung 18 die Freiheitsgrade angepaßt sind, handelt es sich um einen adjustierten Determinationskoeffizient. Für das Beispiel berechnet sich der Wert nach:

$$R^2_{adj} = 1 - \frac{1.62 + 0.74}{3.94} = 0.40$$

Die drei erklärenden Variablen decken demnach ungefähr 40% der Varianz der abhängigen Variable auf.

5. Diskussion

Das Beispiel zeigt, daß die EGLS-Schätzung der Koeffizienten des EC-Modells auch mit SPSS/PC möglich ist. Für eine halbwegs automatische Berechnung ist allerdings eine Vielzahl von Anweisungen möglich. Einmal geschrieben, können sie aber jederzeit als Gerüst für weitere Anwendungen verwendet werden. Sollen etwa zusätzliche Variablen verwendet werden, ist es nur nötig, bei dem im Anhang dokumentierten Programm in den SAVE-, AGGREGATE-, JOIN ADD-, JOIN MATCH- und REGRESSION-Anweisungen zusätzliche Variablen anzuhängen und die Konstanten "p" bzw. "k" entsprechend hochzusetzen. Bei Bedarf kann auch die Konstante "tmax" angepaßt werden, wenn bei der Between-Regression Fälle mit einer anderen Beobachtungshäufigkeit berücksichtigt werden sollen.

Möglich ist ferner die Berücksichtigung von Dummy-Variablen für den jeweiligen Meßzeitpunkt. Die Spezifikation solcher Dummy's bedeutet inhaltlich, daß davon ausgegangen wird, daß das Durchschnittsniveau der abhängigen Variablen unabhängig von möglichen Änderungen bei den zeitveränderlichen Prädiktoren zu jedem Meßzeitpunkt unterschiedlich sein kann. So haben andere Analysen der Daten der Kölner Begleituntersuchung gezeigt,

daß sich zwischen der ersten und der zweiten Panelwelle ein deutlicher Wandel zugunsten einer positiveren Einstellung zur Volkszählung ereignete (vgl. *Scheuch u.a.,* 1989). Diese Veränderung läßt sich dadurch spezifizieren, daß eine Dummy-Variable als weiterer Prädiktor ins Modell aufgenommen wird, die in der ersten Panelwelle den Wert Null und in den Folgewellen den Wert Eins aufweist. Wenn die Einstellungsänderung nicht bereits durch eine Änderung der Beurteilung der Notwendigkeit der Zählung erklärt werden kann, wird diese Dummy-Variable einen signifikanten positiven Effekt aufweisen.

Bei der Spezifikation solcher Dummy-Variablen ist eine gewisse Vorsicht angebracht, da die Dummy-Variablen weder reine zeitvariable noch zeitkonstante Variablen sind. Bei der Berechnung der Mittelwerte über die Zeit ergibt sich nämlich bei allen Untersuchungseinheiten mit gleichem Beobachtungsmuster der gleiche Wert. Bei der Between-Regression lassen sich daher diese Effekte nicht schätzen. Auch bei der Within-Regression kann es Schwierigkeiten, geben, wenn für alle Untersuchungseinheiten die gleiche Anzahl von Meßzeitpunkten vorliegt und mehr als eine Dummy-Variable spezifiziert wird. Am einfachsten ist es, wenn die Dummy-Variablen bei der Schätzung der Varianzkomponenten nicht berücksichtigt werden und erst im letzten Schritt in die Regressionsgleichung aufgenommen werden.[8]

So einfach Analysen mit dem EC-Modell sind, so sei doch auch auf Probleme hingewiesen. Sichtbar werden diese, wenn man die Koeffizientenschätzung der Within-Regression mit der EGLS-Lösung vergleicht. Theoretisch handelt es sich jeweils um konsistente Schätzungen derselben Koeffizienten. Trotzdem weisen die Schätzwerte erhebliche Unterschiede auf.[9] Woran mag dies liegen? Der wichtigste Unterschied zwischen den beiden Schätzverfahren besteht darin, daß die Effekte der Within-Regression allein durch die Veränderungen bei den Werten der zeitveränderlichen Variablen zu verschiedenen Meßzeitpunkten induziert werden. Insbesondere bei dem politischen Interesse ist es aber denkbar, daß dieses genaugenommen zeitkonstant ist, Veränderungen also alleine durch mangelnde Reliabilität hervorgerufen werden. Wenn dem so wäre, würden bei der Within-Regression nur Meßfehler in die Schätzung des Effekts dieser Variable eingehen. Tatsächlich wird bei der Within-Regression wie bei allen linearen Regressionsmodellen für beobachtbare Variablen unterstellt, daß die erklärenden Variablen meßfehlerfrei erfaßt werden. Ist dies nicht der Fall, können die Ergebnisse sehr stark verzerrt sein.

---

[8] Vor der Schätzung der Regressionskoeffizienten müssen allerdings auch die Werte der Dummy-Variablen nach Gleichung (13) transformiert werden.

[9] Die Unterschiede zwischen der Within- und der EGLS-Schätzung kann in statistischen Tests ausgenutzt werden, die in gewisser Hinsicht die Angemessenheit des spezifizierten EC-Modells testen (vgl. *Hübler,* 1990: 72f.). Unglücklicherweise ist bei einem signifikanten Testergebnis i.a. nicht sicher, welche Annahmen denn nun tatsächlich verletzt sind.

Falls also mit einem erheblichen Anteil von Meßfehlern zu rechnen ist, sollte statt des EC-Modells für beobachtete Daten auf Modelle zurückgegriffen werden, die solche Meßfehler berücksichtigen können. Entsprechende Modelle werden in der Arbeit von ***Arminger und Müller*** (1990) diskutiert. Die größere Komplexität jener Modelle bringt es jedoch mit sich, daß sie nicht mehr so einfach mit SPSS geschätzt werden können wie das hier behandelte EC-Modell. Als Strategie bietet sich daher möglicherweise an, zunächst mit einfachen Modellen wie dem EC-Modell zu beginnen und erst in einem späteren Analysestadium auf komplexere Modelle umzusteigen.

6. Literatur

Arminger, G. und F. Müller (1990), Lineare Modelle zur Analyse von Paneldaten. Opladen: Westdeutscher Verlag.

Hübler, O. (1990), Lineare Panelmodelle mit alternativer Störgrößenstruktur. **In:** ***G. Nakhaeizadeh und K.-H. Vollmer*** (Hrsg.), Neuere Entwicklungen in der Angewandten Ökonometrie. Beiträge zum 1. Karlsruher Ökonometrie-Workshop. Heidelberg: Physica-Verlag.

Rendtel, U. (1992), Handouts zur Vorlesung "Panelanalyse" beim Frühjahrsseminar 1992 des Zentralarchiv für empirische Sozialforschung, Köln.

Scheuch, E.K., Graf, L. und S. Kühnel (1989), Volkszählung, Volkszählungsprotest und Bürgerverhalten. Ergebnisse der Begleituntersuchung zur Volkszählung 1987. Stuttgart: Metzler-Poeschel.

ANHANG: Dokumentation der SPSS/PC-Anweisungen

```
set more off.
*** EGLS-Schätzung des EC-Modells mit SPSS/PC (V4.01) ***.
TITLE "EGLS-Schätzung des 'Error-Component-Models' (EC-Modell)".

** Vorbereitung: Re-Arrangieren der Paneldaten
   Die zu analysierende Datei "ecbeisp.sys" enthält die Variablen:
   IDNR      Fallnummer
   Y1        abhäng. Variable, 1. Panelwelle
   Y2        abhäng. Variable, 2. Panelwelle
   Y3        abhäng. Variable, 3. Panelwelle
   X11       zeitabh. Prädiktor X1, 1. Welle
   X12       zeitabh. Prädiktor X1, 2. Welle
   X13       zeitabh. Prädiktor X1, 3. Welle
   X21       zeitabh. Prädiktor X2, 1. Welle
   X22       zeitabh. Prädiktor X2, 2. Welle
   X23       zeitabh. Prädiktor X2, 3. Welle
   Z         zeitkonstanter Prädiktor.
SUBTITLE "Rearrangieren der Daten".
GET FILE="ecbeisp.sys".
* 1. Welle.
COUNT nmissing=y1 x11 x21 z (MISSING).
COMPUTE zeit=1.
PROCESS IF (nmissing EQ 0).
SAVE OUT="temp1.sys"
    /KEEP=idnr zeit y1 x11 x21 z /RENAME (y1 x11 x21=y x1 x2).
* 2. Welle.
GET FILE="ecbeisp.sys".
```

```
COUNT nmissing=y2 x12 x22 z (MISSING).
COMPUTE zeit=2.
PROCESS IF (nmissing EQ 0).
SAVE OUT="temp2.sys"
    /KEEP=idnr zeit y2 x12 x22 z /RENAME (y2 x12 x22=y x1 x2).
* 3. Welle.
GET FILE="ecbeisp.sys".
COUNT nmissing=y3 x13 x23 z (MISSING).
COMPUTE zeit=3.
PROCESS IF (nmissing EQ 0).
SAVE OUT="temp3.sys"
    /KEEP=idnr zeit y3 x13 x23 z /RENAME (y3 x13 x23=y x1 x2).
* Wiedereinlesen und sortieren der Fälle.
JOIN ADD /FILE="temp1.sys" /FILE="temp2.sys" /FILE="temp3.sys".
SORT CASES BY idnr zeit.
* Festlegung von Konstanten:
   p:    Anzahl zeitveränderlicher Prädiktoren (hier: 2)
   k:    Anzahl zeitkonstanter Prädiktoren (hier: 1)
   tmax: maximale Anzahl von Beobachtungen pro Fall über die Zeit (hier 3).
COMPUTE p=2.
COMPUTE k=1.
COMPUTE tmax=3.
* Zwischenspeichern der Ausgangsdaten.
SAVE OUT="temp1.sys" /COMP.

***** Schritt 1: Schätzen der Varianzen der Fehlerkomponenten *****.
** Berechnung der Mittelwerte über die Zeit.
SUBTITLE "Berechnung der Mittelwerte über Zeit durch Aggregation".
* die Mittelwerte über die Zeit sind durch ein 'q' nach dem Variablennamen
  gekennzeichnet, also  'yq' für 'Mittelwert von Variable y'.
* Auch die Konstanten p,k und tmax müssen in die Datei mit den
  aggregierten Fällen übertragen werden!.
AGGREGATE OUT=*
    /PRES /BREAK=IDNR
    /yq=MEAN(y)
    /x1q=MEAN(x1)
    /x2q=MEAN(x2)
    /zq=MEAN(z)
    /ti=NU(y)
    /p=MEAN(p)
    /k=MEAN(k)
    /tmax=MEAN(tmax).
* Berechnung der Hilfsgröße 'const' und Speicherung der Mittelwerte
  in Datei 'tempm.sys'.
COMPUTE const=1.
SAVE OUT="tempm.sys" /COMP.

**  Zwischenspeichern der Fallzahl in 'temp2.sys".
AGGREGATE OUT="temp2.sys" /BREAK=const /NFALL=NU(idnr).

**  Regression über die Mittelwerte: Between-Regression über vollständige
    Fälle; Sichern der Residuen in Variable 'bresid'.
SUBTITLE "Berechnung 'between'-Regression' über vollständige Fälle".
SELECT IF (ti=tmax).
REGRESSION VAR yq x1q x2q zq
    /DEP yq
    /ENT
    /SAVE=RESID(bresid).

** Berechnung und Speicherung der quadrierten Between-Residuen und der
   Fallzahl der Regressions in die Variablen 'sseb' und 'nbetw' in Datei
   'tempb.sys'.
SUBTITLE "Berechnung u. Speicherung SSE-'beetween' und Fallzahl".
COMPUTE bres2=bresid*bresid.
AGGREGATE OUT="tempb.sys" /PRES /BREAK=const
 /nbetw=NU(idnr)
 /sseb=SUM(bres2)
```

```
 /p=MEAN(p)
 /k=MEAN(k)
 /tmax=MEAN(tmax).

** Schätzung within-Regression.
*  Die mittelwertsbereinigten Daten sind durch ein 'w' hinter den
   Variablennamen gekennzeichnet, also 'yw' für die mittelwertsbereinigte
   Variable 'y'.
SUBTITLE "Mittelwertsbereinigung für 'within'-Regression".
JOIN MATCH  TABLE="tempm.sys"
          /FILE="temp1.sys"
          /BY idnr.
COMPUTE yw=y-yq.
COMPUTE x1w=x1-x1q.
COMPUTE x2w=x2-x2q.
*  within-Regression.
*  Die Residuen werden in Variable 'wresid' festgehalten.
SUBTITLE "within-Regression (ohne Konstante!)".
REGRESSION VAR yw x1w x2w
         /ORIGIN /DEP=yw /ENT
         /SAVE=RESID(wresid).

** Speicherung der quadrierten Residuensumme ('ssew') und der Anzahl
   Beobachtungen (über Fälle und Zeit: 'nges') in Datei 'tempw.sys'.
SUBTITLE "Berechnung u. Speicherung SSE-'within' und Beobachtungszahl".
COMPUTE wres2=wresid*wresid.
SAVE OUT="temp1.sys" /COMP.
AGGREGATE OUT="tempw.sys" /PRES /BREAK=const
  /nges=NU(idnr)
  /ssew=SUM(wres2).

** Berechnung der Varianzen 'sigma2e' und 'sigma2d' der Fehlerkomponenten.
SUBTITLE "Berechnung der geschätzten Fehlerkomponenten".
JOIN MATCH FILE="tempb.sys" /FILE="tempw.sys" /FILE="temp2.sys".
COMPUTE sigma2e=ssew/(nges-nfall-p).
COMPUTE sigma2d=sseb/(nbetw-p-k-1) - sigma2e/tmax.
* Nullsetzen von 'sigma2d' bei negativer Schätzung.
IF (sigma2d LT 0) negvar=1.
IF (sigma2d LT 0) sigma2d=0.
FORMATS sigma2e sigma2d (F12.5).
SUBTITLE "***** Varianzen der Fehlerkomponenten:".
LIST VARIABLES= nfall nges nbetw sigma2e sigma2d negvar.

***** Schritt 2: Berechnung der EGLS-Lösung über OLS transformierter Daten.
** Berechnung der Transformationsformel 'theta'.
SUBTITLE "Berechnung von 'theta'".
JOIN MATCH TABLE=*   /FILE="tempm.sys" /BY const.
COMPUTE theta=1-SQRT(sigma2e/(sigma2e+sigma2d*ti)).

** Transformation der Ursprungsdaten und EGLS-Regression.
*  Die transformierten Daten sind durch ein 't' nach dem Variablennamen
   gekennzeichnet, so 'constt' für die transformierte Konstante.
SUBTITLE "EGLS-Regress. über OLS der transform. Daten".
JOIN MATCH TABLE=* /FILE="temp1.sys" /BY idnr.
COMPUTE constt=1-theta.
COMPUTE yt=y-theta*yq.
COMPUTE x1t=x1-theta*x1q.
COMPUTE x2t=x2-theta*x2q.
COMPUTE zt=z-theta*z.
REGRESSION VAR=yt constt x1t x2t zt
   /ORIGIN /DEP=yt /ENT.
*********************************************************************.
```

RZOO: Ein Hilfsmittel zur Verwaltung sozialwissenschaftlicher Daten

In der sozialwissenschaftlichen Forschungspraxis sind häufig große Datenmengen zu verarbeiten, die erhebliche Speicherkapazitäten erfordern. Es liegt deshalb nahe, solche Daten in komprimierter Form abzulegen. Das kann mit zahlreichen Archivierungsprogrammen erreicht werden; zum Beispiel mit den Programmen PKZIP, ARC, LHARC und ZOO, die als Freeware oder Shareware allgemein verfügbar sind. Die damit erreichbaren Kompressionsraten sind beträchtlich. Zum Beispiel lassen sich die Daten des Sozio-ökonomischen Panels (zur Zeit für die ersten sechs Wellen, 1984-1989) auf ungefähr ein Zehntel ihres ursprünglichen Umfangs reduzieren (von etwa 105 auf etwa 12 Mega Bytes).

Ein Nachteil der Verwendung komprimierter Datenarchive liegt darin, daß sämtliche Datenfiles, auf die ein Benutzer zugreifen möchte, zuvor wieder expandiert werden müssen. Das ist auch dann erforderlich, wenn nur ein kleiner Teil der archivierten Daten für weitere Auswertungen erforderlich ist. In der praktischen Arbeit ist dies der Regelfall; und es wäre deshalb hilfreich, wenn direkt und selektiv auf vom Benutzer spezifizierbare Teile eines komprimierten Datenarchivs zugegriffen werden könnte.

Um dies zu erreichen, wurde das Programm RZOO entwickelt. Als Ausgangspunkt wurde das von Rahul Dhesi geschriebene Archivierungsprogramm ZOO gewählt. Es ist nicht nur eines der leistungsfähigsten Archivierungsprogramme, es ist auch frei verfügbar und bereits auf zahlreichen Rechenanlagen installiert. Auch der Sourcecode des Programms ist allgemein zugänglich und konnte somit für die Entwicklung von RZOO genutzt werden. Die wichtigsten Merkmale von RZOO sind:

- Es kann direkt und selektiv auf Daten zugreifen, die in Form komprimierter ZOO Archive gespeichert sind. Eine separate Expansion der Daten ist nicht erforderlich.
- Infolgedessen ist nicht nur der Speicherplatzbedarf für die Daten gering, sondern es reduziert sich auch die für das Lesen der Daten von einer Festplatte erforderliche Zeit.
- Die gespeicherten Daten können als Variablen beschrieben werden, so daß dann mithilfe von Variablennamen beliebige Teile des Archivs herausgelesen werden können.
- Es gibt eine verhältnismäßig flexible Syntax zur Selektion von Records aus den gespeicherten Datenfiles.
- Es ist möglich, Records aus unterschiedlichen Datenfiles zu verknüpfen und in einem Ausgabe-Datenfile zu mischen.

Im Vergleich zu üblichen Datenbankprogrammen sind die Fähigkeiten von RZOO natürlich sehr beschränkt. Dabei sollte jedoch bedacht werden, daß in den meisten Fällen die Struktur sozialwissenschaftlicher Datensätze sehr einfach und eine komplizierte (relationale) Datenbankstruktur nicht erforderlich ist. Dies gilt zum Beispiel auch für das Sozio-ökonomische Panel. Dieser Datensatz ist zwar umfangreich und auf zahlreiche unterschiedliche Datenfiles verstreut; seine Struktur ist jedoch einfach: es gibt nur drei grundlegende Kategorien - Personen, Haushalte und Erhebungswelle -, nach denen die Daten organisiert sind. Wie sich gezeigt hat, ist der Aufbau eines ZOO Archivs für diesen Datensatz auf einfache Weise möglich und liefert einen effizienten und für die meisten Anwendungsfälle hinreichenden Datenzugriff.

Exemplarisch sei deshalb anhand des Sozio-ökonomischen Panels (SOEP) die Einrichtung eines ZOO Archivs, das mit RZOO genutzt werden kann, verdeutlicht. Drei Schritte sind erforderlich.

In einem ersten Schritt muß ein ZOO Archiv gebildet werden, das die zu verwendenden Datenfiles enthält. In diesem Beispiel handelt es sich um die Rohdatenfiles, wie sie für das SOEP zur Verfügung gestellt werden. Praktisch geschieht dies mithilfe des ZOO Archivierungsprogramms. Das resultierende Archiv enthält dann den gesamten Datenbestand. Es ist portabel zwischen verschiedenen Rechenanlagen; und falls es erforderlich sein sollte (z.B. für Updates), können die im Archiv enthaltenen Datenfiles jederzeit wieder herausgezogen, modifiziert und erneut ins Archiv übernommen werden.

In einem zweiten Schritt ist es erforderlich, die Variablen zu beschreiben, die in den Datenfiles des Archivs vorhanden sind. Das geschieht, indem ein sogenanntes Variablenbeschreibungsfile erzeugt wird. Es handelt sich um ein einfaches Textfile, das zeilenweise für jede Variable folgende Informationen enthält: Den Namen der Variablen; eine Angabe des Datenfiles, in dem sich die betreffende Variable befindet; Angaben dazu, wo in den Records des Datenfiles die Variable gefunden werden kann (Offset und Feldlänge); und schließlich (optional) eine Beschreibung der Variablen. Für das SOEP kann zum Beispiel ein solches Variablenbeschreibungsfile auf einfache Weise aus den Dictionary-Files der für das SOEP existierenden SIR Datenbank erzeugt werden. Allerdings müssen sämtliche Variablennamen eines RZOO Archivs eindeutig sein. Da dies beim SOEP nicht durchgängig der Fall ist, sind in diesem Fall einige Variablenbenennungen erforderlich.

Als zusätzliche Option kann ein weiteres File erstellt und in das Archiv aufgenommen werden, das nicht nur Variablenbeschreibungen (variable labels), sondern auch Beschreibungen der Kategorien der Variablen (value labels) enthält. Dies ist insbesondere für ein SOEP Archiv hilfreich, da sich die Definitionen einiger Variablen im Laufe der Zeit geändert haben.

In einem dritten Schritt muß schließlich noch die logische Struktur des Archivs beschrieben werden. Dies geschieht wiederum mithilfe einfacher Textfiles, die zunächst angeben, welche Daten- und Variablenbeschreibungsfiles sich im Archiv befinden, und deren Größen, d.h. Anzahl und Länge der Records, definieren. Zusätzlich können die Archiv-Beschreibungsfiles Informationen darüber enthalten, wie die Records unterschiedlicher Datenfiles miteinander zu verknüpfen sind. Dies geschieht im einfachsten Fall durch die Angabe sog. matching variables. Um zum Beispiel die Datenfiles des SOEP miteinander zu verknüpfen, die nach Personennummern organisiert sind, sind diese Personennummern als matching variables anzugeben. Ganz entsprechend für die Datenfiles, die nach Haushaltsnummern organisiert sind. Um Personen und Haushalte zu verknüpfen, können Zeigervariablen verwendet werden, wie sie etwa durch die Pfadvariablenfiles des SOEP zur Verfügung gestellt werden.

Grundsätzlich ist es möglich, und in vielen Fällen auch sinnvoll, ein Archiv auf mehrere unterschiedliche Weisen zu beschreiben. Jede dieser Beschreibungen bildet gewissermaßen einen Filter, durch den das Archiv für das Retrieval-Programm RZOO eine spezifische logische Struktur erhält. Dadurch ist dann ein nicht nur flexibler, sondern auch effizienter Zugriff auf das Archiv möglich.

Nach diesen drei Schritten ist der Aufbau eines ZOO/RZOO Archivs abgeschlossen. RZOO kann dann verwendet werden, um - im Rahmen der jeweils definierten logischen Archivstruktur - beliebige Daten aus dem Archiv herauszulesen. RZOO ist kein interaktives Programm, sondern wird durch Kommandos gesteuert. Mit diesen Kommandos wird etwa angegeben: das zu verwendende Archiv und seine logische Struktur, die aus dem Archiv zu lesenden Variablen, der Name eines Ausgabefiles, das Ausgabeformat, ggf. Bedingungen zur Selektion von Records usw. Beliebige Mengen dieser Kommandos können wiederum in ein File geschrieben und RZOO kann dann mit einem solchen Kommandofile aufgerufen werden.

RZOO existiert zur Zeit in der Version 1.1; es ist, wie auch das Archivierungsprogramm ZOO, frei verfügbar.

Götz Rohwer
European University Institute
Via dei Roccettini, 9
1-50016 San Domenico di Fiesole (FI)
Fax: 0039/55/5092-201, e-mail: ROHWER IFIIUE.FI.CNR.IT.

Eine geschlossene Kollektion empirischer Studien der Hochschule für Architektur und Bauwesen Weimar im Zentralarchiv

von Evelyn Brislinger und Brigitte Hausstein[1]

ASI-Tagung in Weimar

Weimar in Thüringen, als Zentrum der deutschen Klassik und durch seine Nähe zu Buchenwald weltbekannt, ist Sitz einer Hochschule für Architektur und Bauwesen, deren früheste Traditionen bis in das Jahr 1860 reichen. Hier fand im Oktober/November 1991 die wissenschaftliche Jahrestagung der "Arbeitsgemeinschaft Sozialwissenschaftlicher Institute e.V." (ASI) unter Leitung von Prof. *Uta Gerhardt* und *Ekkehard Mochmann* statt. Die Thematik "Gesellschaftlicher Umbruch 1945 und 1990 - Lebensverhältnisse und Re-Demokratisierung" gab auch Anlaß, über die Datenlage zu diskutieren, die von der DDR-Sozialforschung hinterlassen wurde.

Eine konkrete Übersicht über den Umfang und die Datenlage der DDR-Sozialforschung stand zu diesem Zeitpunkt nicht zur Verfügung und dementsprechend breit war das Meinungsspektrum. Eine prononciert vorgetragene Äußerung, die DDR-Gesellschaft ließe sich mangels eigener kontinuierlicher empirischer Analysen nur mit Daten erklären, die durch "westliche" Forschungsinstitute stellvertretend bei "DDR-Reisenden" erhoben wurden, stieß mehrheitlich auf Kritik. Die methodischen Grenzen dieses Ansatzes sind offensichtlich. Tatsächlich steht eine Vielzahl von Daten zu unterschiedlichsten Themen zur Verfügung, die eine kritisch vergleichende Analyse mit der Stellvertreterforschung ermöglicht. Allerdings war es selbst für aufgeschlossene, an vergleichenden Forschungsarbeiten und Sekundäranalysen interessierte Wissenschaftler bisher schwer, einen Überblick über das empirische Datenmaterial aus der ehemaligen DDR zu gewinnen.

Die Sozialforschung in der DDR war bekanntlich durch restriktive Arbeitsbedingungen, durch rigide Beschränkungen in den Forschungsansätzen und die strikte Geheimhaltung des Datenmaterials geprägt. Eine wissenschaftliche Öffentlichkeit konnte sich kaum entwickeln. Nach *Lepsius* sagt der geringe Bestand an Publikationen und Ergebnisdarstellungen aus der empirisch-soziologischen Forschung demzufolge nichts über deren wirklichen Umfang aus[2] und läßt einen realen Überblick nicht zu. Hinzu kommt die gegenwärtige Neustrukturierung der Forschungslandschaft. Der wissenschaftliche Alltag in den neuen Bundesländern ist

---

1 *Evelyn Brislinger* und *Brigitte Hausstein* sind wissenschaftliche Mitarbeiterinnen des Zentralarchivs in der Außenstelle Berlin.

2 Vgl.: *M. Rainer Lepsius:* Zur Lage der Soziologie an den Universitäten der DDR. In: Kölner Zeitschrift für Soziologie und Sozialpsychologie, 2/1990, S.321.

durch Evaluierung und Abwicklung (sprich Auflösung) vieler Institute, durch inhaltliche und personelle Neubestimmung an den Hochschulen und durch marktwirtschaftliche Versuche auch in der Meinungs- und Sozialforschung bestimmt. Zunehmend sind Vereinzelung von Sozialforschern und Rückzug aus der Wissenschaft zu beobachten.

Entsprechend unübersichtlich stellt sich dem an empirischen Ergebnissen aus der ehemaligen DDR interessierten Sozialforscher die Datenlage dar. Der themenspezifische Überblick über empirische Studien läßt sich nur mit hohem Rechercheaufwand gewinnen. Die sekundäranalytische Nutzung der Daten erfordert außerdem den Einsatz umfangreicher technischer und personeller Kapazitäten zur Aufbereitung, weil sie meist in ungenügender Qualität vorliegen. Letztlich ergeben sich Aufwendungen, die für den einzelnen Forscher nicht zu tragen sind. Die Nichtverfügbarkeit solcher empirischer Studien würde den Verlust unersetzbarer Informationen über die Entwicklung der ehemaligen DDR und "ihrer" Sozialforschung bedeuten. Das wiederum würde Äußerungen über die Nichtexistenz empirisch-soziologischer Forschung aus dieser Zeit stützen.

Das Projekt zur Archivierung der Umfragedaten aus der ehemaligen DDR

Um den Gefahren des Verlustes von Daten und Informationen und sich lösender Arbeitskontakte in der Forschungslandschaft der neuen Bundesländer zu begegnen, wurde bereits 1990 durch das Zentralarchiv Köln und Mitarbeiter des Instituts für Soziologie und Sozialpolitik an der Akademie der Wissenschaften der ehemaligen DDR mit der Sicherung, Aufbereitung und Archivierung der Daten begonnen.

Das vorrangige Ziel war dabei von Beginn an der Aufbau einer möglichst umfangreichen, dem Standard der Altbundesländer vergleichbaren, empirischen Datenbasis. Sie soll empirisch-soziologische Studien aus der ehemaligen DDR und aktuelle Forschungsergebnisse aus den neuen Bundesländern umfassen. Nach inhaltlichen, methodischen und technischen Kriterien ausgewählte Studien werden in den Datenbanken

- zur Sozialgeschichte der DDR und
- zum sozialen Wandel in den neuen Bundesländern nach der Wahl im März 1990[3]

organisiert, und so für sekundäranalytische Forschungen, speziell zum sozialen und politischen Wandel in Deutschland und Osteuropa, verfügbar gemacht.

---

3 Vgl.: ***Ekkehard****Mochmann:* Archivierung, Aufbereitung, Dokumentation und Bereitstellung von Umfragen aus der Sozialforschung der DDR, Projektantrag, Köln, Mai 1990 sowie: ***Ekkehard Mochmann:*** Die unveröffentlichte Meinung: Zur Sicherung von Umfragen aus der DDR für Sekundäranalysen: In: *Dieter* ***Jaufmann, Ernst Kistler, Klaus Meier, Karl-Heinz Strech*** (Hrsg.): Empirische Sozialforschung im vereinten Deutschland, Campus-Verlag Frankfurt/New York, 1992.

Bisherige Recherchen haben gezeigt, daß ältere Studien aus dem umfangreichen - thematisch und methodisch breit angelegten - empirischen Datenmaterial[4] nicht mehr verfügbar sind. Die Ursachen hierfür waren, vor allem in der frühen DDR, ökonomischer Art, d.h. permanenter Mangel an Datenträgern und zu hohe Aufwendungen für deren Pflege. Mit dem gesellschaftlichen Umbruch - und den auch in diesen Fragen oft ungeklärten Eigentumsverhältnissen - wurden anläßlich der Schließung von Rechenzentren "alte" Magnetbänder und mit der Auflösung der Institute "unhandliche" Lochkarten und Erfassungslisten säckeweise vernichtet. Auch das fehlende Interesse an einer sekundäranalytischen Aufbereitung vorhandener Daten führte dazu, daß selbst bei profilierten Instituten der Verlust einzelner Studien bis in die Zeit der Wende reicht.

Erfolgreiche Recherchen lassen jedoch auch eine Reihe von Besonderheiten des empirischen Datenmaterials deutlich werden, die aus Sicht der Datengeber einer Übergabe der Studien an das Zentralarchiv scheinbar entgegenstehen.

- Ein großer Teil der empirischen Studien konnte bisher selbst von Primärforschern nur partiell inhaltlich ausgewertet werden. Das führt bei vielen zu der Befürchtung, daß eigene zukünftige Forschungsvorhaben durch eine Aufnahme der empirischen Ergebnisse in die Datenbank des Zentralarchivs gefährdet sind. Gerade hier erhalten die Zugangskategorien, die vom Datengeber jeweils selbst vergeben werden und die Möglichkeit des Zugriffs und sekundärer Analysen durch andere festlegen, besondere Bedeutung[5]. Der Zugang zu den Daten ist demnach nur in Übereinstimmung mit den Forschungsinteressen des Datengebers möglich.
- Die sekundäranalytische Nutzung der in der ehemaligen DDR erhobenen Daten setzt die kritische inhaltliche und methodische Auseinandersetzung mit ihnen voraus. Hierbei geht es nicht vordergründig um eine politische Diskreditierung einzelner Forscher, sondern vielmehr darum, ihr konzeptionelles Wissen zu nutzen, aus Publikationen und Forschungsberichten Informationen zum Studiendesign zusammenzutragen und für eine qualitative Bewertung der Daten allgemein zugänglich zu machen. Eine studienspezifische Aufarbeitung und Dokumentation soll die Besonderheiten, der aus einer konkreten politischen, ökonomischen und sozialen Situation abgeleiteten Forschungsthemen, deutlich machen. Dies

---

4 **Vgl.:** ***Michael Hader:*** Nun kann auch die soziologische Methodik zusammenwachsen, in: ZA-Information 27, November 1990, S.138.

5 Die Zugangskategorien, die jeweils im Datenbestandskatalog vermerkt sind, umfassen folgende Möglichkeiten:
0 Für jedermann freigegeben
A Frei für wissenschaftliche Forschung und Lehre an Hochschulen sowie für öffentlich finanzierte Institute oder Projekte
B Wie A, jedoch bei Veröffentlichung oder weitergehenden Verarbeitung der Ergebnisse Genehmigung des Datengebers vorab erforderlich
C Grundsätzlich nur nach vorheriger Genehmigung des Datengebers frei; die Genehmigung holt das Zentralarchiv unter Angabe des Benutzers und des Auswertungszweckes ein.

schließt die in der DDR-Sozialforschung üblichen Kontrollen und Auflagen vom Studiendesign bis hin zur konkreten Fragestellung ein und berücksichtigt die sozialstrukturelle Spezifik und den verwendeten Begriffsapparat, die Antwortbereitschaft und Besonderheiten des Antwortverhaltens von Befragten.

- Die empirischen Untersuchungsergebnisse liegen in sehr unterschiedlicher technischer Qualität vor; die Studiendokumentationen sind zumeist unvollständig. Die mangelhafte technische Ausstattung der Institute und Hochschulen in der ehemaligen DDR und das Auseinanderbrechen der Forschergruppen haben in vielen Fällen dazu geführt, daß der Zustand der Forschungsunterlagen eine weitere Auswertung selbst für die Primärforscher nicht ohne weiteres zuläßt. Die in den nunmehr geschlossenen Rechenzentren ausgewerteten Daten, die häufig auf verschlissenen Bändern mit kaum noch gebräuchlicher Schreibdichte gespeichert sind, können mit Hilfe der Möglichkeiten des Zentralarchivs bearbeitet werden. Für den PC technisch aufbereitete Dateien stehen im Ergebnis gemeinsamer Bemühungen nicht zuletzt dem Primärforscher selbst zur Verfügung.

Eine geschlossene Kollektion: Die Archivierung empirischer Studien der Hochschule für Architektur und Bauwesen Weimar

Angesichts gravierender Probleme im Umgang mit Daten war die Sicherung und Aufbereitung der empirischen Studien der Hochschule für Architektur und Bauwesen Weimar ein Erfolg.

Diese Hochschule bildet in den universitären Studiengängen Architektur, Bauingenieurwesen, Baustoffverfahrenstechnik, Informatik und Raumplanung (ehemals Sektion Gebietsplanung und Städtebau)[6] aus. Der Lehrstuhl Soziologie, für den es begründete Hoffnungen auf weitere Existenz und stadtsoziologische Forschungsarbeit gibt, gehört zur Fakultät Raumplanung. Im Kontext des Demokratisierungsprozesses entstand im Frühjahr 1990 das Vorhaben, die 1951 geschlossene Fakultät *Bildende Kunst* neu zu gründen, und damit den Bauhaus-Gedanken wieder aufzugreifen. Mit der Zeit des "Staatlichen Bauhauses in Weimar" verbinden sich die Namen solch berühmter Architekten und Maler wie *Walter Gropius, Paul Klee*, Wassily Kandinsky und *Lyonel Feininger*[7].

Die empirischen Studien, die seit 1980 am Lehrstuhl durchgeführt wurden, konnten vollständig datentechnisch aufbereitet und mit den zusammengetragenen Forschungsunterlagen archi-

6 Vgl.: Weimar - Positionen zur Stadtentwicklung aus der Sicht der Hochschule für Architektur und Bauwesen, Wissenschaftliche Zeitschrift der Hochschule für Architektur und Bauwesen - A -, Weimar 37/1991, S. 51.

7 Ebenda, S. 89f.

viert werden (siehe Abbildung 1). Entscheidend hierfür waren das Interesse und das Engagement einiger Wissenschaftler des Lehrstuhls für Soziologie[s] und die Unterstützung durch Mitarbeiter des Zentralarchivs in der Berliner Außenstelle.

Abbildung 1: 13 Datensätze aus den Jahren 1980 bis 1991

Studienbezeichung	Jahr	Befragte	Variablen
Schwerin: Revitalisierung des Stadtteils "Großer Dreesch"	1991	N = 196	83
Europäische Provinz Weimar	1990	N = 487	148
Soziologische Untersuchung Sömmerda	1990	N = 52	80
Stadtentwicklung im Umbruch Dessau	1989	N = 427	119
Soziologische Untersuchung Schwerin-Schelfstadt	1989	N = 84	88
Soziologische Untersuchung Jena	1988	N = 476	100
Stadtentwicklung und Wohnmilieu Brandenburg	1987	N = 512	91
Stadtentwicklung und Wohnmilieu Magdeburg	1986	N = 436	75
Stadt und Siedlungsgruppe Eisenach	1985	N = 400	76
Stadtentwicklung und Wohnmilieu Halle/S. und Halle/Neustadt	1984	N = 360	77
Soziologische Untersuchung von Wohngebieten der Stadt Rostock	1982	N = 661	77
Soziologische Untersuchung Gotha/Innenstadt	1981	N = 271	77
Soziologische Probleme der Rekonstruktion der Nördlichen Innenstadt von Erfurt	1980	N = 233	26

8 Besonderer Dank für ihre Unterstützung gilt Frau Dr. ***Christine Weiske*** am Lehrstuhl für Soziologie der Hochschule, Frau *Uta Schäfer*, jetzt im Graduiertenkolleg an der Universität zu Köln, Frau ***Andrea Gaube,*** jetzt in der Gesellschaft für behutsame Stadterneuerung "Stern".

Die Geschlossenheit dieser Studienkollektion läßt sie als stadtsoziologisches Zeitdokument besonders interessant erscheinen. Der Untersuchungszeitraum reicht von den letzten Jahren der DDR, über die Zeit des radikalen gesellschaftlichen Umbruchs bis in die jüngste Vergangenheit hinein. Das macht es möglich, die Entwicklung nachzuvollziehen, die die Weimarer Stadtsoziologen in ihrer Beziehung zu dem, im Verständnis der Partei- und Staatsführung der ehemaligen DDR, zentralen und sensiblen Thema der Verwirklichung und Folgen des Wohnungsbauprogramms, genommen haben.

Im Vergleich der Untersuchung "Europäische Provinz Weimar 1990" mit früheren Studien in Brandenburg 1988 oder Magdeburg 1987 wird von dem Versuch einer Neubestimmung des Forschungsansatzes gesprochen. Demnach wurde die bisher vorherrschende Orientierung auf den städtebaulichen Planungsprozeß, der weitgehend von politischen Entscheidungsträgern geprägt war, durch eine genauere Beschreibung alltäglicher Lebensbezüge aus "Bewohnersieht" ergänzt[9].

Die Abschaffung politischer Restriktionen hat auch in der Forschung und für einen Teil der Sozialforscher aus der ehemaligen DDR neue Freiräume geschaffen. Dies bringt nicht nur für das Weimarer Forschungskonzept, das für die gesamte Studienkollektion gilt, neue inhaltliche und methodische Probleme mit sich. Es ist der Konflikt zwischen dem Wunsch, den erreichten Standardisierungsgrad in der Methodik auch zukünftig beizubehalten, und der Notwendigkeit, sich den neuen gesellschaftlichen Gegebenheiten durch ein entsprechendes Instrument und neue methodische Zugänge zu nähern.

Die empirischen Untersuchungen der Hochschule wurden bisher im "kommunalen Praktikum" von Studenten der Sektion "Gebietsplanung und Städtebau" unter Leitung von Wissenschaftlern des Lehrstuhls für Soziologie durchgeführt. Charakteristisch für diese Studien ist ihre vergleichbare inhaltliche Ausrichtung. Es sind die Themenkomplexe 'Wohnung", "unmittelbare Wohnumgebung", "Wohngebiet", "gesamtstädtische Probleme", "Stadtzentrum" und 'Verhältnis der Bewohner zueinander", die immer wiederkehren und durch jeweils besondere Fragenkomplexe ergänzt wurden.

In die Untersuchungen wurden ausgewählte Wohnbereiche, die "typische städtebauliche und soziale Situationen" repräsentieren, einbezogen. Auswahlkriterien in der Weimarer Untersuchung 1990 waren zum Beispiel: "Strukturtypen der Bausubstanz, Lagebeziehung zur Stadt und zur Landschaft und Gebietsimage"[10]. Dort, wo Forschungsdokumentationen bereits exi-

---

9 Vgl.: *Uta Schäfer:* III. Methodische Probleme: "Weimar in der Wende - Untersuchungsansatz und Methodenkritik" in: Europäische Provinz Weimar, in der Schriftenreihe der Hochschule für Architektur und Bauwesen Weimar (noch unveröffentlicht).

10 Ebenda

stieren oder die erstellten stadtsoziologischen Gutachten beschafft werden konnten, weisen sie erkannte Abweichungen zwischen Stichprobe und Bewohnerstruktur auch aus und unterscheiden zwischen Aussagen, die die gesamte Stadt repräsentieren und Trendaussagen.

Mit dem Ziel, ein möglichst umfassendes Bild der vielfältigen Beziehung zwischen der Stadt und ihren Bewohnern zu gewinnen, wurde eine Standardmethodik als Kombination und Ergänzung verschiedener soziologischer und städtebaulicher Erhebungsmethoden eingesetzt[11]. Dieses speziell entwickelte Erhebungsdesign ist in seinem komplexen Charakter für den Methodenstandard in der ehemaligen DDR bemerkenswert.

Die angewandte komplexe Methodik setzt sich im standardisierten Teil aus folgenden quantitativen und qualitativen Erhebungsmethoden zusammen[12]:

Befragung

- Nach einem standardisierten Fragebogen wurden Interviews mit zufällig ausgewählten Hauptmietern von Wohnungen zu den Themen: "engerer Wohnbereich", "gesellschaftliche Einrichtungen", "Arbeitsstätten", "öffentliche Freiräume", "sozialräumliche Beziehungen" und "Bewohnerstruktur" durchgeführt.
- Eine Ergänzung hierzu stellen Kurzinterviews mit Passanten und Nutzern "gesellschaftlicher Einrichtungen" bzw. "öffentlicher Freiräume", Expertengespräche mit Vertretern staatlicher und gesellschaftlicher Organisationen, und Gruppeninterviews mit Interessenten und Initiativgruppen nach einem standardisiertem Leitfaden dar.

Beobachtung

- Mit dem Ziel, typische soziale Aktivitäten der Bewohner in ausgewählten "Wohnbereichen" und "öffentlichen Freiräumen" zu erfassen, wurden die Beobachtungen in offener, nichtteilnehmender Form durchgeführt.

Dokumentenanalyse

- Statistische, historische und planerische Dokumente wurden im Hinblick auf "engeren Wohnbereich", "gesellschaftliche Einrichtungen", "Arbeitsstätten" und "Bewohnerstruktur" analysiert.

---

11 Vgl.: *Bernd Hunger*. Soziologische Untersuchungen als Bestandteil der städtebaulichen Planung der Umgestaltung von Altbau Wohngebieten, Hochschule für Architektur und Bauwesen Weimar, 1982, Dissertation A.

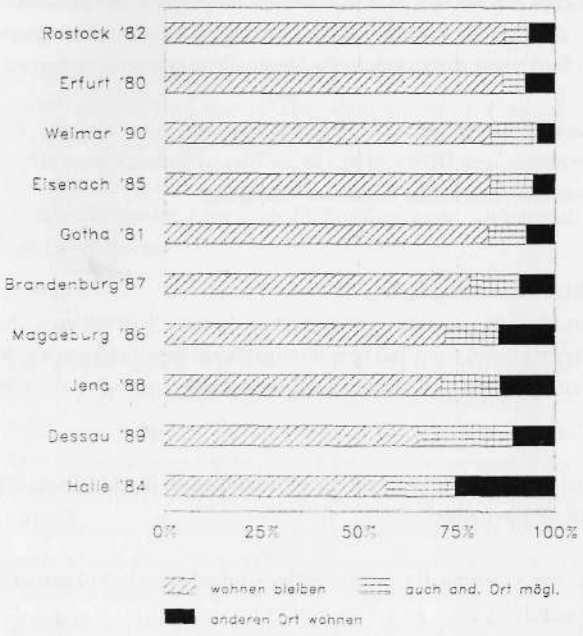
12 In der Darstellung wurden Informationen aus den verschiedenen vorhandenen Studien-Dokumentationen zusammengestellt.

Die Daten der Erhebungen, die in Abbildung 1 aufgeführt sind, werden ins Zentralarchiv aufgenommen. Die exemplarische Beschreibung einer Studie, wie sie in den Datenbestandskatalog aufgenommen wurde, zeigt Abbildung 2 (im Anhang). Die Ergebnisse der kombiniert eingesetzten Methoden, ergänzt um Freihandzeichnungen und Fotodokumentationen, wurden in der Schriftenreihe der Hochschule publiziert.

Beispiel für die Möglichkeit sekundärer Analysen:

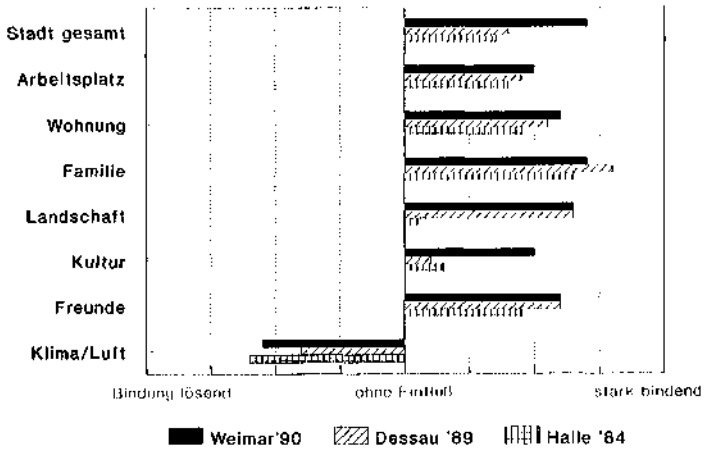
Daten, die einen Zeitraum von mehr als zehn Jahren umfassen, und ein hoher Standardisierungsgrad in der Methodik fordern die komplexe sekundäranalytische Auswertung geradezu heraus. In den Untersuchungen der Hochschule zeigte sich, daß die Merkmale "Heimischfühlen" und "Gebundensein an die Stadt" Ausdruck der emotionalen Bewertung der Lebensbedingungen am Wohnort sind[13]. Wie verschieden diese Merkmale in ausgewählten Orten der ehemaligen DDR bewertet wurden, zeigen die Abbildungen 3 und 4.

Abbildung 3: "Heimischfühlen" - Städte im Vergleich



13 Vgl.: *Christine Weiske:* Heimischfühlen in der Stadt - zur Wechselwirkung von Ortsverbundenheit und Migration, FSU Jena 1984. Dissertation A

Abbildung 4: "Faktoren der Bindung an die Stadt"



Die Graphiken vermitteln einen ersten Einblick in die empirischen Ergebnisse aus den Erhebungen in den unterschiedlichen Städten und zu verschiedenen Zeitpunkten. Sie machen nicht die Differenzierungen deutlich, die vor allem durch Besonderheiten der einbezogenen Wohngebiete (Alt- und Neubau) sowie sozialstrukturelle Merkmale hervorgerufen wurden.

Für sekundäranalytische Beratungs- und Serviceleistungen hinsichtlich dieser und weiterer empirischer Studien aus der ehemaligen DDR und den neuen Bundesländern stehen Mitarbeiter im Zentralarchiv Köln und der Außenstelle Berlin zur Verfügung.

Zusammenarbeit mit weiteren Einrichtungen:

Die empirischen Studien der folgenden (zum Teil bereits aufgelösten) Einrichtungen werden gegenwärtig gesichtet, für eine Archivierung technisch aufbereitet und hinsichtlich ihres Studiendesigns beschrieben:

- Institut für Theorie, Geschichte und Organisation der Wissenschaft an der Akademie der Wissenschaften der DDR, Berlin (aufgelöst)

- Institut für Soziologie an der Akademie für Gesellschaftswissenschaften beim Zentralkomitee der SED, Berlin (aufgelöst)

- Institut für Soziologie und Sozialpolitik an der Akademie der Wissenschaften der DDR, Berlin (aufgelöst)
- Humboldt-Universität zu Berlin, Charité
- Institut für Sozialdatenanalyse e.V., Berlin (ISDA)
- Berliner Institut für sozialwissenschaftliche Studien GmbH (BISS)

Darüber hinaus sollen die Studien der Unabhängige Serviceeinrichtung für Umfragen, Methoden und Analysen (USUMA), des Zentralinstituts für Jugendforschung (ZU) und des Rundfunks/Fernsehens der DDR gesichtet werden.

Abbildung 2: Studienbeschreibung zu der Untersuchung "Europäische Provinz Weimar 1990"

Erhebungszeitraum:
Juni und Juli 1990

Primärforscher und Datenerheber:
Wissenschaftler des Lehrstuhls für Soziologie an der Hochschule für Architektur und Bauwesen, Fakultät Raumplanung, Weimar

Datenerhebung:
Wissenschaftler des Lehrstuhls für Soziologie und Studenten der Studiengänge Städtebau und Gebietsplanung

Inhalt:
Analyse der vielfältigen Lebensansprüche, die sich auf die Stadt als Lebensraum richten.

Themen: Wohndauer; Anzahl der Räume; Zufriedenheit mit der Wohnung; Wissen über Besitzverhältnisse bezüglich des Wohnhauses; Einstellung zur 3-Generationen-Familie und zur Wohngemeinschaft; Eingebundensein in die Nachbarschaft und wichtige gemeinschaftliche Aktivitäten; Möglichkeiten zur Befriedigung ausgewählter Bedürfnisse in der unmittelbaren Wohnumgebung; Wunsch-Wohngegend in Weimar; Nutzung und Einschätzung der Einkaufsmöglichkeiten; Polaritätsprofil zu den Eindrücken vom Wohngebiet; Bereitschaft zu Aktivitäten im Wohngebiet; Orte für Freizeitbeschäftigung und Häufigkeit der Nutzung bestimmter Räumlichkeiten; Häufigkeit privater Kontakte mit ausgewählten Personengruppen; Besitz an Grundstücken und Fahrzeugen; Nutzung von Verkehrsmitteln bei unterschiedlichen Anlässen; Gründe und benötigte Zeit für Besuche des Stadtzentrums; Einschätzung des Weimarer Stadtzentrums; Polaritätsprofil zu den Eindrücken vom Stadtzentrum; Heimischfühlen in der

Stadt; Möglichkeit des Wegziehens; Faktoren für und gegen ein Gebundensein an die Stadt; Arbeitsort und benötigte Zeit dorthin; erlernter Beruf; höchste Qualifikation; ausgeübte Tätigkeit; Charakteristik der Arbeitstätigkeit; Schichtarbeit; Sicherheit des Arbeitsplatzes; mögliche Reaktionen im Falle von Arbeitslosigkeit

Untersuchungsgebiet: Weimar, Zentrum der Stadt, ein Neubaugebiet, eine frühe Werkssiedlung und ein Gründerzeitgebiet der Stadt Weimar (Thüringen)

Grundgesamtheit und Auswahl:

Zweistufige Auswahl von Personen im Alter von 18 bis 75 Jahren

Erhebungsverfahren:

Mündliche Befragung mit standardisiertem Fragebogen

Datensatz:

Anzahl der Einheiten: 487

Weitere Hinweise:

Diese Untersuchung stand in dem Gesamtbemühen, Weimar als Kulturstadt zu fördern. In Kooperation mit Sozialpsychologen der Friedrich-Schiller Universität in Jena wurden als Besonderheit dieser Studie gruppendynamische Prozesse in der Phasen der Feldforschung analysiert.[14]

Literatur

Michael Hader: Nun kann auch die soziologische Methodik zusammenwachsen, in: ZA-Information 27, November 1990.

Bernd Hunger: Soziologische Untersuchungen als Bestandteil der städtebaulichen Planung der Umgestaltung von Altbauwohngebieten, Hochschule für Architektur und Bauwesen Weimar, 1982, Dissertation A.

M. Rainer Lepsius: Zur Lage der Soziologie an den Universitäten der DDR, in: Kölner Zeitschrift für Soziologie und Sozialpsychologie, 2/1990.

Ekkehard Mochmann: Archivierung, Aufbereitung, Dokumentation und Bereitstellung von Umfragen aus der Sozialforschung der DDR, Projektantrag, Köln, Mai 1990.

Ekkehard Mochmann: Die unveröffentlichte Meinung: Zur Sicherung von Umfragen aus der DDR für Sekundäranalysen, in: ***Dieter Jauf mann, Ernst Kistler, Klaus Meier, Karl-Heinz Strech*** (Hrsg.): Empirische Sozialforschung im vereinten Deutschland, Campus-Verlag Frankfurt/New York, 1992.

14 Vgl. ***Christine Weiske,*** Vorankündigung der Untersuchung "Europäische Provinz Weimar", in: Weimar - Positionen zur Stadtentwicklung aus der Sicht der Hochschule für Architektur und Bauwesen, Wissenschaftliche Zeitschrift der Hochschule für Architektur und Bauwesen - A -, Weimar 37/1991, S.12.

Christine Weiske: Heimischfühlen in der Stadt - zur Wechselwirkung von Ortsverbundenheit und Migration, FSU Jena 1984, Dissertation A

Weimar - Positionen zur Stadtentwicklung aus der Sicht der Hochschule für Architektur und Bauwesen, Wissenschaftliche Zeitschrift der Hochschule für Architektur und Bauwesen - A -, Weimar 37/1991.

Zentralarchiv für empirische Sozialforschung (Hrsg.): Daten der empirischen Sozialforschung - Datenbestandskatalog, Campus Verlag Frankfurt/New York, 1991.

Studiendokumentationen in der Schriftenreihe der Hochschule für Architektur und Bauwesen:

Weimar 1990:
"Europäische Provinz Weimar", in der Schriftenreihe der Hochschule für Architektur und Bauwesen Weimar (unveröffentlicht).

Brandenburg 1987
"Die Stadt und ihre Bewohner" Teil I, II, 1989.

Dessau 1989
"Stadtentwicklung im Umbruch" - Situation, Ausblick 1989 - 90".

Eisenach 1985
"Soziologische Untersuchung Eisenach" Stadt- und Siedlungsgruppen 1985 Teil I und II.

Studie zur Bebauungsplanung und zum Ersatzneubau des innerstädtischen Umgestaltungsgebietes Eisenach, 1984, Weiterbildungsinstitut für Städtebau und Architektur.

Halle 1984
"Stadtentwicklung und Wohnmilieu Halle-Neustadt 1985".

Magdeburg 1986
"Stadtentwicklung und Wohnmilieu Magdeburg" Teil I, II, III, IV, 1987.

Rostock 1982
"Soziologische Untersuchung von Wohngebieten der Stadt Rostock", 1983.

Erfurt 1980
"Soziologische Probleme der Rekonstruktion in der nördlichen Innenstadt von Erfurt", 1982 (unveröffentlicht).

Jena 1988
"Lebensbilder aus Jena" - Versuch einer Dokumentation, Hg: Dietmar Ebert, Fotografie von Jürgen Hohmuth, Universitätsverlag Jena mit Unterstützung der Städte Tübingen und Jena, 1990.

Gotha 1981
"Soziologische Untersuchung zur Rekonstruktion der Gothaer Innenstadt", 1983.

Erste Kontakte zum spanischen Zentrum für Soziologische Forschung[1] CIS

Zu einem ersten Informationsaustausch zwischen einem spanischen Archiv für empirische Sozialforschung und dem Zentralarchiv kam es Ende April in Köln anläßlich des Besuchs der Soziologin Margarita Delgado, Direktorin der Datenbank des spanischen Zentrums für Soziologische Forschung, und des Volkswirtschaftlers Manuel Gago. Gesprochen wurde unter anderem über Möglichkeiten einer Zusammenarbeit im Rahmen des "Council of European Social Science Data Archives" (CESSDA) und der "International Federation of Data Organisations for the Social Sciences" (IFDO). Die Gäste aus Madrid interessierten sich insbesondere für die Organisation der Datenbank im Zentralarchiv und die Möglichkeiten der Fragenrecherche.

Die Arbeit des spanischen Zentrums reicht von der Planung und Durchführung sozial-wissenschaftlicher Umfragen über deren Archivierung bis zur Veröffentlichung der Analysen in eigenen Publikationen, einer monographischen Reihe oder der Vierteljahreszeitschrift "Revista Espafiola de Investigaciones Sociológicas" (REIS). Auf Anfrage von Politikern, Wissenschaftlern oder Journalisten werden zudem jeweils gesonderte Ergebnisberichte zusammengestellt. Darüber hinaus sollen künftig verstärkt die Datensätze selbst für Sekundäranalysen zur Verfügung gestellt werden. Die juristischen Voraussetzungen für einen Vertrieb auch außerhalb Spaniens sind dabei allerdings noch nicht geklärt.

Die CIS Datenbank umfaßt zur Zeit fast 900 eigene oder vom Zentrum mitfinanzierte Studien, darunter ein monatliches Politbarometer seit Juli 1979 sowie regelmäßige Vor- und Nachwahlstudien auf nationaler und regionaler Ebene. Zu den frühesten archivierten Untersuchungen zählen drei Befragungen ausländischer Touristen 1964,1965 und 1966. Umfragen anderer, insbesondere auch kommerzieller Institute, werden nicht archiviert. Der ausführliche Datenbestandskatalog kann im Zentralarchiv eingesehen werden.

Präsident des Zentrums für Soziologische Forschung ist seit gut einem Jahr der Sozialwissenschaftler Professor Don Joaquín Arango. Als staatliche Behörde ist es dem "Ministerium für Beziehungen zu Parlament und Regierungssekretariat" (Ministerio de Relaciones con las Cortes y de la Secretaria del Gobierno) zugeordnet.

---

1 Centro de Investigaciones Sociológicas, Montalbán 8, E-28014 Madrid
(Tel 0034-1-5318975 / Fax 0034-1-5318131)

Zur aktuellen Situation von Glauben und Kirche im vereinigten Deutschland: Eine Analyse der Basisumfrage 1991

von Michael Terwey

Abstract

Gegenstand der Untersuchung ist die Lage von Kirche und Glauben in den alten und neuen Bundesländern des vereinigten Deutschlands. Kirchliche Partizipation, Gottesglauben und Glauben an weitere vorwiegend christlich geprägte Glaubensvorstellungen werden in den neuen Bundesländern sehr viel seltener angegeben als in den alten. Geringere Unterschiede ergeben sich hinsichtlich der Verbreitung von Glauben an andere paranormale Phänomene. Abschließend wird in einer multivariaten Analyse überprüft, welche Charakteristika gottesungläubige und gottesgläubige Befragte auszeichnen.

The article focuses on religious affiliations and beliefs in the western and the eastern parts of today's Germany. Data support predictions that bygone socialism in the newfederal states has led to a remarkable diminution of traditional religious commitments and beliefs, whereas socialism is still regarded as a good idea by the majority of eastern respondents. Both parts of Germany show smaller differences in the frequencies of superstition. Finally, the question which people still have faith in God is investigated by multivariate analysis.

In Deutschland hatten sich vor der Wiedervereinigung sehr verschiedene Gesellschaftssysteme etabliert. Im Westen führte die soziale Rekonstruktion nach dem Zweiten Weltkrieg zur Entfaltung einer pluralistisch demokratischen Gesellschaft mit einer prosperierenden freien Marktwirtschaft. In der ehemaligen DDR hatte dagegen Einflußnahme der sowjetischen Siegermacht die Gründung eines nominell demokratischen, staatssozialistischen Systems zur Folge, dessen politische und wirtschaftliche Entwicklung im Zusammenbruch des Systems mit staatlicher Auflösung mündete. Über die verschiedenen Auswirkungen der politischen und organisatorischen Unterschiede beider deutscher Staaten lagen insbesondere hinsichtlich der neuen Bundesländer bisher nur wenig differenzierte empirische Daten auf Individualebene vor, weil in der ehemaligen DDR eine von Staatseinflüssen freie Umfrageforschung kaum durchführbar war.

Als Folge der Wiedervereinigung ist vornehmlich in den neuen Bundesländern ein erheblicher sozialer Wandel zu erwarten. Im Interesse der möglichst raschen Erfassung von sozial-

wissenschaftlichen Daten über die besondere historische Situation wurde deshalb mit der **"Basisumfrage 1991"** eine Sondererhebung im Programm der "Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften" (ALLBUS) durchgeführt.[1] Diese Daten eröffnen einer breiten wissenschaftlichen Öffentlichkeit die Chance, soziale Strukturen und Einstellungen in den alten und neuen Bundesländern kurz nach der Wiedervereinigung zu vergleichen. Aus dem großen Spektrum der möglichen Fragestellungen wird hier die Analyse von zur Zeit existierenden Differenzen und Gemeinsamkeiten in religiösen Orientierungen aufgegriffen.

Kirchliche Partizipation in den alten und neuen Bundesländern

Die Teilung Deutschlands hatte auch für Religion und Kirche in Ost und West weitreichende Folgen. In der alten Bundesrepublik spielten traditionelle Religiosität und Kirchen eine gesellschaftlich überwiegend wohl akzeptierte, wenn auch in ihrer Bedeutung etwas zurückgehende Rolle. Tendenzen, die in modernen westlichen Gesellschaften den überlieferten Glaubensrichtungen zuwiderlaufen, können mit Entwertung überkommener Mythen, Säkularisierung, Entkirchlichung und Individualisierung der Lebensstile umschrieben werden. In der ehemaligen DDR stießen Kirchen und christlicher Glauben darüberhinaus aus grundsätzlichen politischen Gründen auf ungünstige Voraussetzungen. Die ursprüngliche marxistische Ideologie zielte darauf ab, Religion als "Opium des Volkes" mindestens langfristig zu eliminieren. Im Zuge einer pragmatischen Politik wurde in der DDR dennoch formale Religionsfreiheit gewährt. Es entwickelte sich auch eine zurückhaltende Akzeptanz von Kirche und Religion unter dem Konzept der alle konstruktiven gesellschaftlichen Kräfte vertretenden "Sozialistischen Einheitspartei", doch war Wertschätzung für die "bessere" sozialistische Gesellschaft eine Voraussetzung für die Zusammenarbeit *(Klient* 1989). Grundsätzlich kooperierte der Staat in geringerem Ausmaß mit den Kirchen als im Westen. Als Beispiele nennt *Weichert* 1990: 306 die fehlende staatliche Hilfe beim Einzug von Kirchensteuern und das fehlende Recht der Kirchen, Religionsunterricht in Schulräumen zu erteilen. Zudem entwickelte sich im Sozialismus ein eigenes Instrumentarium nicht-christlicher Riten und Sozialisationsmechanismen. Unter diesen Voraussetzungen ist eine starke Verringerung der **Kirchenmitgliedschaft** zu vermuten.

---

1 Die Basisumfrage 1991 wurde von der Deutschen Forschungsgemeinschaft **(DFG)** finanziert (DFG-Projekt Mo 331/2-1). Ihre Konzeption, Durchführung und Aufbereitung wurde von den am ALLBUS beteiligten Institutionen und Personen getragen. DFG-Antragsteller für die Basisumfrage waren ***Peter Ph. Mohler, Michael Braun, Erwin K. Scheuch*** und ***Michael Häder.*** Die Datendokumentation und die Daten sind beim Zentralarchiv für empirische Sozialforschung erhältlich.

Tabelle 1: Welcher Konfessionsgemeinschaft gehören Sie an?

(Deutsche Befragte, Berechnung nach der Basisumfrage 1991)

	Alte Bundesländer %	Neue Bundesländer %
Der römisch-katholischen Kirche	41,9	5,5
Der evangelischen Kirche (ohne Freikirchen)	45,0	27,0
Einer evangelischen Freikirche	1,3	1,8
Einer anderen christlichen Religionsgemeinschaft	1,2	0,8
Einer anderen nicht christlichen Religionsgemeinschaft	0,1	0,1
Keiner Religionsgemeinschaft	10,6	64,7
n = 100 %	1477	1537

Cramer's V = .60

Nach Tabelle 1 können in den neuen Bundesländern für 1991 noch 5,5 Prozent der Befragten in der katholischen Kirche und 28,8 Prozent in den evangelischen Kirchen (einschließlich Freikirchen) festgestellt werden, wogegen der Anteil der Konfessionslosen mit 64,7 Prozent sehr hoch ist.[2] Im Jahre 1950 waren dagegen auf dem Gebiet der ehemaligen DDR noch 11 Prozent der Bevölkerung katholisch, 80,6 Prozent evangelisch und nur 7,6 Prozent konfessionslos *(Weichert* 1990: 308). Die von *Weichert* für die DDR von 1989 geschätzten Mitgliedschaftszahlen entsprechen annähernd den Daten aus der Basisumfrage. Insgesamt ist somit im Osten ein enormer Mitgliederverlust für die Kirchen feststellbar, wovon die katholische Kirche - relativ gesehen - etwas weniger betroffen war. In den alten Bundesländern sind nach der Basisumfrage 1991 nur 10,6 Prozent aller Befragten ohne Konfessionsmitgliedschaft, während die großen Volkskirchen fast 90 Prozent aller Befragten zu annähernd gleichen Teilen auf sich vereinen. Als weiterer Basisindikator für die kirchliche Integration kann die **Kirchgangshäufigkeit** gesehen werden.

2 Die in der Basisumfrage ebenfalls interviewten deutschsprachigen Ausländer werden bei den hier vorgestellten Analysen nicht berücksichtigt. Insgesamt konnten nur 41 Ausländerinterviews realisiert werden, davon 4 in den neuen Bundesländern. Eine gesonderte Auswertung dieser heterogenen Untergruppe ist bei diesen kleinen Fallzahlen für unsere Fragestellung kaum sinnvoll.

Tabelle 2: Wie oft gehen Sie im allgemeinen in die Kirche

(Deutsche Befragte, Berechnung nach der Basisumfrage 1991)

	Alte Bundesländer %	Neue Bundesländer %
Mehr als einmal die Woche	3,1	0,9
Einmal in der Woche	11,6	2,6
Ein- bis dreimal im Monat	9,7	3,3
Mehrmals im Jahr	21,9	11,3
Seltener	32,7	21,7
Nie	21,0	60,2
n = 100 %	1472	1528

Cramer's V = .42

Die naheliegende Konsequenz der weniger verbreiteten Kirchenmitgliedschaft in den neuen Ländern ist eine deutlich geringere Kirchgangshäufigkeit. Hinzu kommt für die Beurteilung der Spaltendifferenzen in Tabelle 2 aber auch, daß der Anteil von Protestanten, die im allgemeinen seltener als Katholiken die Kirche besuchen, im Osten höher ist. Somit wären unter Umständen noch deutlichere Unterschiede für die Kirchgangshäufigkeit zu erwarten gewesen. Innerhalb der verschiedenen Konfessionsgruppen reduzieren sich die aus Tabelle 2 global ablesbaren Unterschiede schließlich erheblich. So weit sich Häufigkeitsaussagen aus der zur Messung verwendeten Skala ablesen lassen, gehen die Katholiken im Osten etwas häufiger und die Lutheraner ähnlich oft zur Kirche wie die Vergleichsgruppen im Westen. Die wenigen Angehörigen der evangelischen Freikirchen und die Konfessionslosen besuchen in den alten Bundesländern die Kirche geringfügig häufiger.

Fraglich ist, inwieweit der Rückgang in der formalen Kirchenmitgliedschaft auch Auswirkungen auf die Verbreitung inhaltlicher Glaubensvorstellungen gehabt hat. Denkbar wäre, daß ein Teil der Kirchenaustritte eine äußerliche Konzession an das sozialistische System darstellte und der Vermeidung von Chancennachteilen für Ausbildung und berufliche Karriere diente (vgl. *Nowossadeck* 1991: 96). Zur Klärung wird nachfolgend die Akzeptanz verschiedener Glaubensvorstellungen untersucht.

Glauben, Aberglauben und Bekenntnis zum Sozialismus

Die Basisumfrage 1991 enthält neben den Fragen zur konfessionellen Mitgliedschaft und zur' Kirchgangshäufigkeit auch eine größere Anzahl detaillierter Fragen zur Religion, die Bestandteil der schriftlichen Zusatzerhebung im Rahmen des ISSP sind[3]. Zentral für die inhaltliche Akzeptanz der meisten Religionsformen im engeren Sinne ist das Vorhandensein eines **Gottesglaubens.** Dies gilt auch für religiöse Personen, die nicht in den großen Volkskirchen integriert sind. Von den verschiedenen möglichen Fragen zur Erfassung von Gottesglauben wurde für die folgende Untersuchung ein Item ausgewählt, in dem die Frage nach Gott allgemein formuliert und ohne spezifizierende Zusätze, die aus einem besonderen christlichen Religionsverständis herrühren, gestellt wurde. Zusätzlich wird dabei auch die Entwicklung des Gottesglaubens retrospektiv erfaßt.

Tabelle 3: Welche dieser Aussagen beschreibt Ihren Glauben an Gott am Besten?

(Deutsche Befragte, Berechnung nach der ISSP-Zusatzbefragung in der Basisumfrage 1991)

	Alte Bundesländer %	Neue Bundesländer %
Ich glaube nicht an Gott und habe niemals an ihn geglaubt.	9,6	50,7
Ich glaube nicht an Gott, habe aber früher an ihn geglaubt.	23,4	24,9
Ich glaube an Gott, habe aber früher nicht an ihn geglaubt.	9,0	5,0
Ich glaube an Gott und habe immer an ihn geglaubt.	57,9	19,4
n = 100 %	1260	1453

Cramer's V = .49

3 Im "International Social Survey Programme" (ISSP) werden Befragungen durchgeführt, die internationale Vergleiche zu verschiedenen Schwerpunktthemen ermöglichen. Die Realisierung des ISSP 1991 ist in wenigstens vierzehn Ländern gesichert. Abweichend vom mündlichen Hauptinterview der Basisumfrage fand die deutsche ISSP-Erhebung 1991 wie in anderen Ländern mit einem schriftlichen "drop-off '-Fragebogen statt. Insgesamt konnten 2832 schriftliche Interviews realisiert werden. Die Daten aus allen am ISSP 1991 beteiligten Ländern werden vom Zentralarchiv insgesamt gesondert aufbereitet und dokumentiert. Dieser international vergleichende Datensatz wird nach Abschluß der Integrations- und Aufbereitungsarbeiten beim Zentralarchiv als gesonderte Studie erhältlich sein.

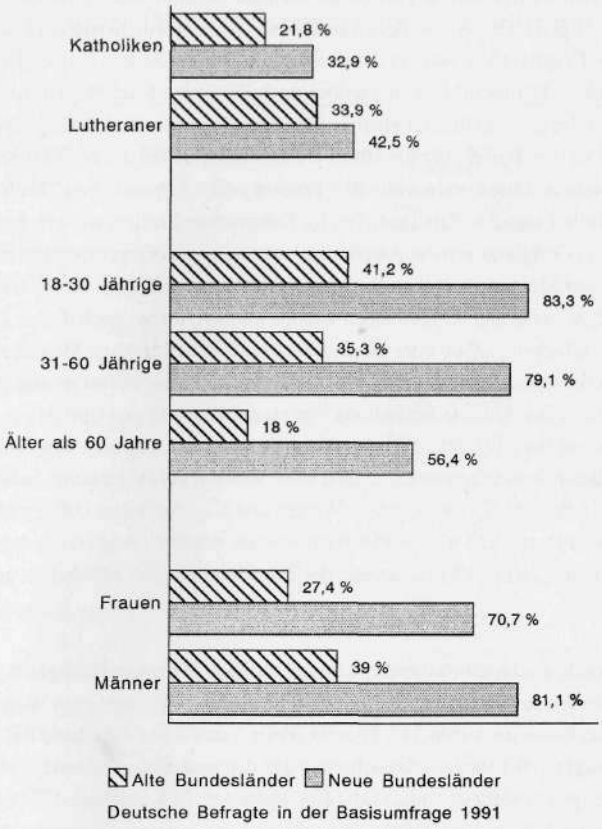
Tabelle 3 zufolge bestehen nicht nur hinsichtlich des Ausmaßes kirchlicher Orientierung deutliche Unterschiede zwischen den alten und neuen Bundesländern. Auch die Akzeptanz des Gottesglaubens differiert in hohem Ausmaß. Während im Westen nur eine Minderheit von 9,6 Prozent angab, nie an Gott geglaubt zu haben, stimmten im Osten 50,7 Prozent dieser Fragekategorie zu. Der Anteil derer, die sich im Lebensverlauf vom Gottesglauben abwandten, ist dagegen im Osten nur geringfügig höher als im Westen (24,9 Prozent vs. 23,4 Prozent) [4]. Deutlich weniger Befragte gaben dagegen den entgegengesetzten Wechsel **hin** zum Gottesglauben an, wobei der Anteil in den neuen Bundesländern (5 Prozent) nunmehr geringer ausfällt als in den alten (9 Prozent). Schließlich stehen den 57,9 Prozent stets Gottesgläubiger im Westen nur noch 19,4 Prozent im Osten gegenüber. Es war bei weitem nicht nur die Vermeidung von Chancennachteilen im Sozialismus, die zu einer äußerlich gesehen geringeren kirchlichen Partizipation geführt hat. Vielmehr ist auch hinsichtlich des zentralen Glaubens an Gott eine weit verbreitete inhaltliche Abwendung unter dem Einfluß konkurrierender Ideologien zu vermuten.

Für die weitere Untersuchung lassen wir die zusätzlichen Informationen darüber, ob früher eine Änderung im Gottesglauben stattgefunden hat, außer acht und untersuchen nur noch die grundlegende Dichotomie: **Ungläubige** (vgl. Zeile 1 und 2 in Tabelle 3) vs. **Gläubige** (vgl. Zeile 3 und 4 in Tabelle 3).

Hinsichtlich der grundlegenden Differenzierung zwischen Ungläubigen und Gläubigen sind auch innerhalb der großen Konfessionen deutliche Unterschiede festzustellen: 32,9 Prozent der Katholiken und 42,5 Prozent der Lutheraner in den neuen Bundesländern glauben nicht an Gott (siehe Schaubild 1). In den alten Bundesländern sind es dagegen "nur" 21,8 Prozent und 33,9 Prozent. Das für die großen Kirchen bestehende Problem einer zwar formal gegebenen Mitgliedschaft, die aber keinen Rückhalt im zentralen Glaubensgrundsatz hat, ist nach diesem Zahlenvergleich auf dem Gebiet der ehemaligen DDR eher noch größer. Auch für die weiteren demographischen Vergleichsgruppen in Schaubild 1 läßt sich ablesen, daß die

4 Vielleicht wäre in den neuen Bundesländern ein höherer Anteil an Befragten zu erwarten gewesen, die sich vom Gottesglauben abgewandt haben. Die Möglichkeit einer mangelhaften Rückerinnerung an einen früheren Gottesglauben unter heutigen Atheisten ist in Betracht zu ziehen. In diesem Zusammenhang ist zu berücksichtigen, daß auch im Nationalsozialismus eine säkulare Quasi-Religion propagiert wurde, die mit den traditionellen Religionen und Kirchen konkurrierte bzw. ihnen oft feindlich gegenüberstand. Fraglich ist, ob das mit anderen, heute meist negativ besetzten Ideologieelementen vermischte Gottesverständnis, das in einer nationalsozialistisch geprägten Sozialisation ggfs. vermittelt wurde, heute noch als solches verstanden wird. Der schnelle Übergang vom Nationalsozialismus zum Sozialismus hat im Osten die Chancen einer Sozialisation im Sinne des Christentums langfristig beeinträchtigt. In der alten Bundesrepublik waren die ersten Jahre nach dem Zusammenbruch des Nationalsozialismus durch niedrige Kirchenaustrittszahlen gekennzeichnet, während im Deutschen Reich kurz vor 1939 und in den ersten Kriegsjahren vergleichsweise viele Personen zumindest die evangelische Kirche verließen (vgl. Statistisches Bundesamt; *Hanselmann* et al. 1984: 35 f.).

Schaubild 1: Anteile von Gottesungläubigen in verschiedenen Bevölkerungsgruppen



Atheistenanteile im Westen durchgängig niedriger ausfallen. Unter Außerachtlassung dieser generellen Relation bestehen aber darin Gemeinsamkeiten, daß in beiden Teilen Deutschlands ältere Menschen öfter als jüngere und Frauen öfter als Männer an Gott glauben.

Tabelle 4 ergänzt das Bild geringerer Akzeptanz von traditioneller Religiosität an Hand von fünf Glaubensfragen aus dem weiteren Transzendenzbereich. Hinsichtlich jeder der fünf verschiedenen Glaubensvorstellungen finden wir wiederum weniger zustimmende Antworten auf dem Gebiet der ehemaligen DDR. Am stärksten sind diese Unterschiede hinsichtlich des **Lebens nach dem Tode** (Cramer's V = .44), an das in den alten Bundesländern vergleichsweise viele Befragte glauben. Darüberhinaus sind jedoch auch Parallelen festzustellen. An ein Leben nach dem Tod oder an den Himmel glauben in beiden Teilen Deutschlands jeweils relativ mehr Befragte als an den **Teufel** oder die **Hölle.** Dem Gottesglauben mit 66,9 Prozent im Westen und 24,4 Prozent im Osten, stehen nur 20,1 Prozent und 5,7 Prozent von "Teufelsgläubigen" gegenüber (Zeile 1 und 2 in Tabelle 4, Teil b). Entsprechend zeigen weitere Aufgliederungen, daß auch bei Gottgläubigen die Annahme eines negativen "Gegenspielers" zu Gott insgesamt nur bei einer Minderheit vorhanden ist[5]. Glaubensvorstellungen, die als individuell wünschenswert bzw. heilbringend verstanden werden können, finden nach diesen Daten mehr Akzeptanz. Der Glauben an Gott oder an ein Leben nach dem Tod kann Menschen in einem positiven Existenzverständnis bestärken. Von der Hölle oder dem Teufel ist dagegen direkt kein Heil zu erwarten. Die tradierte Vorstellung von einem allmächtigen Gott des Heils, der sich mit Liebe und Barmherzigkeit für den Menschen einsetzt, zugleich den Teufel aber als Versucher oder Bestrafenden zumindest zuläßt, stößt vielfach auf Ablehnung. In dieses Bild paßt, daß nur 11,1 Prozent der Befragten im Westen und 6,3 Prozent im Osten meinen, "Die Bibel ist das Wort Gottes und muß wörtlich genommen werden". Auch unter den Personen, die an Gott glauben, ist diese Einschätzung der Bibel nur bei einer Minderheit gegeben.

Bemerkenswert ist in Tabelle 4 schließlich die vergleichsweise hohe **Wundergläubigkeit,** die auch in den neuen Bundesländern besteht. Während im Westen 51,2 Prozent zum Wunderglauben neigen, sind es im Osten immerhin 34,9 Prozent (Zeile 1 und 2 aus Tabelle 4, Teil e). Weitere Berechnungen zeigen: 26,4 Prozent der Atheisten aus den neuen Bundesländern akzeptieren Wunder.als wenigstens wahrscheinlich (alte Bundesländer: 32,8 Prozent). Gibt es somit auch in dem Bereich der ehemaligen DDR einen vergleichsweise beachtlichen Bevölkerungsanteil, in dem losgelöst von zentralen christlichen Vorstellungen an übernatürliche Phänomene geglaubt wird?

---

5 Aus den jeweils beachtlichen Befragtengruppen, die auf derartige Statements mit "Kann ich nicht sagen" geantwortet haben, könnte unter Umständen noch ein Teil zu Transzendenzaussagen der hier vorgestellten Art neigen. Sei es, daß sie zwar partiell zu den Glaubensvorstellungen tendieren, sich aber keine dezidierte Meinung gebildet haben, sei es, daß beispielsweise zum Teufelsglauben neigende Menschen sich scheuen, darüber eine explizite Angabe zu machen. Selbst unter dieser Annahme gilt jedoch weiterhin, daß der Gottesglauben stärker verbreitet ist als der Teufelsglauben.

Tabelle 4: Akzeptanz von fünf Formen traditioneller Transzendenzvorstellungen
(Deutsche Befragte, Berechnungen nach der ISSP-Zusatzbefragung in der Basisumfrage 1991)

		Alte Bundesländer %	Neue Bundesländer %
a) Glauben Sie, daß es ein Leben nach dem Tod gibt:	Ja, ganz sicher	20,3	5,2
	Ja, wahrscheinlich	24,1	7,0
	Nein, wahrscheinlich nicht	16,7	12,0
	Nein, sicher nicht	22,4	62,1
	Kann ich nicht sagen	16,5	13,6
n = 100 %		1304	1471
	Cramer's V = .44		
b) Glauben Sie, daß es den Teufel gibt:	Ja, ganz sicher	7,7	3,2
	Ja, wahrscheinlich	12,4	2,5
	Nein, wahrscheinlich nicht	14,7	8,2
	Nein, sicher nicht	47,8	74,7
	Kann ich nicht sagen	17,4	11,3
n = 100 %		1283	1445
	Cramer's V = .30		
c) Glauben Sie, daß es den Himmel gibt:	Ja, ganz sicher	14,8	8,9
	Ja, wahrscheinlich	20,2	8,5
	Nein, wahrscheinlich nicht	14,8	8,6
	Nein, sicher nicht	32,8	61,1
	Kann ich nicht sagen	17,5	13,0
n = 100 %		1288	1443
	Cramer's V = .29		
d) Glauben Sie, daß es die Hölle gibt:	Ja, ganz sicher	7,4	2,3
	Ja, wahrscheinlich	12,9	3,4
	Nein, wahrscheinlich nicht	16,5	7,4
	Nein, sicher nicht	45,1	73,9
	Kann ich nicht sagen	18,0	13,0
n = 100 %		1282	1442
	Cramer's V = .31		
e) Glauben Sie, daß es Wunder gibt:	Ja, ganz sicher	19,6	10,3
	Ja, wahrscheinlich	31,6	24,6
	Nein, wahrscheinlich nicht	17,1	14,6
	Nein, sicher nicht	19,2	38,2
	Kann ich nicht sagen	12,6	12,3
n = 100 %		1294	1451
	Cramer's V = .23		

Tabelle 5 enthält weitere Fragen zu Glaubensvorstellungen, die oft zum Bereich des "Aberglaubens" gerechnet werden. Die Werte für Cramer's V weisen in Tabelle 5 mit .03 bis .12 generell deutlich geringere Unterschiede zwischen den beiden Teilstichproben aus als in Tabelle 4 (.23 bis .44). **Glücksbringer, Wahrsager, Wunderheiler** und **Sternzeichen** werden jeweils von ca. 20 Prozent bis 30 Prozent der Befragten für wenigstens wahrscheinlich existent oder wirksam gehalten. Rationale Aufklärung, Erkenntnisse der Wissenschaft oder sozialistische Ideologie haben in diesem Sinne offenbar relevante Bevölkerungsteile nicht von derlei paranormalen Vorstellungen abbringen können. Etwas geringere Akzeptanz haben in den neuen Bundesländern lediglich die Vorstellungen von Wahrsagern und astrologischen Einflüssen. Während die Daten von *Noelle-Neumann/Piel* 1983: 123 noch teilweise die Möglichkeit eines schwachen Rückgangs astrologischer Überzeugungen in der Bundesrepublik zwischen 1950 und 1982 andeuteten, erwecken die neueren Daten eher den Eindruck, daß wenigstens im Westen ein relativ stabiler Bevölkerungsanteil von ca. 25 Prozent "sternengläubig" ist.

Als säkulare Alternative zu der traditionellen Religiosität kann der Glauben an eine politische Idee gelten. Wirksamkeitsverlust der Kirchen kann zu einer quasi-religiösen Aufladung anderer gesellschaftlicher Teilsysteme führen - nicht zuletzt im politischen Bereich *{Schmidtchen* 1978: 26). Wie bereits frühere Untersuchungen in der Europäischen Gemeinschaft ansatzweise gezeigt haben, ist die Einstellung zu Gott nicht unabhängig von einer Grundeinstellung zur Gesellschaft, die sie sich in der Selbsteinstufung auf der Links-Rechts-Skala zeigt (EG 1985: 49, 51). Allerdings ist die komplexe inhaltliche Bedeutung einer globalen Links-Rechts-Einstufung je nach Land und Untersuchungszeitraum relativ schwierig einzuordnen. Uns interessiert an dieser Stelle primär, wie sich die Einstellung zur sozialistischen Ideologie als Alternative zur traditionellen Religiosität auswirkt. In der Basisumfrage finden wir die Ansicht der Befragten darüber, ob der **Sozialismus** im Grunde eine gute Idee ist, die nur schlecht ausgeführt wurde.

Die Daten in Tabelle 6 zeigen für die alten Bundesländer eine Mehrheit von 54,4 Prozent, die zur Idee des Sozialismus eher negativ eingestellt ist (Zeile 3 und 4 zusammengefaßt). Ganz im Gegensatz dazu steht in den neuen Bundesländern mit 69,6 Prozent weiterhin eine deutliche Mehrheit mit einer positiven Beurteilung (Zeile 1 und 2 zusammengefaßt). Die Erwartung, daß sich dort nach den Erfahrungen mit dem real existierenden Sozialismus eine negative Grundeinstellung zu dieser Idee durchgesetzt hat, kann zur Zeit noch nicht bestätigt werden. Darüberhinaus stufen sich die Bürger in den neuen Bundesländern auch im Vergleich zu denen in den alten Bundesländern häufiger als links ein.

Tabelle 5: Glauben an paranormale Phänomene außerhalb des engeren Bereichs
(Deutsche Befragte, Berechnungen nach ISSP-Zusatzbefragung in der Basisumfrage 1991)

	Alte Bundesländer %	Neue Bundesländer %
a) Glücksbringer bringen manchmal tatsächlich Glück:		
Stimmt sicher	4,3	5,3
Stimmt wahrscheinlich	22,7	24,3
Stimmt wahrscheinlich nicht	23,5	21,7
Stimmt sicher nicht	38,4	35,2
Kann ich nicht sagen	11,1	13,5
n = 100 %	1313	1480
Cramer's V = .05		
b) Es gibt Wahrsager, die die Zukunft wirklich vorhersehen können:		
Stimmt sicher	5,8	2,9
Stimmt wahrscheinlich	22,1	16,1
Stimmt wahrscheinlich nicht	23,6	22,8
Stimmt sicher nicht	38,2	44,3
Kann ich nicht sagen	10,4	13,9
n = 100 %	1313	1480
Cramer's V = .12		
c) Manche Wunderheiler verfügen wirklich über übernatürliche Kräfte:		
Stimmt sicher	6,8	7,4
Stimmt wahrscheinlich	25,7	26,3
Stimmt wahrscheinlich nicht	22,9	20,3
Stimmt sicher nicht	31,8	32,0
Kann ich nicht sagen	12,8	14,1
n = 100 %	1313	1480
Cramer's V = .03		
d) Das Sternzeichen bzw. das Geburtshoroskop eines Menschen hat einen Einfluß auf den Verlauf seines Lebens:		
Stimmt sicher	5,9	3,2
Stimmt wahrscheinlich	21,9	17,3
Stimmt wahrscheinlich nicht	24,5	24,4
Stimmt sicher nicht	35,8	41,9
Kann ich nicht sagen	12,0	13,2
n = 100 %	1316	1482
Cramer's V = .10		

Tabelle 6: Der Sozialismus ist im Grunde eine gute Idee, die nur schlecht ausgeführt wurde (Deutsche Befragte, Berechnung nach der Basisumfrage 1991)

	Alte Bundesländer %	Neue Bundesländer %
Stimme voll zu	11,6	34,7
Stimme eher zu	24,1	34,9
Stimme eher nicht zu	25,0	13,2
Stimme überhaupt nicht zu	29,4	9,3
Weiß nicht	9,8	7,9
n = 100 %	1474	1537

Cramer's V = .37

Um ansatzweise zu ermitteln, wie sich die Beziehungen zwischen verschiedenen in diesem Abschnitt vorgestellten Glaubensformen zueinander verhalten, sollen diese Variablen im Sinne einer heuristischen Exploration mit der Faktorenanalyse untersucht werden. Diese Analysen beschränken sich auf Befragte, welche zu den verschiedenen Glaubensbezügen eine dezidierte Meinung vertreten haben (Ausschluß der Befragten mit der Angabe 'Kann ich nicht sagen'). Gewählt wurde die Methode der Hauptkomponentenanalyse mit Varimax-Rotation und paarweisem Ausschluß fehlender Werte. Bei anderen Optionen zur Behandlung fehlender Werte wurden analoge Ergebnisse erzielt.

Die Faktorenanalyse ergibt ein sehr ähnliches Bild in beiden Teilen von Tabelle 7. Die im weiteren Sinne traditionellen Transzendenzvorstellungen laden mit Werten über .50 auf dem ersten Faktor. Die paranormalen Annahmen laden auf dem zweiten Faktor und Einschätzung der Idee des Sozialismus ebenso wie die Links-Rechts-Selbsteinstufung auf einem dritten. Ein vierter Faktor ist zur Beschreibung dieser Daten nach dem Kriterium Eigenwert >1 nicht notwendig. Vergleichsweise schwach ist jeweils die Ladung des Glaubens an Wunder auf dem ersten Faktor, verbunden mit einer etwas gehobenen Nebenladung auf dem zweiten - eine Besonderheit, die im Osten noch etwas stärker auftritt.

Tabelle 7: Glaubensvorstellungen in einer explorativen Faktorenanalyse
(Deutsche Befragte, Berechnung nach der Basisumfrage 1991)

A) Alte Bundesländer:

	Faktor 1	Faktor 2	Faktor 3	Kommunalität
Glauben an ein Leben nach dem Tod	.803	.051	.002	.648
Glauben an den Teufel	.889	.012	-.105	.802
Glauben an den Himmel	.868	.016	-.119	.767
Glauben an die Hölle	.897	.001	-.105	.816
Glauben an Wunder	.666	.326	-.016	.550
Glauben an Glücksbringer	.011	.743	-.073	.557
Glauben an Wahrsager	.138	.835	.051	.719
Glauben an Wunderheiler	.185	.799	.111	.684
Glauben an Einfluß des Sternzeichens	-.062	.766	-.044	.592
Glauben an die Idee des Sozialismus	-.032	.026	.796	.635
Links-Rechts-Selbsteinstufung	-.151	-.014	.746	.580

	Eigenwert	Erklärte Variation	Kumulierte Variationserklärung
Faktor 1	3.792	34.5	34.5
Faktor 2	2.394	21.8	56.2
Faktor 3	1.164	10.6	66.8

B) Neue Bundesländer:

	Faktor 1	Faktor 2	Faktor 3	Kommunalität
Glauben an ein Leben nach dem Tod	.820	.060	-.117	.690
Glauben an den Teufel	.909	-.013	-.044	.828
Glauben an den Himmel	.777	.164	-.083	.638
Glauben an die Hölle	.907	.018	-.036	.824
Glauben an Wunder	.518	.464	-.087	.492
Glauben an Glücksbringer	.004	.745	-.028	.555
Glauben an Wahrsager	.089	.812	-.025	.668
Glauben an Wunderheiler	.091	.767	.038	.597
Glauben an Einfluß des Sternzeichens	.061	.736	-.060	.549
Glauben an die Idee des Sozialismus	-.089	.067	.822	.688
Links-Rechts-Selbsteinstufung	-.098	-.131	.801	.669

	Eigenwert	Erklärte Variation	Kumulierte Variationserklärung
Faktor 1	3.701	33.6	33.6
Faktor 2	2.237	20.3	54.0
Faktor 3	1.260	11.5	65.4

Die Ergebnisse dieser Analyse sollen in erster Linie Anregungen vermitteln, wie die Beziehungen der Daten untereinander tendenziell einzuordnen und weiter zu analysieren sind. Am Rande bemerkt, ergeben sich beispielsweise die höchsten Ladungen auf dem ersten Faktor für den Teufels- und Höllenglauben. Wird das Modell durch zusätzliche Indikatoren des Gottesglaubens oder der Kirchlichkeit erweitert, so ergibt sich ein vierter Faktor für diese letztgenannten Items mit Nebenladungen für Glauben an ein Leben nach dem Tod und Glauben an den Himmel. Den vorläufigen Ergebnissen der explorativen Faktorenanalysen zufolge ist eine ähnliche Beziehungsstruktur zwischen den Glaubensvorstellungen bei Befragten in den neuen und alten Bundesländern festzustellen, obgleich bei den meisten Glaubensaussagen die Verbreitung im Osten sehr viel geringer ist.

Multivariate Untersuchung zum Gottesglauben in Ost und West

Die bisherigen Häufigkeitsübersichten haben globale Unterschiede und Gemeinsamkeiten aufgezeigt, die zwischen beiden Teilen Deutschlands in Bezug auf Kirchlichkeit und verschiedene Glaubensvorstellungen bestehen. Sie eröffnen eine Reihe von Fragestellungen, die letztlich nur mit komplexen multivariaten Verfahren in übersichtlicher Form beantwortet werden können. Dies betrifft beispielsweise die grundsätzliche Berücksichtigung des differierenden konfessionellen Hintergrundes in den neuen und alten Bundesländern, auf dessen Bedeutung bisher bereits stellenweise verwiesen wurde. Auch die in Schaubild 1 angedeutete bivariate Bedeutung von Alter und Geschlechtszugehörigkeit ist weiter zu hinterfragen. Abschließend soll daraufhin Gottesglauben in den alten und neuen Bundesländern mit dem multivariaten Verfahren der logistischen Regression untersucht werden. Eingesetzt wurde vorwiegend das von Steffen M. *Kühnel* erstellte SPSS-Matrix-Programm (vgl. *Kühnel* 1990). Als abhängige Variable verwenden wir wiederum die Dichotomie "gottesungläubig vs. gottesgläubig". Die Konstruktion der in Tabelle 8 vorgestellten Modelle wurde in mehreren Schritten vorbereitet. Im ersten Schritt wurden allgemeine demographische Merkmale überprüft. Prädiktoren, welche sich in wenigstens einem Untersuchungsbereich als relevant erwiesen, wurden beibehalten. Im zweiten Schritt wurde das Modell mit Indikatoren kirchlicher Partizipation analog erweitert. Abschließend wurden mehrere Einstellungs- und Glaubensfragen überprüft.

Tabelle 8: Erklärungsfaktoren des Gottesglaubens in der logistischen Regression
(Deutsche Befragte, Berechnung nach der Basisumfrage 1991)

A) Alte Bundesländer

	χ^2	p	β	Effekt	Effekt$_S$	P^2_{brutto}	P^2_{netto}
Alter	23.44	.000	.026	1.026	1.554	2.3	1.7
Geschlecht = Mann	4.16	.041	-.341	.711	1/1.186	1.2	.3
Wohnortsgröße	10.20	.001	-.091	.913	1/1.331	3.5	.7
Konfessionslosigkeit	11.04	.001	-.889	.411	1/1.325	8.8	.8
Abnehmende Kirchgangshäufigkeit	97.15	.000	-1.046	.351	1/4.281	26.5	9.2
Glauben an die Idee des Sozialismus	.73	.393	-.069	.933	1/1.074	.7	.1
Lebenswende zur Religion	5.00	.025	.585	1.795	1.234	1.1	.4
Glauben an ein Leben nach dem Tod	32.86	.000	1.150	3.157	1.772	12.4	2.4
Glauben an den Himmel	27.62	.000	1.283	3.606	1.840	13.8	2.1
Glauben an Wunder	.61	.436	.143	1.153	1.074	5.4	.1
Konstante	56.80	.000	4.836	-	-	-	-
P^2-Gesamtmodell (n = 1138)			38.5				

B) Neue Bundesländer

	χ^2	p	β	Effekt	Effekt$_S$	P^2_{brutto}	P^2_{netto}
Alter	16.08	.000	.027	1.028	1.519	4.7	1.2
Geschlecht = Mann	.74	.389	-.183	.833	1/1.095	1.0	.1
Wohnortsgröße	.03	.870	-.008	.992	1/1.017	1.2	.1
Konfessionslosigkeit	58.06	.000	-1.857	.156	1/2.412	36.4	4.3
Abnehmende Kirchgangshäufigkeit	75.74	.000	-1.175	.309	1/3.448	40.4	6.4
Glauben an die Idee des Sozialismus	9.87	.002	-.333	.717	1/1.378	3.2	.7
Lebenswende zur Religion	7.66	.006	.853	2.348	1.301	5.6	.6
Glauben an ein Leben nach dem Tod	19.44	.000	1.382	3.984	1.567	18.1	1.4
Glauben an den Himmel	10.36	.001	.873	2.394	1.384	13.7	.8
Glauben an Wunder	11.39	.001	.743	2.102	1.421	7.5	.8
Konstante	35.17	.000	4.746	-	-	-	-
P^2-Gesamtmodell (n = 1332)			55.6				

Zunächst sind die Zusammenhänge einiger demographischer Merkmale mit dem Vorhandensein des Gottesglaubens in Tabelle 8 aufgeführt. Exemplarisch wird die Bedeutung der verschiedenen Tabellenspalten anhand des Prädiktors **Alter** beschrieben.

Die Größen χ^2 und p weisen für Alter unter der Bedingung, daß alle konkurrierenden Prädiktoren statistisch konstant gehalten werden, in beiden Untersuchungsbereichen Werte aus, die auch bei einem noch strengeren α als .05 für Signifikanz sprechen. Die unstandardisierten Regressionskoeffizienten β ergeben mit .026 und .027 einen jeweils ähnlichen Einfluß auf das geschätzte Wahrscheinlichkeitsverhältnis zugunsten des Gottesglaubens. Als Effekt wird hier die Exponentiation von β benannt. Die Chancen, daß ein Befragter gottesgläubig ist, steigen durchschnittlich bei einem zusätzlichen Altersjahr um den Faktor 1.026 bzw. 1.028. Mit Effekts werden hier standardisierte Effekte bezeichnet, die darlegen, wie sich die Wahrscheinlichkeitsverhältnisse verschieben, wenn ein Prädiktor seinen Wert um +1 Standardabweichung ändert. Bei Effektwerten <1 wird für Effekts der Kehrwert angegeben. Effekts bietet jeweils *innerhalb* der beiden Modelle in Tabelle 8 eine Möglichkeit, die Effektgrößen unabhängig von den verschiedenen Meßeinheiten zu vergleichen. Auch die standardisierten Effekte von 1.554 bzw. 1.519 indizieren eine relativ große Bedeutung des Alters für das Vorhandensein des Gottesglaubens. P^2_{brutto} beschreibt die Devianzreduktion, welche in einer bivariaten logistischen Regression allein mit dem Prädiktor Alter erzielt werden kann, wobei als Basismodell das Konstantenmodell herangezogen wurde, so daß die Devianzreduktion seitens der Regressionskonstante außer acht gelassen wird (vgl. *Hensher/Johnson* 1981: **51; *Kühnel et*** al. 1989: 61 ff.; ***Kühnel*** 1990: 100). P^2 variiert zwischen 0 (keine Devianzreduktion) und 1 (100 Prozent Devianzreduktion) und kann annähernd analog interpretiert werden wie ein quadrierter Korrelationskoeffizient in der normalen OLS-Regression, wobei allerdings u.a. zu berücksichtigen ist, daß P^2 seltener vergleichbar hohe Werte erreicht wie R^2. P^2_{netto} schließlich weist aus, wie stark sich P^2 im multivariaten Modell reduziert, wenn der in der entsprechenden Tabellenzeile beschriebene Prädiktor aus dem jeweiligen Gesamtmodell entfernt wird.

Während sich für Alter in beiden Modellen ein annähernd ähnlicher positiver Einfluß auf das wahrscheinliche Vorhandensein von Gottesglauben ergibt, erweisen sich im Gesamtmodell die Merkmale **Geschlechtszugehörigkeit** und **Wohnortsgröße** nur in den alten Bundesländern als signifikant. Zwar sind beide Merkmale auch im Osten bivariat signifikant, so daß auch hier davon gesprochen werden kann, daß Männer und Befragte in größeren Orten seltener an Gott glauben, doch werden die Einflüsse dieser Prädiktoren bei dem in Tabelle 8 gezeigten komplexen Modell weitgehend substituiert[6].

Als sonstige demographische Prädiktoren wurden überprüft: Schulbildung, hauptberufliche Erwerbstätigkeit und die Angabe, ob ein Befragter aus einem Teil des wiedervereinigten Deutschlands in den anderen Teil übergesiedelt ist. Überraschenderweise konnte dabei nicht

6 Das schwächere Abschneiden der Ortsgröße in den neuen Bundesländern mag dadurch mit verursacht sein, daß hier nur eine siebenstufige Klassifikation im Datensatz verfügbar ist, während in den alten Bundesländern eine zehnstufige Einteilung nach ***Boustedt*** vorliegt.

festgestellt werden, daß mit höherer Schulbildung sich der Gottesglauben signifikant verringert. Geringfügige Effekte reduzieren sich bereits nach Kontrolle des Alters. Schulbildung vermittelt demzufolge kaum rationales Wissen, das den Gottesglauben systematisch ablösen würde[7]. Hinter der Überprüfung der Erwerbsbeteiligung steht die Annahme, daß das wiederholt festgestellte häufigere Auftreten religiöser Anschauungen bei Frauen auf deren seltenere Integration ins Arbeitsleben zurückzuführen sei. Erfahrungen, Identifikationen und Werte, die im alltäglichen Arbeitsleben erworben werden, können die Neigung, sich kirchlichem Engagement und religiösen Wertsetzungen anzuvertrauen, reduzieren *(De Vaus/ Mcallister* 1987: 480). Wird die Arbeitswelt von Männern dominiert, können erwerbstätige Frauen ferner dazu neigen, sich auch in ihren religiösen Einstellungen den Männern anzupassen (ebenda). Tatsächlich führt der Prädiktor Erwerbsbeteiligung auch bei einfacheren Modellen insbesondere in den neuen Bundesländern nicht zu den erwarteten Auswirkungen auf den Effekt der Geschlechtsvariable. Befragte in den alten Bundesländern, die früher im Gebiet der ehemaligen DDR gelebt haben, zeigen in Modellen mit ausschließlich demographischen Prädiktoren einen moderaten Effekt zuungunsten des Gottesglaubens. Für die neuen Bundesländer ist bei Übersiedlern aus dem Westen kein nennenswerter Effekt festzustellen, wobei allerdings zu berücksichtigen ist, daß die Zahl dieser Übersiedler im Sample sehr gering ist.

Insgesamt kann aus einfacheren multivariaten Modellen mit demographischen Prädiktoren abgeleitet werden: In beiden Teilen Deutschlands nimmt bei männlichem Geschlecht, in größeren Orten und bei geringerem Alter die geschätzte Wahrscheinlichkeit für den Gottesglauben signifikant ab. Dennoch sind die Devianzreduktionen von jeweils ca. 7 Prozent, die sich mit den bisher genannten demographischen Merkmalen erzielen lassen, leider noch recht unbefriedigend.

Wir gehen somit bei der Suche nach weiteren Erklärungsfaktoren zu Indikatoren **kirchlicher Partizipation** über. Global betrachtet, sind unter den Lutheranern größere Befragtenanteile als unter den Katholiken eigentlich atheistisch (vgl. Schaubild 1). Bereits einfachere Modelle zeigen aber, daß die Mitglieder der großen christlichen Konfessionen sich hinsichtlich der relativen Häufigkeit des Gottesglaubens weniger unterscheiden, wenn wir neben demographischen Merkmalen die ebenfalls differierende Kirchgangshäufigkeit kontrollieren[8]. Aktiv am Kirchenleben teilnehmende Kernmitglieder in den beiden großen Bekenntnisgemeinschaften ähneln sich also in dieser Hinsicht. Befragte ohne Konfessionsmitgliedschaft sind jedoch insbesondere in den neuen Bundesländern eine gesondert zu berücksichtigende Gruppe. Da auch nach einer ebenfalls überprüften Aufteilung der Kirchgangshäufigkeit in Dummy-Varia-

7 Operationalisiert wurde Schulbildung mit folgenden Ausprägungen: Maximal Hauptschulabschluß, mittlerer Schulabschluß (z.B. Mittlere Reife), Hochschulreife und Hochschulabschluß. Auch alternative Operationalisierungen mit Hilfe von Dummy-Variablen ergaben keinen sinnvoll interpretierbaren Bildungseffekt.

8 Geringe Unterschiede zwischen Protestanten und Katholiken ergeben sich multivariat - insbesondere nach Berücksichtigung der unterschiedlichen Kirchgangshäufigkeit - auch bei anderen Untersuchungen von Religiosität in den alten Bundesländern (z.B. *Terwey* 1988; ***Lukatis/Lukatis* 1989).**

blen Konfessionslosigkeit mit hoher Relevanz im Modell verbleibt, kann die Bedeutung dieses Prädiktors nicht in Gänze durch die geringe Kirchgangshäufigkeit der Konfessionslosen erklärt werden. Somit verbleiben in Tabelle 8 jeweils der Kontrast Konfessionsmitglieder vs. Konfessionslose und abnehmende Kirchgangshäufigkeit als vergleichsweise starke Prädiktoren. Insgesamt lassen sich mit konfessioneller Mitgliedschaft und Kirchgangshäufigkeit allein bereits Devianzreduktionen von 28,2 Prozent im Westen und 48,4 Prozent im Osten erzielen.

Wenn wir zu den weiteren Prädiktoren in Tabelle 8 übergehen, verlassen wir den Bereich der Variablen aus der Standarddemographie der Basisumfrage und gehen zu einigen Einstellungs- bzw. Glaubensfragen über, die in Zusammenhang mit dem Gottesglauben stehen. Wie bereits in Tabelle 6 gezeigt, wurde in den neuen Bundesländern merklich häufiger geäußert, der **Sozialismus** sei eine gute Idee, die bisher nur schlecht ausgeführt wurde. Unter den Bürgern aus der ehemaligen DDR steht der Glauben an die sozialistische Idee auch multivariat noch in signifikanter negativer Beziehung zum Gottesglauben. Diesen immer noch verbliebenen Befürwortern des Sozialismus schwebt mehrheitlich weiterhin eine atheistisch geprägte Idee vor und nicht etwa die eines frühchristlichen Sozialismus oder einer revolutionären Theologie. In den alten Bundesländern diskriminiert diese Einstellung demgegenüber kaum nennenswert zwischen Gläubigen und Ungläubigen wie der geringe Wert von .7 für P^2_{brutto} zeigt.

Abschließend werden in Tabelle 8 noch weitere Prädiktoren aus dem Religionsmodul des ISSP in der Basisumfrage eingeführt. Hierbei handelt es sich zunächst um die Frage: "Gab es jemals in ihrem Leben einen Wendepunkt, an dem Sie eine neue und persönliche Bindung zur Religion eingegangen sind?" und um einige der bereits in Tabelle 4 vorgestellten Glaubensvorstellungen aus dem weiteren christlichen Bereich[9] . Die explizite Angabe einer **Lebenswende zur Religion** ergibt in den alten Bundesländern eine Änderung des Wahrscheinlichkeitsverhältnisses zugunsten des Gottesglaubens von Effekt = 1.795. Mit 2.348 ist der Effekt in den neuen Bundesländern jedoch noch höher. Akzeptanz des Glaubens an ein **Leben nach dem Tod** und an den **Himmel** sind in beiden Teilsamples relevant, doch tragen sie bemerkenswerterweise in den alten Bundesländern relativ stark zur Devianzreduktion im multivariaten Modell bei (P^2_{netto} = 2.4 bzw. 2.1). Schließlich stoßen wir in den neuen Bundesländern nochmals auf die besondere Rolle der Akzeptanz von **Wundern,** die multivariat nur dort statistisch signifikant in Beziehung zum Gottesglauben steht. Wir wollen hier nicht

9 Diese Variablen wurden zu Dummies zusammengefaßt. Befragte, die eine Wende zur Religion angaben (16,7 % in den alten und 11,1 % in den neuen Bundesländern), und Befragte, welche die weiteren Glaubensvorstellungen als sicher oder wahrscheinlich akzeptieren, werden bei den jeweiligen Variablen mit 1 codiert. Alle anderen Befragten erhalten den Code 0. Insbesondere dadurch, daß auch Personen ohne valide Angaben, die in der schriftlichen ISSP-Befragung vergleichsweise häufig sind, unter dem Code 0 erfaßt sind, wird der Einfluß dieser Variablen in der Regel etwas unterschätzt. Der Vorteil dieser Verfahrensweise liegt jedoch darin, daß wir die verbleibende Zahl der validen Fälle im multivariaten Modell nicht drastisch reduzieren.

darüber spekulieren, ob es etwa Befragte sind, die besonders depriviert oder enttäuscht sind, welche ihre Zuflucht zum Wunderglauben nehmen.

Erweiterungen des in Tabelle 8 vorgestellten Modells sind mit dem reichhaltigen Datenmaterial der Basisumfrage selbstverständlich möglich. Allerdings kann, wie ergänzende Analysen gezeigt haben, für die generelle Frage nach dem Gottesglauben auf die Einführung der unterschiedlichen Akzeptanz von paranormalen Phänomenen wie **Glücksbringer, Wahrsager, Wunderheiler** und **Astrologie** gänzlich verzichtet werden. Nach den bisherigen Versuchen des Verfassers dürfte es auch schwierig sein, diese Merkmale als abhängige Variablen in erklärende Verfahren mit größeren Vorhersageerfolgen zu überführen. Hinsichtlich des Gottesglaubens können dagegen die P^2-Werte von 38.5 und 55.6 als relativ hoch eingeschätzt werden. Wenn wir einen diesbezüglichen Vergleich der beiden unterschiedlichen Teilstichproben zulassen, ist in den neuen Bundesländern die Prädiktion demzufolge etwas erfolgreicher. Wenden wir das anschaulichere, aber weniger strenge Kriterium der korrekten Klassifikation an, so können mit dem Variablensatz aus Tabelle 8 im Osten 89,3 Prozent aller Fälle hinsichtlich ihres Gottesglaubens richtig klassifiziert werden (alte Bundesländer 80,7 Prozent aller Fälle).

Zusammenfassung und Ausblick

Die neuen Bundesländer zeichnen sich im Vergleich zur ehemaligen Bundesrepublik durch einen deutlich geringeren Grad kirchlicher Integration aus. Besonders starke Einbrüche in der kirchlichen Partizipation verzeichnete die evangelische Kirche in der DDR, so weit bekannt, in den 70er Jahren *(Weichen* 1990: 310). Diese Einbrüche liegen zeitgleich mit der von *Weichert* 1990: 306 beschriebenen vorübergehenden Annäherung von Kirche und Staat in einigen Bereichen. Andererseits ist zu berücksichtigen, daß auch in der alten Bundesrepublik Kirchenaustritte während der 70er Jahre deutlich zunahmen (vgl. Statistisches Bundesamt; *Hanselmann* et al. 1984: 36), wozu die sozialkritischen Bewegungen seit den späten 60er Jahren beigetragen haben dürften.

Große Unterschiede bestehen zwischen den beiden Teilstichproben der Basisumfrage in der Verbreitung des Gottesglaubens und weiterer vorwiegend christlich besetzter Transzendenzannahmen. Geringere Unterschiede lassen sich hinsichtlich der Akzeptanz anderer paranormaler Phänomene verzeichnen. Dagegen ist in den neuen Bundesländern hohe Wertschätzung für die Idee des Sozialismus auch 1991 noch deutlich verbreiteter als im Westen. Diese Wertschätzung steht insbesondere im Osten in einem negativen Zusammenhang mit Religiosität. Die Beziehungen zwischen verschiedenen Glaubensvariablen und politischen Grundeinstellungen sind in faktoranalytischen Explorationen überwiegend ähnlich. Die abschließende Su-

che nach Erklärungsfaktoren für die unterschiedliche Verbreitung des Gottesglaubens als eines zentralen inhaltlichen Indikators für Religiosität ergab in beiden Teilsamples erst nach Berücksichtigung von Merkmalen kirchlicher Integration und anderen religiösen Einstellungen zufriedenstellende Resultate. Die in den neuen Bundesländern heute noch feststellbare relative Verbreitung des Gottesglaubens ist nach den Ergebnissen noch stärker auf diejenigen Gruppen zurückgedrängt worden, die entgegen dem verbreiteten Trend zur kirchlichen Desintegration in den Volkskirchen verblieben sind und aktiv partizipieren.

Über die hier analysierten Fragen hinaus erweisen auch weitere Daten der Basisumfrage in vieler Hinsicht eine größere Distanz zu den Kirchen unter den neuen Bundesbürgern insgesamt. Dieses läßt sich etwa bei der Frage nach dem Vertrauen zu Kirchen und religiösen Organisationen ablesen. Die oppositionelle Rolle, welche die Kirchen teilweise in der Auseinandersetzung mit dem Staat und bei dessen Ablösung gespielt haben, hat nicht zu einer generellen Bewertungsangleichung in Ost und West geführt. Es gibt jedoch auch Daten, die in Teilbereichen dieses Bild relativieren. Die Befragten in den neuen Bundesländern sprechen sich im Vergleich seltener für eine grundsätzliche Freigabe von Büchern oder Filmen aus, welche die Gefühle tief religiöser Menschen verletzen. Dagegen stößt eine Stellungnahme der Kirchen zu politischen Fragen auf mehr Akzeptanz. Schließlich wird der Aussage, daß die Kirchen und religiösen Organisationen in unserem Land zu viel Macht haben, von den neuen Bundesbürgern seltener zugestimmt, wobei die Erinnerung an die bisherigen faktischen Verhältnisse eine Rolle gespielt haben mag.

Literatur:

De Vaus, David und Ian McAllister 1987:
Gender differences in religion: A test of the structural location theory, in: American Sociological Review 52: 472-481.

EG (Kommission der Europäischen Gemeinschaften) 1985:
EURO-Barometer Nr. 24: Die öffentliche Meinung in der europäischen Gemeinschaft, in Spanien und Portugal Ende 1985, Brüssel.

Hanselmann, Johannes, Helmut Hild, Eduard Lohse (Hg.) 1984:
Was wird aus der Kirche? Ergebnisse der zweiten EKD-Umfrage über Kirchenmitgliedschaft, Gütersloh: Mohn.

Hensher, David A., Lester W. Johnson 1981:
Applied Discrete-Choice Modelling, London: Croom Helm.

Kliem, Wolfgang 1989:
Kirche im Sozialismus - aus der Sicht des Staates, in: Kommission SPD und Kirchen, Kirche im Sozialismus - Kirche im pluralen Staat, Mühlheim/Ruhr: Evangelische Akademie: 47 -55.

Kühnel, Steffen M. 1990:
Lassen sich mit SPSS^{x}-Matrix anwenderspezifische Analyseprobleme lösen? Ein Anwendungstest am Beispiel der multinomialen logistischen Regression, in: ZA-Information 27: 89 -109.

Kühnel, Steffen M., Wolfgang Jagodzinski, Michael Terwey 1989:
Teilnehmen oder boykottieren. Ein Anwendungsbeispiel der binären logistischen Regression mit SPSS^{X}, in: ZA-Information 25: 44 - 75.

Lukatis, Ingrid, Wolfgang Lukatis 1989:
Protestanten, Katholiken und Nicht-Kirchenmitglieder. Ein Vergleich ihrer Wert- und Orientierungsmuster, in: Karl-Fritz Daiber (Hg.). Religion und Konfession. Studien zu politischen, ethischen und religiösen Einstellungen von Katholiken, Protestanten und Konfessionslosen in der Bundesrepublik Deutschland und in den Niederlanden, Hannover: Lutherisches Verlagshaus: 17 - 71.

Noelle-Neumann, Elisabeth, Edgar Fiel (Hg.) 1983:
Allensbacher Jahrbuch der Demoskopie 1978 - 1983, München: K.G. Saur.

Nowossadeck, Sabine 1991:
Religion, in: Michael Hader (Hg.), Denken und Handeln in der Krise. Die DDR nach der "Wende": Ergebnisse einer empirisch soziologischen Studie, Berlin: Akademie: 95 - 99.

Schmidtchen, Gerhard 1978:
Die gesellschaftlichen Folgen der Entchristlichung, in: Wilhelm F. Kasch (Hg.), Entchristlichung und religiöse Desozialisation, Paderborn: Ferdinand Schöningh: 17-28.

Statistisches Bundesamt (verschiedene Jahrgänge):
Statistisches Jahrbuch für die Bundesrepublik Deutschland, Wiesbaden: Kohlhammer.

Terwey, Michael 1988:
Traditionelle Religiosität in der Bundesrepublik Deutschland, Papier präsentiert auf der 13. Jahresfachtagung des Arbeitskreises religions- und kirchensoziologischer Forschungsinstitute (ArkiF) am 18.4./20.4.88 in Rummelsberg.

Weichert, Brigitte 1990:
Gesellschaftliche Beteiligung, in: Gunnar Winkler (Hg.), Sozialreport '90. Daten und Fakten zur sozialen Lage in der DDR, Berlin: Die Wirtschaft: 279 - 312.

Regressionsanalyse mit Zeitreihendaten: Zur Interpretation dynamischer Modelle

von Helmut Thome

Abstract

Zeitreihendaten ermöglichen (unter bestimmten Voraussetzungen) die Schätzung ***dynamischer*** *Regressionsmodelle, in denen der Zeitpfad der Anpassung eines Systems (repräsentiert durch die Kriteriumsvariable) an eine veränderte Bedingungskonstellation (im Vektor der Prädiktoren) spezifiziert wird. Der Artikel erläutert die formale Struktur und den analytischen Gehalt derartiger Modelle.*

Time series data open up the possibility (given certain prerequisites) to estimate ***dynamic*** *regression modeis that specify the temporal dimension in which a system (the dependent variable) responds and adjusts itself to changes in its environment represented by the vector of predictor variables. The article expounds the formal structure and analytical substance of such "distributed lag" regression modeis.*

1. Einleitung

Auf zeitreihenanalytische Untersuchungen stößt man in der empirischen Sozialforschung nur selten. Das liegt zum einen daran, daß geeignete Datenstrukturen rar sind. Es liegt aber wohl auch daran, daß das entsprechende Methodenwissen in der Soziologie nicht weit verbreitet ist. Anwendungen zeitreihenanalytischer Verfahren findet man noch am ehesten in bestimmten Sparten der politischen Soziologie (siehe z. B. *Norpoth* 1984, *Kirchgässner* 1986) und der historisch orientierten Kriminalsoziologie (siehe z. B. ***Cohen/Land*** 1990, ***Devine*** et al. 1988). Die dort vorgestellten Regressionsmodelle variieren in der Regel folgende Gleichungsform:

$$(1) \qquad Y_t = \alpha + \delta Y_{t-1} + \beta X_t + \gamma Z_t + \varepsilon_t ,$$

wobei mit t=l,2,... die gleichabständigen Meßzeitpunkte der erhobenen Variablen indiziert sind und ε_t eine stochastische Fehlerkomponente repräsentiert. Zum Beispiel untersucht

Steven Stack (1987) die Mordrate (homicide) Y_t in Abhängigkeit von der um ein Intervall zurückliegenden Mordrate Y_{t-1} sowie mehreren exogenen Variablen wie der Arbeitslosenrate X_t und dem Vorkommen von Nachrichten über vollzogene Todesstrafen Z_t. In einer Fußnote (S. 535) erklärt er, daß er den verzögerten endogenen Term Y_{t-1} in die Schätzgleichung aufgenommen habe, um die Autokorrelation des Fehlerterms ε_t zu mindern. Die standardisierten Regressionskoeffizienten von *X* und Z deutet er ohne Berücksichtigung von δ als spezifische Einflußgewichte der jeweiligen exogenen Variablen. *Stacks* Interpretation ist nicht korrekt. Derartige Fehlinterpretationen findet man aber recht häufig in der soziologischen Literatur, obwohl dynamische Modelle wie Gleichung (1) in der Ökonometrie seit langem bekannt sind. Es scheint deshalb sinnvoll, noch einmal zu zeigen, worin der analytische Gehalt solcher Modelle besteht und wie ihre Parameter zu interpretieren sind. Auf Schätzprobleme gehe ich nur am Rande ein.

Das Verständnis dynamischer Modelle wird erleichtert, wenn man sich zuvor einer Voraussetzung erinnert, die bei der Interpretation der Regressionsanalyse mit Querschnittsdaten oft nicht gebührend beachtet wird. Gegeben sei das Modell

$$(2) \quad Y_i = \alpha + \beta X_i + \gamma Z_i + \varepsilon_i$$
$$E(\varepsilon_i)=0,\ Var(\varepsilon_i)=\sigma^2,\ Cov(\varepsilon_i,\varepsilon_j)=0,\ i\neq j,\ (i,j=1,2,\ldots,n)$$
$$Cov(X_i,\varepsilon_i)=0,\ Cov(Z_i,\varepsilon_i)=0$$

Die Interpretation der Steigungskoeffizienten β und γ als Kausalgewichte der exogenen Variablen wird üblicherweise in Form eines Bedingungssatzes gegeben: Wenn *X (Z)* um eine Einheit zunimmt (zunähme), nimmt (nähme) *Y* durchschnittlich um β (γ) Einheiten zu. Die Veränderung selbst kann jedoch mit Querschnittsdaten nicht beobachtet werden; sie wird quasi simuliert, indem man bei den verschiedenen Untersuchungsobjekten unterschiedliche Werte für die einzelnen Variablen beobachtet und voraussetzt, daß bei allen Untersuchungseinheiten der gleiche strukturelle Zusammenhang vorliegt. Die Kausalinterpretation ist aber nur dann sinnvoll, wenn sich die Daten zum Zeitpunkt der Messung in Ruhelage (Äquilibrium) befinden. Das heißt, zum Zeitpunkt der Messung muß jede vor der Messung abgelaufene (aber nicht beobachtete) Änderung von *X*- oder Z-Werten sich bereits voll auf *Y* ausgewirkt haben, andernfalls ist mit einer Verzerrung der geschätzten Regressionskoeffizienten zu rechnen. Wenn die (nicht beobachtete) Veränderungsrate von *Y* mit einer unabhängigen Variablen korreliert, wird der entsprechende Steigungskoeffizient, andernfalls nur der Ordinatenabschnitt verzerrt geschätzt. Die Gleichgewichtsvoraussetzung, die in Lehrbüchern und in der Forschungspraxis oft unterschlagen wird, ist in

vielen sozialwissenschaftlichen Untersuchungen sicherlich nicht erfüllt.

2. Übergang von der statischen zur dynamischen Zeitreihenanalyse

Zeitreihenanalyse ist eine Form der dynamischen Analyse, bei der man die Äquilibriumsvoraussetzung fallen läßt oder wenigstens fallen lassen kann. Das heißt, man erhält die Möglichkeit, die Anpassung eines Systems in der Merkmalsdimension *Y* an eine veränderte Bedingungskonstellation (im Vektor der Prädiktoren) über Zeit zu rekonstruieren. Allerdings werden auch Zusammenhänge zwischen Zeitreihen gelegentlich in **statischer** Weise analysiert, beispielsweise mit folgendem Modell:

$$(3) \quad Y_t = \alpha + \beta X_t + \gamma Z_t + \varepsilon_t$$

Im Unterschied zur Querschnittsanalyse werden die Messungen jetzt nicht an *n* Objekten, sondern an einem einzigen Objekt über *[T]* diskrete Zeitpunkte *t=l,2,...,T* vorgenommen. Die Annahme unkorrelierter Fehler ist bei Zeitreihendaten in der Regel nicht haltbar - mit den bekannten Konsequenzen für die Schätzung der Parameter und ihrer Standardfehler (s. z. B. *Kmenta* 1986). Die Interpretation der Steigungskoeffizienten ändert sich im Vergleich zur Querschnittsanalyse jedoch nicht: die Koeffizienten ß und *y* geben an, um wieviel Einheiten sich Y_t in Reaktion auf Einheitsänderungen in *X*, bzw. *Z*, ändert. Im Modell (3) wird angenommen, daß sich diese Reaktion "sofort", also noch innerhalb des gleichen Meßintervalls, vollzieht, in dem die Änderungen in *X* bzw. Z auftraten. Das heißt, die Äquilibriumsvoraussetzung wird auch hier praktisch als erfüllt betrachtet; eine Veränderungsmessung ist nicht Bestandteil des Modells. Einen ersten Schritt in Richtung **dynamischer** Analyse unternimmt man hingegen mit der folgenden Gleichung:

$$(4) \quad Y_t = \alpha + \beta X_{t-k} + \gamma Z_{t-l} + \varepsilon_t \text{, wobei } 0 < k,l < T$$

In diesem Modell wird angenommen, daß sich eine Änderung in *X (Z)* zum Zeitpunkt *t* erst nach *k (l)* Zeitintervallen in *Y* auswirkt, daß nach diesem Zeitpunkt aber keine zusätzlichen Wirkungen mehr eintreten. Man kann dieses Modell dadurch erweitern, daß man die Wirkung von *X* oder/und Z auf mehrere Intervalle ausdehnt, z. B.

$$(5) \quad Y_t = \alpha + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \beta_3 X_{t-3} + \gamma_t Z_{t-l} + \varepsilon_t$$

Jetzt wird angenommen, daß eine Änderung in *X* zum Zeitpunkt *t* mehrere Intervalle benötigt, um sich voll auf *Y* auszuwirken; sie wird erstmals sichtbar nach einem Intervall und erhält weitere Zuwächse in den beiden folgenden Intervallen. Man beachte, daß die Stei-

gungskoeffizienten **partielle** Koeffizienten darstellen, d. h., die volle (erwartete) Änderung in *Y* auf einen dauerhaften Einheitszuwachs in *X* ist mit $\beta = \beta_1+\beta_2+\beta_3$ gegeben. Modelle wie Gleichung (5) werden auch als "dynamisch in *X*" bezeichnet. Der Begriff der "dynamischen Regression" oder "distributed lag regression" wird meistens jedoch für Modelle reserviert, die mindestens einen verzögerten endogenen Term, z. B. Y_{t-1}, in die rechte Gleichungsseite mit aufnehmen, wie oben in Gleichung (1). Um solche Modelle verstehen zu können, betrachten wir zunächst folgende Gleichung:

$$(6) \qquad Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \cdots + \beta_\infty X_{t-\infty} + \varepsilon_t$$

Hinsichtlich des Fehlerterms ε_t sollen (einstweilen) die Voraussetzungen analog zu Modell (2) gelten. Modell (6) bezeichnet man als "general infinite lag structure" *(Dhrymes* 1971, S. 30 f.); es ist die allgemeinste Form des Regressionsmodells mit verteilten Verzögerungen (Lags) bei nur einer exogenen Variablen. Es ist klar, daß man die unendlich große Zahl der Parameter eines solchen Modells nicht schätzen kann. Man kann es aber als Basismodell benutzen, aus dem sich mittels Parameter-Restriktionen realistische Modelle gewinnen lassen. Oft reicht aber das theoretische Wissen nicht aus, um im voraus sagen zu können, über wieviele Beobachtungsintervalle sich eine Wirkung von *X* auf *Y* entfaltet. Möglicherweise gibt es aber eine begründete Vorstellung über die Form des Zeitpfads dieser Wirkung. Man könnte z. B. annehmen, daß sie innerhalb des ersten Beobachtungsintervalls am stärksten ist und daß die in den nächsten Perioden hinzukommenden Effekte geometrisch abnehmen.

Die Annahme der geometrischen Abflachung der Wirkungskurve läßt sich formal ausdrükken in der Gleichung

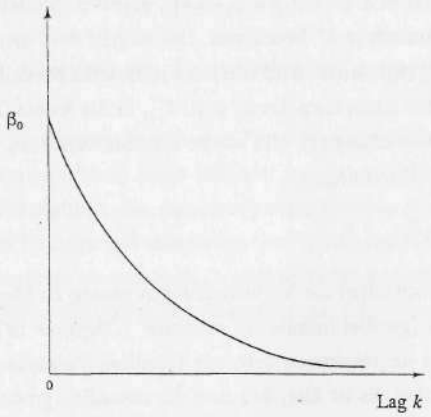
$$(7) \qquad \beta_k = \beta_0 \delta^k \;;\; k = 1,2,\cdots \;,\; 0<\delta<1$$

Nach der Anfangswirkung β_0 lassen sich alle nachfolgenden β's als Produkt aus β_0 und einer noch zu bestimmenden Potenzzahl δ errechnen. Wenn man die Koeffizientenrestriktion für die Beta-Gewichte gemäß Gleichung (7) in Gleichung (6) einführt, erhält man

$$(8) \qquad Y_t = \alpha+\beta_0 X_t+\beta_0\delta X_{t-1}+\beta_0\delta^2 X_{t-2}+\cdots+\beta_0\delta^{\infty} X_{t-\infty} + \varepsilon_t$$

Schematisch läßt sich die Abfolge der (partiellen!) Regressionskoeffizienten wie in Abb. 1 darstellen.

Abb. 1: Muster der Regressionskoeffizienten eines geometrischen Lag-Modells



Der Abschwung führt um so steiler gegen Null, je kleiner das Gewicht δ ist. Betrachtet man die Koeffizienten β_k als Funktion des Lags k = 0,1,2,..., so spricht man allgemein von einer "Antwortfunktion" oder einer "impulse-response-function". Modell (8) enthält nur noch drei unbekannte Parameter: α, β_0 und δ. Wie können sie, wie kann insbesondere δ geschätzt werden? Die Schätzung wird durch eine Reparametrisierung *(Koyck-Transforma.*-tion) der Gleichung (8) ermöglicht. Wenn Gleichung (8) gilt, dann gilt auch

$$(9) \qquad Y_{t-1} = \alpha + \beta_0 X_{t-1} + \beta_0 \delta X_{t-2} + \cdots + \varepsilon_{t-1}$$

Gleichung (9) multipliziert mit dem Faktor δ führt zu

$$(10) \qquad \delta Y_{t-1} = \alpha\delta + \beta_0 \delta X_{t-1} + \beta_0 \delta^2 X_{t-2} + \cdots + \delta\varepsilon_{t-1}$$

Im nächsten Schritt subtrahieren wir Gleichung (10) von Gleichung (8)

$$(11) \qquad Y_t - \delta Y_{t-1} = (\alpha - \delta\alpha) + \beta_0 X_t + (\varepsilon_t - \delta\varepsilon_{t-1})$$

Wenn wir $\alpha^* = (\alpha - \delta\alpha)$ und $\varepsilon^*_t = \varepsilon_t - \delta\varepsilon_{t-1}$ (Fehler-Autokorrelation 1. Ordnung!) schreiben und die Ausdrücke in Gleichung (11) neu ordnen, erhalten wir die Schätzgleichung

$$(12) \qquad Y_t = \alpha^* + \delta Y_{t-1} + \beta_0 X_t + \varepsilon^*_t \quad , \quad 0<\delta<1$$

Wenn δ und β_0 bekannt bzw. geschätzt sind, lassen sich die einzelnen Beta-Gewichte der Ausgangsgleichung (6) gemäß Gleichung (7) ermitteln und somit das Anpassungsverhalten des Systems über Zeit beschreiben. Das übliche Kleinstquadratverfahren (OLS-

Regression) ist jedoch auf Gleichung (12) nicht anwendbar. Die Kombination autokorrelierter Fehler und verzögerter endogener Variablen als Regressor auf der rechten Seite der Modellgleichung läßt die OLS-Schätzer inkonsistent werden *(Johnston* 1972, S. 307); wie Gleichung (9) zu entnehmen ist, sind der Regressor Y_{t-1} und der Fehlerterm ε_{t-1} miteinander korreliert. Diese Problematik möchte ich jedoch hier nicht erörtern (zu alternativen Schätzverfahren siehe *Hibbs* (1974) und *Judge* et al. (1980, S. 170 ff.).

Gleichung (12) ähnelt bereits stark Gleichung (1), die unsere Betrachtungen auslöste. Die inhaltliche Bedeutung des Modells (12) erschließt sich vielleicht am ehesten anhand einer Beispielrechnung, in der wir die erwartbare, zeitlich gestreckte Reaktion von *Y* auf drei unterschiedliche Formen der Input-Änderung in *X* für das Modell

$$E(Y_t) = 0{,}5Y_{t-1} + 0{,}5X_t$$

ermitteln. Anhand der folgenden Tabelle 1 betrachten wir die Y-Reaktion auf (a) einen *Pulse-Input* (Spalte *a),* der das X-Niveau zu einem bestimmten Zeitpunkt *t* vorübergehend von 0 auf 1 anhebt, (b) einen Step-Input (Spalte *b),* der das X-Niveau zu einem bestimmten Zeitpunkt *t* dauerhaft von 0 auf 1 erhöht, sowie (c) einen allgemeinen, stochastischen Input mit ständig wechselnden X-Werten.

Tabelle 1

Output-Entwicklung bei verschiedenen Input-Formen im Modell
$E(Y_t) = 0{,}5Y_{t-1} + 2{,}5X_{t-1}$

	a		b		c		d		e	
t	x_t	y_t	x_t	y_t	x_t	y_t	x_t	y_t	x_t	y_t
	Pulse Input	Output	Step Input	Output	Input	Output	Input	Output	General Input	Output
0	0	0	0	0	0	0	2,0	8,0	0	0
1	1	0	1	0	1,5	0	3,5	9,0	1,5	0
2	0	2,5	1	2,5	0,5	3,75	2,5	13,25	0,5	3,75
3	0	1,25	1	3,75	2,0	3,12	4,0	12,87	2,0	3,12
4	0	0,62	1	4,38	2,0	6,56	4,0	16,44	1,0	6,56
5	0	0,31	1	4,69	2,0	8,28	4,0	16,22	-2,5	5,78
6	0	0,16	1	4,84	2,0	9,14	4,0	19,11	0,5	-3,36
7	0	0,08	1	4,92	2,0	9,57	4,0	19,56	1,0	0,43
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
.	.	Y_g=0,0	.	Y_g=5,0 =G·1	.	Y_g=10,0 =G·2	.	Y_g=20,0 =G·4	.	.

Quelle: ***Box*** und ***Jenkins*** (1976, S. 354) sowie eigene Berechnungen

Tabelle 2 zeigt das Berechnungsschema für den Pulse-Input; die Modifikationen für die anderen Input-Formen sind offensichtlich, so daß sie hier nicht demonstriert werden müssen.

Tabelle 2

Zeitpfad des Modells $E(Y_t) = 0{,}5Y_{t-1} + 2{,}5X_t$ bei einem Pulse-Input

Zeitperiode t	Y_t	Y_{t-1}	X_t	Lag k
0	0	0	0	
1	$\beta_0 = 2{,}5$	0	1	0
2	$\beta_0\delta = 1{,}25$	β_0	0	1
3	$\beta_0\delta^2 = 0{,}62$	$\beta_0\delta$	0	2
4	$\beta_0\delta^3 = 0{,}31$	$\beta_0\delta^2$	0	3
.	...	...	.	.

Bei einem Pulse-Input der Größe $|1|$ ist der Zeitpfad der Anpassungsreaktion in *Y* unmittelbar durch die Folge der β-Koeffizienten der Gleichung (6) gegeben, die sich über die Restriktionsgleichung (7) aus den Parametern unseres Beispielmodells berechnen lassen. Bei einem Step-Input paßt sich *Y* allmählich an ein neues Niveau, einen neuen Gleichgewichtswert Y_g an. Der Gleichgewichtszuwachs *G* pro Einheitszuwachs in *X* ergibt sich als Grenzwert aus der Addition der (partiellen) Steigungskoeffizienten, wie ein Vergleich der entsprechenden Ergebnisspalten in Tab. 1 (Spalten *a* und *b)* zeigt:

$$(13) \quad G = \beta_0 + \beta_1 + \ldots + \beta_\infty = \beta_0 \sum_{k=1}^{\infty} \delta^k \quad , \quad 0<\delta<1$$

Da die Summanden in Gleichung (13) eine geometrische Reihe bilden, läßt sich der Grenzwert unmittelbar aus den Parametern des reduzierten Modells (12) errechnen:

$$(14) \quad G = \frac{\beta_0}{1 - \delta} = \frac{2,5}{1 - 0,5} = 5$$

Die Beispiele in den Spalten *c* und *d* der Tabelle 1 zeigen, daß (in einem linearen Modell) der Gleichgewichtszuwachs in Y pro dauerhaftem X-Zuwachs um 1 Einheit völlig unabhängig ist von den Anfangswerten der beiden Variablen und der numerischen Größe des Niveausprungs in *X*. Der erwartete Gleichgewichtswert Y_g läßt sich somit als Funktion des Gleichgewichtswerts X_g (falls existent) darstellen:

$$E(Y_g) = \alpha + G \cdot X_g$$

Der Koeffizient *G* entspricht also dem Steigungskoeffizienten β im statischen Regressionsmodell für Querschnittsdaten. Die drei Input-Formen von *X* führen zu unterschiedlich präzisen Schätzungen des G-Koeffizienten. Es ist auch damit zu rechnen, daß ein soziales System jeweils andere Anpassungsmechanismen entwickelt, je nachdem, ob die Bedingungskonstellationen einmalig, dauerhaft oder ständig wechseln, so daß der errechnete Gleichgewichtszuwachs eine etwas unsichere analytische Konstruktion ist, wenn er auf der Basis stochastischer Input-Werte ermittelt wird. Beispiele für komplexere Anpassungsreaktionen werden weiter unten präsentiert.

Bisher haben wir zwei Elemente des analytischen Gehalts der dynamischen Regressionsgleichung (12) herausgearbeitet:
(a) Sie läßt sich als Reparametrisierung der Gleichung (6) verstehen, wenn die β-Koeffizienten in bestimmter Weise restringiert werden. Das bedeutet, das kausale Gewicht von *X* hängt nicht nur von dem Parameter β_0, sondern auch von dem Parameter δ ab und drückt sich numerisch in dem Gleichgewichtskoeffizienten *G* aus. Der endogene Regressor Y_{t-1} repräsentiert in dem theoretischen Modell (12) also keine externen Einflußgrößen außer *X*. Falls das Modell (12) unterspezifiziert ist, wird der Koeffizient δ in der Regel einiges von dem Einflußgewicht der ausgelassenen Variablen aufsaugen. Das bedeutet, daß er in diesem Falle nicht konsistent geschätzt wird. Es bedeutet nicht, daß der Mangel der Unterspezifikation auf diese Weise zu beseitigen oder auch nur zu mindern wäre. Auch sollte der Regressor Y_{t-1} nicht in der bisher vorgestellten Weise mit dem Zweck in das Schätzmodell eingeführt werden, die Autokorrelation der Fehler zu verringern. Dazu unten mehr.

(b) Der zweite Aspekt des bisher herausgearbeiteten analytischen Gehalts bestand darin, zu zeigen, wie der Zeitpfad der Wirkung von *X* auf *Y* sowie das neue (hypothetische) Gleich-

gewichtsniveau anhand der Modellparameter ermittelt werden kann.

Als nächstes ist zu erläutern, wie sich geometrische Lag-Modelle in Form der Gleichung (12) substanztheoretisch begründen lassen. Dabei geht es nicht um den spezifischen Inhalt einer substanzwissenschaftlichen Theorie, sondern lediglich darum, die Übersetzung theoretischer Überlegungen in dynamische Regressionsmodelle allgemein zu veranschaulichen.

3. Eine substanztheoretische Begründung für geometrische Lag-Modelle

In der Ökonomie wird in mancherlei Zusammenhängen das Verhalten im Sinne eines adaptiven Erwartungsmodells ("adaptive expectation model") gedeutet. Dessen Grundstruktur läßt sich wie folgt formalisieren:

$$(15) \quad Y_t = \alpha + \beta X^*_{t+1} + \varepsilon_t ,$$

wobei für den stochastischen Fehlerterm wiederum die üblichen Bedingungen analog zu Modell (2) gelten sollen. Die abhängige Variable Y_t (z. B. die gegenwärtig nachgefragte Gütermenge) soll von dem *erwarteten* Wert X^*_{t+1} der Prädiktorvariable (z. B. dem für die Zukunft erhofften Einkommen) abhängen. Wenn man den erwarteten Wert nicht direkt beobachten kann, benötigt man eine theoretische Vorstellung, wie die Erwartungen sich als Funktion beobachtbarer Größen entwickeln. Es wird z. B. angenommen, daß sich das für die Periode *t+1* erwartete Einkommen als gewichteter Durchschnitt aus dem gegenwärtig, zum Zeitpunkt *t,* erzielten und dem für diesen Zeitpunkt erwarteten Einkommen ergibt:

$$(16) \quad X^*_{t+1} = \delta X^*_t + (1-\delta)X_t \quad , \quad \delta < 1$$

Offensichtlich gilt dann auch

$$(17) \quad X^*_t = \delta X^*_{t-1} + (1-\delta)X_{t-1}$$

Durch Einsetzen von Gleichung (17) in Gleichung (16) erhält man

$$(18) \quad \begin{aligned} X^*_{t+1} &= \delta[\delta X^*_{t-1} + (1-\delta)X_{t-1}] + (1-\delta)X_t \\ &= (1-\delta)(X_t + \delta X_{t-1}) + \delta^2 X^*_{t-1} \end{aligned}$$

Durch fortgesetzte Substitution der gesternten X-Terme ergibt sich

$$(19) \quad X^*_{t+1} = (1-\delta)(X_t + \delta X_{t-1} + \delta^2 X_{t-2} + \ldots)$$

Dieser Ausdruck, eingesetzt in Gleichung (15), führt unmittelbar zu dem geometrischen Lag-Modell

$$(20) \qquad Y_t = \alpha + \beta(1-\delta)(X_t + \delta X_{t-1} + \delta^2 X_{t-2} + \ldots) + \varepsilon_t$$

das identisch ist mit Gleichung (8), wenn man den Koeffizienten $\beta(1-\delta) = \beta_0$ setzt.[1]

Die modellthoeretischen Überlegungen könnten noch weitergetrieben werden, indem man zeigte, unter welchen Bedingungen das Modell adaptiver Erwartungen sich als Spezialfall des allgemeineren Modells der rationalen Erwartungsbildung darstellt. Hierzu sei aber lediglich auf die einschlägige Literatur verwiesen (siehe z. B. *Judge* et al. 1980, S. 626 ff. oder ***Kmenta*** 1986, S. 529, 544 ff. und das sozialwissenschaftliche Anwendungsbeispiel in *Kirchgässner* 1985 a). Hinzuweisen ist auch auf das "partial adjustment" bzw. "habit persistence model", in dem eine verhaltenstheoretische Hypothese ebenfalls in ein geometrisches Lag-Modell analog zu Gleichung (12) übersetzt wird, diesmal allerdings mit der Möglichkeit, eine unkorrelierte Fehlerstruktur für die Schätzgleichung annehmen zu können (s. *Dhrymes* 1971, S. 58; *Judge* et al. 1980, S. 626 f., 630; *Kmenta* 1986, S. 529 f.)

4. Autokorrelatives Fehlermodell vs. geometrische Lagstruktur des strukturellen Zusammenhangs

Wie allgemein bekannt, führt die Methode der kleinsten Quadrate bei autokorrelierten Fehlern zu nicht-effizienten Schätzungen der Regressionsparameter und zu Verzerrungen in der Schätzung ihrer Standardfehler. Alternative Schätzmethoden beinhalten häufig Variablentransformationen, die im Endeffekt ebenfalls einen verzögerten endogenen Term in die Regressionsgleichung einführen. Vermutlich verführt dies manche Forschungspraktiker dazu, die Größe Y_{t-1} auch dann in die Regressionsgleichung einzuführen, wenn nicht eine geometrische Lag-Struktur für die Beziehung zwischen abhängiger und unabhängiger Variable postuliert, sondern lediglich eine autoregressive Fehlerstruktur bewältigt werden soll - wie etwa:

---

[1] Da das Erwartungsmodell (16) zunächst auf der Individualebene formuliert wird, die Zeitreihendaten in der Regel aber für ein Kollektiv erhoben werden, müßte man zusätzliche Annahmen einführen, um das Modell auch auf der Aggregatebene begründen zu können.

$$(21) \quad Y_t = \beta X_t + \varepsilon_t$$
$$\varepsilon_t = \rho \varepsilon_{t-1} + u_t \quad , \quad |\rho| < 1$$

(mit den Standard-Annahmen über den stochastischen Term u_t). Diese beiden Modellkomponenten sind jedoch konzeptuell und formal strikt getrennt. Für das häufig angewandte *Cochrane/Orcutt-Schätzverfahren* läßt sich dies z. B. in folgender Weise zeigen (vergl. ***Dhrymes*** 1971, S. 59 ff.; ***Kmenta*** 1986, S. 314 ff.): In einem ersten Schritt werden die Residuen ε_t in (21) durch die OLS-Residuen e_t ersetzt, und der Koeffizient ρ wird anhand der Stichproben-Autokorrelation (unter Verlust des ersten Beobachtungswertes) geschätzt[2]:

$$(22) \quad \hat{\rho} = \frac{\sum e_t e_{t-1}}{\sum e_{t-1}^2} \quad ; \quad t = 2,3,\ldots,T$$

Mit Hilfe dieses Schätzers wird die Regressionsgleichung wie folgt transformiert:

$$(23) \quad Y_t = \alpha + \beta X_t + \varepsilon_t \quad , \quad \varepsilon_t = \rho \varepsilon_{t-1} + u_t$$
$$(Y_t - \hat{\rho} Y_{t-1}) = \alpha - \hat{\rho}\alpha + \beta (X_t - \hat{\rho} X_{t-1}) + \hat{u}_t$$
$$Y_t^* = \alpha^* + \beta X_t^* + \hat{u}_t \quad ; \quad Y_t^* = Y_t - \hat{\rho} Y_{t-1} \; , \; X_t^* = X_t - \hat{\rho} X_{t-1}$$

Mit den so transformierten Daten ("generalized differences") können die Modellparameter α und β mit dem gewöhnlichen Kleinstquadratverfahren (OLS-Regression) geschätzt werden[3]. Die Transformation in Gleichung (23) hat die Struktur der Beziehung zwischen *Y* und *X* nicht verändert, da beide Variablen in gleicher Weise (mit dem gleichen "Filter") transformiert wurden. Daß dieses statische Modell mit autoregressiver Fehlerstruktur gelegentlich mit einem dynamischen Modell verwechselt wird, liegt vermutlich an der möglichen Umstellung der Terme der zweiten Zeile in Gleichung (23), die sich wie folgt schreiben läßt:

---

[2] Die OLS-Residuen sind bei kleinen Stichproben (Daumenregel: T<50) verzerrt; dennoch ist der Schätzer für p konsistent. Über die Eigenschaften verschiedener Schätzer siehe *Judge* et al. 1980, S. 182 ff. in Verbindung mit S. 212 ff. sowie ***Kmenta*** 1986, S. 315).

[3] Die oft praktizierte Transformation $Y^*_t = Y_t - Y_{t-1}$ und $X^*_t = X_t - X_{t-1}$ (einfache Differenzenbildung) führt zu einer verzerrten Schätzung der Standardfehler des Steigungskoeffizienten, ist also nicht empfehlenswert, wenn p nicht tatsächlich in der Nähe von eins liegt (s. ***Kmenta*** 1986: 322).

(23 a) $Y_t = \alpha^* + \rho Y_{t-1} + \beta X_t - \beta\rho X_{t-1} + \hat{u}_t \ , \ \alpha^* = \alpha - \rho\alpha$

Dieses Modell scheint sich formal von dem geometrischen Lag-Modell (12) nur durch die Hinzunahme des verzögerten Terms der exogenen Variablen zu unterscheiden. Man beachte aber die Parameterrestriktion: Der Koeffizient von X_{t-1} ist gleich dem negativen Produkt der Koeffizienten von X_t und Y_{t-1}. Was diese Restriktion inhaltlich bedeutet, läßt sich leicht erkennen, wenn man (unter Voraussetzung der Stationarität des Prozesses) Gleichung (23) um ein Intervall verschiebt:

(23 b) $Y_{t-1} = \alpha + \beta X_{t-1} + \varepsilon_{t-1}$

Wenn wir diesen Ausdruck für Y_{t-1} in Gleichung (23 a) einsetzen, erhalten wir

(23c)
$$\begin{aligned} Y_t &= \alpha^* + (\rho\alpha + \rho\beta X_{t-1} + \rho\varepsilon_{t-1}) + \beta X_t - \rho\beta X_{t-1} + \hat{u}_t \\ &= (\alpha - \rho\alpha) + \rho\alpha + \rho\beta X_{t-1} - \rho\beta X_{t-1} + \beta X_t + \rho\varepsilon_{t-1} + \hat{u}_t \\ &= \alpha + \beta X_t + \varepsilon_t \end{aligned}$$

Da Y_{t-1} selbst eine Funktion von βX_{t-1} ist, heben sich die beiden Ausdrücke durch die ρ-Gewichtung mit entgegengesetztem Vorzeichen auf. Dies wird weiter verdeutlicht, wenn wir das folgende dynamische Modell (24) betrachten, das eine geometrische Lag-Struktur enthält, die sich von der des Modells (12) lediglich durch die Hinzunahme des verzögerten Terms der *exogenen* Variablen, X_{t-1}, unterscheidet:

(24) $Y_t = \alpha + \delta Y_{t-1} + \beta_0 X_t + \beta_1 X_{t-1} + \varepsilon_t \ , \ 0<\delta<1$

Dieses dynamische Modell ähnelt noch stärker als Modell (12) dem statischen Regressionsmodell der Gleichung (23). Der entscheidende Unterschied besteht darin, daß in Gleichung (24) der Parameter β_1 nicht restringiert ist. Dadurch ergibt sich eine über Zeit verteilte Wirkung von *X* nach *Y,* ein Zeitpfad, dessen Verlauf Tabelle 3 darstellt.

Tabelle 3

Zeitpfad des Modells $E(Y_t) = \delta Y_{t-1} + \beta_0 X_t + \beta X_{t-1}$ bei einem Pulse-Input

Zeitperiode t	Y_t	Y_{t-1}	X_t	X_{t-1}	Lag k
0	0	0	0	0	
1	β_0	0	1	0	0
2	$\beta_0\delta + \beta_1$	β_0	0	1	1
3	$(\beta_0\delta + \beta_1)\delta$	$\beta_0\delta + \beta_1$	0	0	2
4	$(\beta_0\delta + \beta_1)\delta^2$	$(\beta_0\delta + \beta_1)\delta$	0	0	3
5	$(\beta_0\delta + \beta_1)\delta^3$	$(\beta_0\delta + \beta_1)\delta^2$	0	0	4
.	...	...	.	.	.

Gegenüber dem Modell (12) beginnt der geometrische Abfall des Zeitprofils der "impulse-response-function" (also der Wirkungszuwächse) jetzt ein Intervall später. In Modell (24) ist der maximale Lag der *X*-Variable *k=1;* theoretisch könnte der maximale Lag irgendeinen Wert zwischen 0 und *T* annehmen: $0 \leq k < T$. Allgemein gilt, daß der durch den Koeffizienten δ gesteuerte geometrische Abschwung bei dem Intervall *k+1* beginnt. Der Unterschied zwischen einem autoregressiven Fehlermodell 1. Ordnung - wie in Gleichung (23) - und einem geometrischen Lag-Modell wie in Gleichung (24) wird durch Tab. 3 nochmals verdeutlicht: Im Falle der Koeffizientenrestriktion $\beta_1 = -\beta_0\delta$ wird die Summe $\beta_0\delta + \beta_1$ gleich Null.

Nun ist der Sozialforscher aber häufig nicht in der Lage, sich aus theoretischen Gründen a priori entweder für ein dynamisches Regressionsmodell wie in Gleichung (24) oder ein statisches Regressionsmodell mit autoregressiver Fehlerstruktur wie in Gleichung (23) zu entscheiden. Mit Stichprobenschätzern wird man aber kaum jemals die Identität $\beta_0\delta + \beta_1 = 0$ in den geschätzten Koeffizienten realisieren, auch wenn sie im "wahren" Modell gegeben ist.[4] *Rao/Miller* (1971, S. 174) empfehlen für diesen Fall, sowohl Modell (23) als auch Modell (12) zu schätzen, die jeweiligen empirischen

[4] Zur Testproblematik siehe *Judge* et al. (1985: 321 f.)

Zeitprofile ("impulse response function") zu ermitteln und anhand der Muster, die sie bilden, sich für den einen oder anderen Modelltyp zu entscheiden. Falls das wahre Modell eine geometrische Lag-Struktur beinhaltet, werden die Gleichungen (23) und (12) zu ähnlichen Zeitprofilen führen. Falls das wahre Modell keine geometrische Lag-Struktur für die Beziehung zwischen *X* und *Y* aufweist, werden die Schätzer des Modells (12) - zurücktransformiert in die Schätzer des Modells (6), also dessen Antwortfunktion - wegen des ausgelassenen Terms X_{t-1} dennoch eine verteilte Wirkung mit abflachendem Muster nahelegen, während die Schätzer des Modells (23) ab Lag *k=1* unregelmäßig nahe Null schwanken. Eine allgemeinere Strategie zur Modellidentifikation haben *Box/Jenkins* (1976, Kap. 10) mit ihrer Interpretation des sog. Kreuzkorrelogramms (siehe unten, Abschn. 6) vorgeschlagen.

5. Geometrischer Lag bei zwei Input-Variablen

Bisher haben wir geometrische Lag-Modelle lediglich für eine einzige Input-Variable *X* spezifiziert. Dieses Modell wollen wir nun um eine zweite Input-Variable Z erweitern, die ebenfalls mit geometrischer Lag-Struktur auf *Y* wirken soll:

$$(25) \quad Y_t = \alpha + \beta_0(X_t + \delta X_{t-1} + \delta^2 X_{t-2} + \ldots) + \gamma_0(Z_t + \lambda Z_{t-1} + \lambda^2 Z_{t-2} + \ldots) + \varepsilon_t \;;\; 0<\delta<1 \;,\; 0<\lambda<1$$

Zur Reduzierung der Parameter muß die Koyck-Transformation - siehe oben Gleichungen (8) bis (12) - jetzt zweimal angewandt werden. Im ersten Schritt wird (25) um ein Intervall zurückversetzt, mit dem Faktor 8 multipliziert und anschließend von der Ausgangsgleichung (25) subtrahiert. Dies führt zu

$$(26) \quad Y_t = \alpha(1-\delta) + \delta Y_{t-1} + \beta_0 X_t + \gamma_0[Z_t + (\lambda-\delta)Z_{t-1} + \lambda(\lambda-\delta)Z_{t-2} + \lambda^2(\lambda-\delta)Z_{t-3} + \ldots] + \varepsilon_t - \delta\varepsilon_{t-1}$$

Wenn die Koeffizienten δ und λ gleich groß sind, wenn also *Y* auf die Veränderungen in *X* mit der gleichen Trägheit wie auf Veränderungen in Z reagiert, entfallen in Gleichung (26) alle verzögerten Z-Terme; sie reduziert sich zu

$$(27) \quad Y_t = \alpha^* + \delta Y_{t-1} + \beta_0 X_t + \gamma_0 Z_t + \varepsilon_t - \delta\varepsilon_{t-1}$$

Dies ist, abgesehen von der Fehlerkomponente, genau das eingangs zitierte Modell (1) von *Stack* (1987), der aber die einschneidende Parameterrestriktion $\delta = \lambda$ überhaupt nicht expliziert und nicht deutet. Es drängt sich natürlich sofort die Frage auf,

wie denn ein Modell aussieht, in dem eine Variable (X) eine geometrisch abfallende Anpassungsreaktion in *Y* auslöst, die andere Input-Variable (Z) aber nur instantan wirkt. Ein solches Modell läßt sich am leichtesten unter Verwendung des Lag-Operators ableiten, was weiter unten erläutert wird.

Wenn die Restriktion $\delta = \lambda$ in Modell (26) nicht gelten soll, ist eine zweite *Koyck*-Transformation durchzuführen, indem man Gleichung (26) um ein Intervall zurückversetzt, sie sodann mit dem Faktor λ multipliziert und die neu erhaltene Gleichung von der Zwischengleichung (26) subtrahiert. Dies ergibt schließlich

$$(28) \quad Y_t = \alpha(1-\delta)(1-\lambda) + (\delta+\lambda)Y_{t-1} - \delta\lambda Y_{t-2} + \beta_0 X_t - \beta_0\lambda X_{t-1} + \gamma_0 Z_t - \gamma_0\delta Z_{t-1} + \eta_t^* ,$$
$$\eta_t^* = \varepsilon_t - (\delta + \lambda)\varepsilon_{t-1} + \delta\lambda\varepsilon_{t-2}$$

Dieses Modell spezifiziert ein geometrisch abfallendes Zeitprofil für die Anpassungsreaktionen von *Y* auf *beide* Input-Variablen, sieht aber einen unterschiedlich raschen Vollzug der Anpassungsreaktion vor, je nachdem, ob es sich um den *X*- oder den Z-Input handelt. Die Anpassungsparameter δ und λ sind jedoch mit der Schätzgleichung (28) nicht eindeutig zu bestimmen. Zwar lassen sich aus den Koeffizienten für X_t und X_{t-1} bzw. Z_t und Z_{t-1} Delta- und Lambda-Koeffizienten errechnen; wenn einer von ihnen bestimmt ist, stehen aber auch noch die Koeffizienten der beiden verzögerten Y-Terme zur Berechnung von δ bzw. λ zur Verfügung. Eine numerische Lösung dieses Problems (in einem iterativen Schätzverfahren) ist nur möglich, weil λ und δ nicht nur (beide) bezüglich Y_{t-1}, sondern außerdem λ (ohne δ) bezüglich X_{t-1} und δ (ohne λ) bezüglich Z_{t-1} zu optimieren sind.

Die Gleichungen dynamischer Regressionsmodelle lassen sich viel einfacher manipulieren und interpretieren, wenn sie mit Hilfe des sog. Shift-Operators *B* geschrieben werden. Da die Fachliteratur zur Zeitreihenanalyse durchgängig diesen Operator (auch "Lag-Operator" genannt) verwendet, sei er hier kurz eingeführt. Er ist wie folgt definiert:

(29) $B^p Y_t = Y_{t-p}$, $p=1,2,...$

$(1-B)Y_t = Y_t - Y_{t-1}$; $(1-B)^{-1} = (1+B+B^2+...)$

$B^0 = 1$; $B^p c = c$, $c = const.$

Mit dem Shift-Operator *B* kann gemäß den obigen Definitionen wie mit einer algebraischen Größe gerechnet werden, so daß sich das geometrische Lag-Modell (8) wie folgt reformulieren läßt:

(30)
$$\begin{aligned} Y_t &= \alpha + \beta_0 [1 + \delta B + (\delta B)^2 + (\delta B)^3 + \ldots]X_t + \varepsilon_t \\ &= \alpha + \beta_0 (1 - \delta B)^{-1} X_t + \varepsilon_t \\ &= \alpha + \frac{\beta_0}{(1 - \delta B)} X_t + \varepsilon_t \end{aligned}$$

$$(1 - \delta B)Y_t = \alpha(1 - \delta) + \beta_0 X_t + (1 - \delta B)\varepsilon_t ,$$

was exakt der Gleichung (11) entspricht. Ein Modell mit geometrischer Lag-Struktur für zwei Input-Variablen kann mit dem Shift-Operator ohne umständliche Ableitungen direkt formuliert werden:

(31)
$$Y_t = \alpha + \frac{\beta}{(1-\delta B)} X_t + \frac{\gamma}{(1-\lambda B)} Z_t + \varepsilon_t$$

$$(1-\delta B)(1-\lambda B) Y_t = \alpha^* + (1-\lambda B)\beta X_t + (1-\delta B)\gamma Z_t + (1-\delta B)(1-\lambda B)\varepsilon_t$$

Nach Ausmultiplizieren der Klammerausdrücke erhält man Gleichung (28). Setzt man in Gleichung (31) die Parameter $\delta = \lambda$, so erhält man

(32)
$$Y_t = \alpha + \frac{\beta X_t + \gamma Z_t}{(1-\delta B)} + \varepsilon_t$$

$$(1-\delta B) Y_t = \alpha^* + \beta X_t + \gamma Z_t + (1-\delta B)\varepsilon_t$$

Sie entspricht der Gleichung (27). Nun können wir auch leicht ein Modell angeben, das einen geometrischen Lag für eine der beiden Input-Variablen und eine nur instantane Wirkung für die andere spezifiziert:

$$(33) \qquad Y_t = \alpha + \frac{\beta}{(1-\delta B)} X_t + \gamma Z_t + \varepsilon_t$$

$$(1-\delta B)Y_t = \alpha(1-\delta) + \beta X_t + \gamma Z_t(1-\delta B) + \varepsilon_t(1-\delta B)$$

$$Y_t = \alpha^* + \delta Y_{t-1} + \beta X_t + \gamma Z_t - \gamma\delta Z_{t-1} + \varepsilon_t^*$$

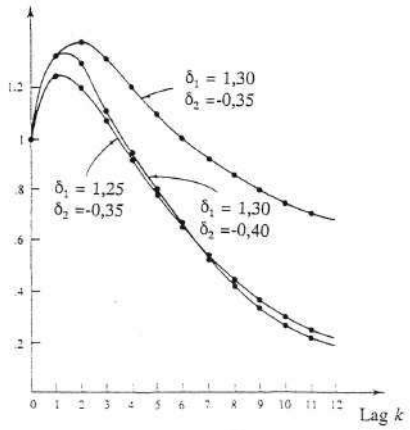
Das ausmultiplizierte Modell (33) unterscheidet sich von Modell (27) durch den gelagten Z-Term, dessen Koeffizient auf den negativen Betrag des Produkts $\lambda\delta$ restringiert ist (vergl. die Diskussion zu Modell (23)).

Bisher haben wir nur geometrische Lag-Strukturen spezifiziert. Man kann diese Modelle aber formal leicht erweitern, indem man zusätzliche verzögerte Terme der endogenen und der exogenen Variablen in die Gleichung mit einbezieht:

$$(34) \qquad Y_t - \delta_1 Y_{t-1} - \cdots - \delta_r Y_{t-r} = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \cdots + \beta_s X_{t-s} + \varepsilon_t$$

Diese allgemeine Modellform bezeichnet man, wie bereits erwähnt, auch als "rational distributed lag model". Mit ihm kann im Prinzip fast jedes beliebige Zeitprofil der Wirkung von *X* auf *Y* spezifiziert werden, indem man eine geeignete Menge von Delta- und Beta-Parametern gleich Null bzw. ungleich Null setzt. Abb. 2 zeigt drei Zeitprofile für das Modell $Y_t = \delta_1 Y_{t-1} + \delta_1 Y_{t-2} + \beta X_t + \varepsilon_t$.

Abb. 2: Zeitpfade verschiedener dynamischer Modelle 2. Ordnung:



Quelle: Rao/Miller (1971: 172); modifiziert

Die allgemeine Modellstruktur der Gleichung (34) läßt sich leicht auf zusätzliche Input-Variablen ausdehnen. Üblicherweise benutzt man dazu eine verkürzte Schreibweise, die wir anhand der Gleichung (34) hier einführen. Wenn man den Ordinatenabschnitt $\alpha = 0$ setzt und der Konvention folgt, die Koeffizienten der verzögerten Terme mit einem negativen Vorzeichen zu schreiben, wird aus (34)

$$
\begin{aligned}
(35) \qquad Y_t - \delta_1 Y_{t-1} - \cdots - \delta_r Y_{t-r} &= \beta_0 X_t - \beta_1 X_{t-1} - \cdots - \beta_s X_{t-s} + \varepsilon_t \\
(1 - \delta_1 B - \delta_2 B^2 - \cdots - \delta_r B^r) Y_t &= (\beta_0 - \beta_1 B - \cdots - \beta_s B^s) X_t + \varepsilon_t \\
Y_t &= \frac{\beta_0 - \beta_1 B - \cdots - \beta_s B^s}{1 - \delta_1 B - \cdots - \delta_r B^r} X_t + \varepsilon_t \\
Y_t &= \delta^{-1}(B)\beta(B) X_t + \delta^{-1}(B)\varepsilon_t
\end{aligned}
$$

Mit der Kurzform der Lag-Polynome, wie sie die letzte Gleichungszeile einführt, läßt sich nun das allgemeine "rational distributed lag model" für mehrere Input-Variablen in folgender Weise schreiben:

$$
(36) \qquad Y_t = \frac{\beta^{(1)}(B)}{\delta^{(1)}(B)} X_{1,t} + \frac{\beta^{(2)}(B)}{\delta^{(2)}(B)} X_{2,t} + \cdots + \frac{\beta^{(m)}(B)}{\delta^{(m)}(B)} X_{m,t} + \varepsilon_t
$$

Das Modell kann darüber hinaus erweitert werden, indem man (a) die Möglichkeit einräumt, daß die (sichtbare) Wirkung einer exogenen Variablen auf Y erst mit einer Verzögerung von $0 \le b < T$ Intervallen beginnt und (b) außerdem vorsieht, daß der Fehlerterm eine Abhängigkeitsstruktur besitzt, die sich in Form eines ARMA-Modells spezifizieren läßt:

$$
\begin{aligned}
(37) \qquad Y_t &= \frac{\beta^{(1)}(B)}{\delta^{(1)}(B)} B^{b_1} X_{1,t} + \frac{\beta^{(2)}(B)}{\delta^{(2)}(B)} B^{b_2} X_{2,t} + \cdots + \frac{\beta^{(m)}(B)}{\delta^{(m)}(B)} B^{b_m} X_{m,t} + \varepsilon_t \\
\varepsilon_t &= \frac{(1 - \theta_1 - \theta_2 - \ldots - \theta_q)}{(\phi_1 - \phi_2 - \ldots - \phi_p)} u_t \\
Y_t &= \frac{\beta^{(1)}(B)}{\delta^{(1)}(B)} B^{b_1} X_{1,t} + \frac{\beta^{(2)}(B)}{\delta^{(2)}(B)} B^{b_2} X_{2,t} + \ldots + \frac{\beta^{(m)}(B)}{\delta^{(m)}(B)} B^{b_m} X_{m,t} + \frac{\theta(B)}{\phi(B)} u_t
\end{aligned}
$$

Solche Modelle lassen sich im Prinzip mit Hilfe der von *Liu/Hanssens* (1982) vorgeschlagenen "corner method" empirisch identifizieren.

Der Gleichgewichtszuwachs G kann für jede X_k-Variable (k=l,2,...,m) gemäß

$$(38) \quad G = \frac{(\beta_0 - \beta_1 - \ldots - \beta_s)}{(1 - \delta_1 - \delta_2 - \ldots - \delta_r)}$$

bestimmt werden, wobei die Anzahl *s* der Beta-Parameter und die Anzahl *r* der Delta-Parameter für jede Input-Variable unterschiedlich sein kann.

Bei der Identifikation und Schätzung der Modelle sind neben der Stationarität der einzelnen Zeitreihen die Stabilitätsbedingungen des Systems zu beachten. Ein System ist dann stabil, wenn es nach jeder Input-Veränderung dazu tendiert, zu seinem alten oder einem neuen Gleichgewichtsniveau zurückzukehren. Die formale Bedingung hierzu ist, daß die Wurzeln des Polynoms $(1-\delta_1 B-\delta_2 B^2-\ldots-\delta_r B^r)$ alle außerhalb des Einheitskreises liegen müssen. Für ein dynamisches Modell 2. Ordnung läßt sich diese Regel in folgende Ungleichungen übersetzen:

$$\delta_1 + \delta_2 < 1$$
$$\delta_2 - \delta_1 < 1$$
$$-1 < \delta_2 < 1$$

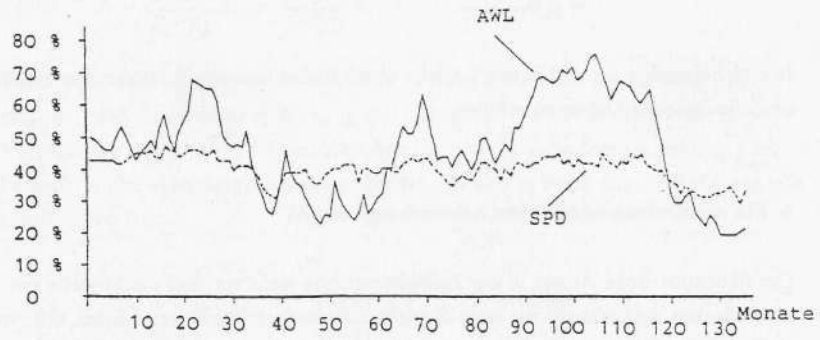
In den Beispielen der Abbildung 2 (siehe oben) finden sich nur Parameterkombinationen, die diese Bedingungen erfüllen.

6. Ein sozialwissenschaftliches Anwendungsbeispiel

Der ökonometrische Ansatz in der Zeitreihenanalyse sieht vor, daß die Modelle mit theoretischen Argumenten aus einer Substanzwissenschaft begründet werden. Oft, vor allem in den Sozialwissenschaften, liegen geeignete Theorien aber nicht vor. Der rein statistische Ansatz der Zeitreihenanalyse, wie er prominent von *Box und Jenkins* vertreten wird, hat ein Instrumentarium vorgeschlagen, mit dem die Modelle unter bestimmten Voraussetzungen auf empirische Weise "identifiziert" werden können. Voraussetzungen (vor allem die Stationarität der Reihen) und Instrumentarium können hier nicht im einzelnen erörtert werden. Ich möchte den dreistufigen Vorgang der Identifikation, der Schätzung und der Diagnose eines dynamischen Regressionsmodells ("Transferfunktionsmodells" in der *Box/Jenkins-Terminologie)* lediglich anhand eines Beispiels kurz vorstellen. Ich entnehme es einer Arbeit von *G. Kirchgässner*

(1985 b), dessen Daten ich reanalysiert habe[3]. Dabei geht es um den Zusammenhang zwischen der Einschätzung der allgemeinen wirtschaftlichen Lage (AWL) und den Parteipräferenzen (Votum für CDU oder für SPD). Die entsprechenden Zeitreihen mit monatlichen Intervallen sind demoskopischen Umfragen entnommen, die von "Infratest" (München) und dem Institut für Demoskopie Allensbach durchgeführt wurden. Sie umfassen die Periode von Februar 1971 bis April 1982 (135 Beobachtungen bzw. linear interpolierte Fälle). Ich benutze hier lediglich die Infratest-Daten zur allgemeinen wirtschaftlichen Lageeinschätzung (Anteil derer, die sie als "gut" oder "sehr gut" ansehen) und zur SPD-Präferenz (Anteil derer, die "am nächsten Sonntag" , SPD wählen würden, falls Wahlen stattfänden). Abb. 3 gibt den Verlauf der beiden Zeitreihen wieder.

Abb. 3: Präferenzanteile für die SPD sowie Einschätzung der wirtschaftlichen Lage, Februar 1971 bis April 1982



Sie läßt ausgeprägte lokale Trends erkennen. Nach einfacher Differenzenbildung beider Zeitreihen können die Stationaritätsbedingungen als erfüllt gelten[6].

---

[5] Ich danke Herrn Prof. *Kirchgässner*, der mir diese Zeitreihen mit Genehmigung des Bundespresseamtes zur Verfügung gestellt hat.

[6] Die Problematik der Differenzenbildung kann ich im Rahmen dieses Papiers nicht behandeln.

Es ist zu vermuten, daß die Einschätzung der wirtschaftlichen Lage einseitig auf die Parteipräferenz wirkt[7]. Damit ist "AWL" als Input-, "SPD" als Output-Reihe festgelegt. Das Instrument für die Identifizierung eines dynamischen Regressionsmodells ("Transferfunktionsmodells") ist die sog. Kreuzkorrelationsfunktion (KKF), die - in Stichprobennotation - wie folgt definiert ist:

$$(39) \qquad r_{yx}(k) = \frac{c_{yx}(k)}{s_y s_x} = \begin{cases} \dfrac{\frac{1}{n}\sum_{t=1}^{n-k}(x_t-\bar{x})(y_{t+k}-\bar{y})}{\sqrt{\frac{1}{n}\sum_{t=1}^{n}(x_t-\bar{x})^2\,\frac{1}{n}\sum_{t=1}^{n}(y_t-\bar{y})^2}}, & k=0,1,2,\ldots \\ \dfrac{\frac{1}{n}\sum_{t=1}^{n+k}(y_t-\bar{y})(x_{t-k}-\bar{x})}{\sqrt{\frac{1}{n}\sum_{t=1}^{n}(x_t-\bar{x})^2\,\frac{1}{n}\sum_{t=1}^{n}(y_t-\bar{y})^2}}, & k=0,-1,-2,\ldots \end{cases}$$

Die obere Gleichung behandelt die X-Reihe, die untere die Y-Reihe als bedingende Variable; der Buchstabe *k* steht für die Anzahl der "Lags" bzw. "Leads". Damit die (empirische) KKF als Instrument der Modellidenfikation verwendbar ist, muß die Input-Reihe *X* einen reinen Zufallsprozeß ("weißes Rauschen", "white noise") darstellen oder zu einem solchen transformiert werden ("prewhitening"). Der gleiche *prewhitening-Filter* ist sodann auch auf die Y-Reihe anzuwenden. Diese Schritte kann ich hier nicht darstellen (siehe die Ausführungen in *Box/Jenkins* 1976, Kap. II)[8]. Mit dem Quotienten der (geschätzten) Standardabweichungen der beiden Reihen (S_y/s_x) lassen sich die Kreuzkorrelationskoeffizienten für die verschiedenen Lags k=0,l,2,... als nicht-effiziente Schätzer der Koeffizienten $ß_k$ in dynamischen Regressionsmodellen gemäß Gleichung (6) interpretieren. Unter günstigen Voraussetzungen (so z. B. muß *X* einen genügend großen Varianzanteil in *Y* "erklären") enthüllt somit das Muster der Kreuzkorrelationskoeffizienten den zunächst nicht bekannten Modelltyp der dynamischen Beziehung der beiden Reihen.

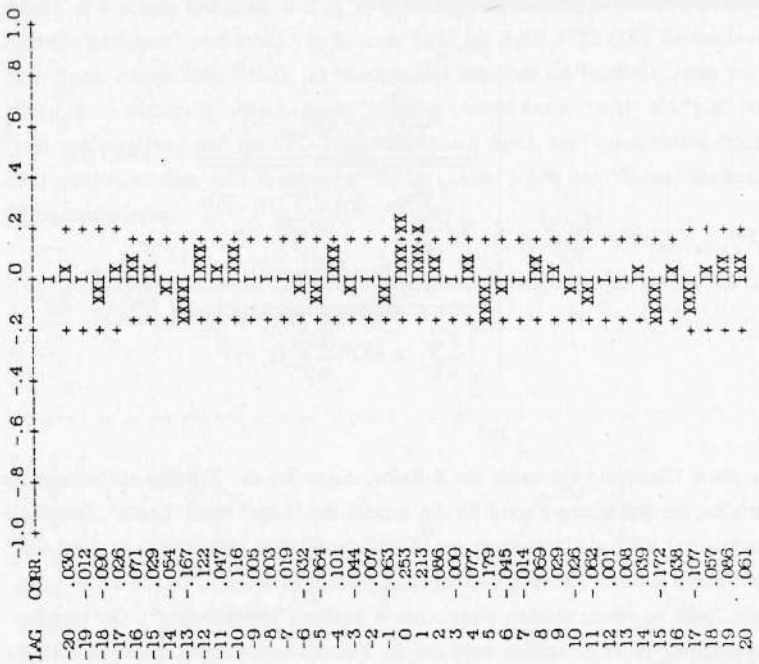
---

[7] Verfahren, mit denen derartige Hypothesen getestet werden können, stellt *Kirchgässner* (1981) vor; siehe auch die kurze Zusammenfassung in *Thome* (1988).

[8] Das bereits erwähnte Verfahren von *Liu/Hanssens* (1982) arbeitet ohne diese Form des Prewhitening.

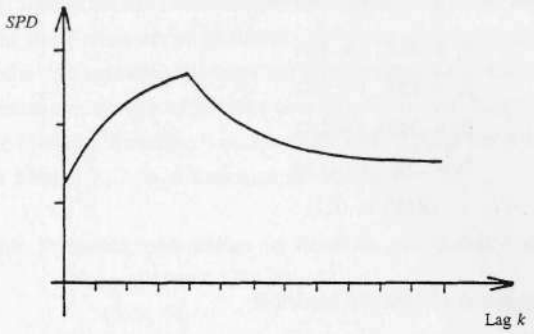
Das Kreuzkorrelogramm (die graphische Darstellung der KKF) der beiden transformierten Reihen in unserem Beispiel ist in Abb. 4 wiedergegeben.

Abb. 4: Kreuzkorrelogramm der transformierten AWL- und SPD-Reihen



Es zeigt einen signifikanten Koeffizienten bei k=0 und ein darauf folgendes Muster mit abfallenden Werten. Es läßt sich im Sinne eines dynamischen Modells 1. Ordnung (eines geometrischen Lag-Modells) interpretieren, das oben ausführlich besprochen wurde. Nun tritt aber bei "Lag" k=5 ein weiterer, diesmal negativer Koeffizient auf, der signifikant sein könnte und dem man mit etwas Phantasie ebenfalls ein abfallendes Muster zuordnen kann. Theoretisch ließe sich diese Konstellation im Sinne einer zurückschwingenden Anpassungsreaktion interpretieren: Eine positiv veränderte Einschätzung der ökonomischen Lage kommt zunächst der Regierungspartei SPD zugute. Dieser Effekt wird aber nach einer gewissen Zeit der Gewöhnung, hier nach etwa 4 Monaten, bei einem Teil der (potentiellen) Wähler wieder zurückgenommen. Eine solche Verlaufsform für den Äquilibriumszuwachs läßt sich schematisch wie in Abb. 4 a darstellen.

Abb. 4a: Muster des Zeitpfads der **SPD-Reaktion auf** einen **AWL-Step-Input**



Ein solches Modell wird durch folgende Gleichung spezifiziert:

$$(40) \quad \text{Modell A: } \nabla SPD_t = \frac{\beta_0 - \beta_5 B^5}{1 - \delta_1 B} \nabla AWL_t + N_t$$

Alternativ dazu schätzen wir ein weiteres Modell, in dem wir die negative Transferkomponente (also das Zurückschwingen der Anpassungsreaktion) ausschließen:

$$(41) \quad \text{Modell B: } \nabla SPD_t = \frac{\beta_0}{1 - \delta_1 B} \nabla AWL_t + N_t$$

In beiden Gleichungen steht N_t ("Noise") für die Fehlerkomponente, für die ebenfalls ein Modell zu spezifizieren ist, falls eine Abhängigkeitsstruktur der Fehler anzunehmen ist. Ich übernehme hier das von *Kirchgässner* identifizierte *moving-average* Modell 1. Ordnung: $N_t = \theta a_{t-1} + a_t = (1-\theta B)a_t$. Das Symbol ∇ steht für den (einfachen) Differenzenoperator: $\nabla SPD_t = SPD_t - SPD_{t-1} = (1-B)SPD_t$. Da sowohl die Input-Reihe (AWL) als auch die Output-Reihe (SPD) in gleicher Weise "differenziert" wurden, bleibt die Interpretation der Modellparameter von dieser Operation unberührt[9]. Im einzelnen werden für die beiden Modelle folgende Parameter geschätzt:

---

[9] Dies wird häufig übersehen, das heißt, die Gleichungen

(a) $Y_t = a + bX_t + e_t$

(b) $\nabla Y_t = a^* + b(\nabla X_t) + \nabla e_t$

werden so interpretiert, als formalisierten sie inhaltlich unterschiedliche Beziehungen.

Tabelle 4: Parameterschätzungen und t-Werte für Modell A

$\delta_1 = .482$, t = 4.44
$\beta_0 = .133$, t = 5.58
$\beta_5 = .043$, t = 2.42
$\theta_1 = .303$, t = 3.55
(RMS = .021)

, **Tabelle 5:** Parameterschätzungen und t-Werte für Modell B

$\delta_1 = .382$, t = 3.19
$\omega_0 = .126$, t = 4.76
$\theta_1 = .278$, t = 3.30
(RMS = .022)

Die Koeffizienten beider Modelle sind statistisch signifikant und liegen innerhalb der Stationaritätsregionen. Auch die Interkorrelationen der Parameter (hier nicht abgedruckt) sind in beiden Fällen zufriedenstellend. Die *t*-Werte lassen erkennen, daß die Parameterschätzungen für Modell A etwas stabiler sind als diejenigen für Modell B. Die RMS-Werte differieren aber nur geringfügig. Sie sind beide erheblich geringer als die RMS=0.166 für das univariate SPD-Modell.

Die Frage ist, ob sich empirisch anhand weiterer Diagnose-Kriterien entscheiden läßt, welches der beiden Modelle, A oder B, den beobachteten Daten eher entspricht. Die übliche Residuenanalyse (hier nicht dokumentiert) weist beide Modelle als akzeptabel aus mit

---

Der Steigungsparameter "b" besagt jedoch in beiden Gleichungen dasselbe. Das wird deutlich, wenn man von Gleichung (a) die inhaltsgleiche Beziehung

(a*) $Y_{t-1} = a + bX_{t-1} + e_{t-1}$

subtrahiert. Es resultiert die Gleichung (b) mit unverändertem Steigungskoeffizienten *b*. Allerdings stellt (b) eine andere Schätzgleichung als (a) dar; es verändert sich nicht nur der Ordinatenabschnitt (er wird gleich Null), sondern es verändert sich auch die Autokorrelation der Fehler. Außerdem ist darauf zu achten, daß die beiden Reihen nicht "überdifferenziert" werden (siehe ***Gottman*** 1981: 260 ff.)

leichten Vorteilen für Modell A. Der zweite entscheidende Test für die Güte der (vorläufig) identifizierten Transferfunktionsmodelle wird in der Forschungspraxis häufig ausgelassen: Die Kreuzkorrelationsfunktion zwischen den Residuen und der "vorgeweißten" Input-Reihe darf keinerlei Hinweise auf eine strukturelle Beziehung enthalten; die Abfolge der Kreuzkorrelationskoeffizienten über eine Serie von "Lags" sollte einem reinen Zufallsprozeß ("weißes Rauschen") entsprechen. Abb. 5 zeigt das entsprechende Kreuzkorrelogramm für Modell A, Abb. 6 dasjenige für Modell B.

Abb. 5: Kreuzkorrelogramm der Residuen aus Modell A mit dem "vorgeweißten" AWL-Input

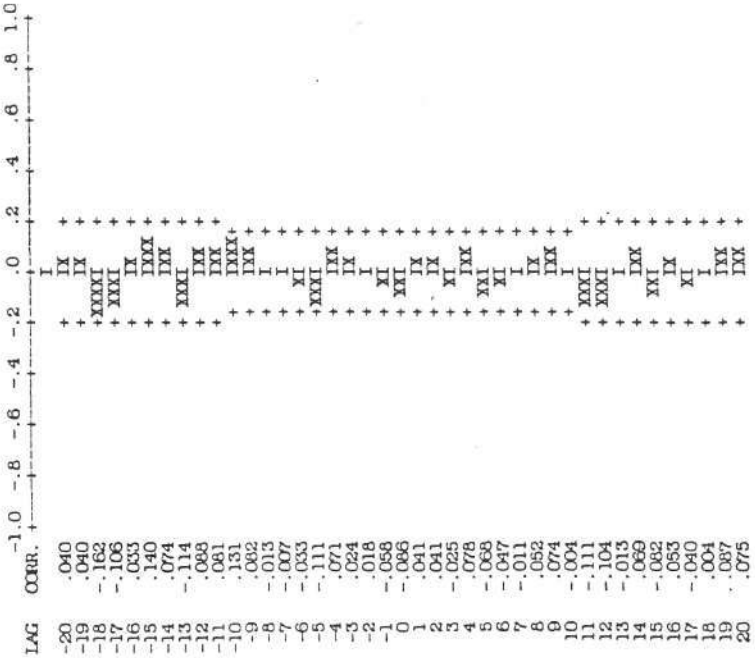
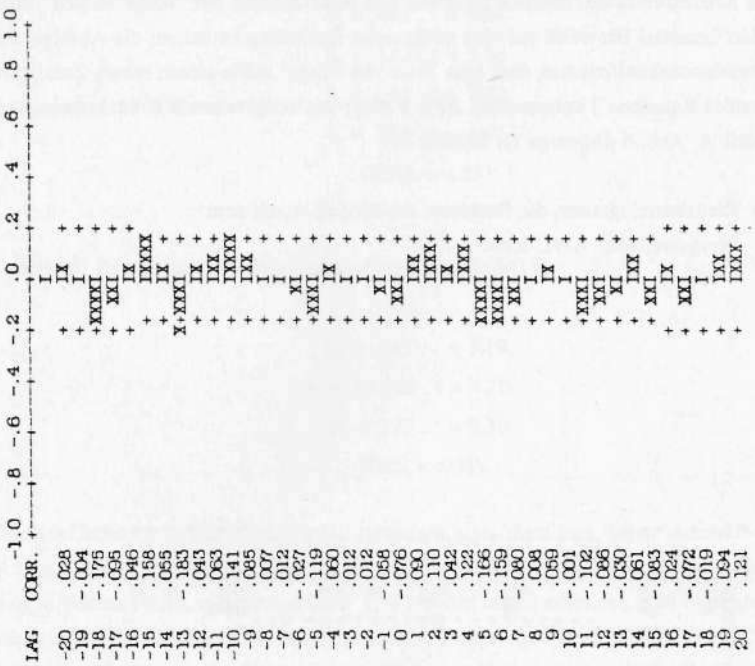


Abb. 6: Kreuzkorrelogramm der Residuen aus Modell B mit dem "vorgeweißten" AWL-Input



In Abb. 7 sind keinerlei signifikante Koeffizienten zu erkennen. Auch wenn wir nicht nur einzelne Koeffizienten, sondern Gruppen von Koeffizienten testen, ist keine statistische Signifikanz feststellbar: Analog zur *Ljung-Box* Statistik für die Autokorrelationsfunktion univariater Zeitreihen (Test auf "white noise") ist auch ein Portmanteau-Test für die KKF des "vorgeweißten" Inputs mit den Residuen des Gesamtmodells definiert:

$$(42) \qquad Q(K) = m(m+2)\sum_{k=0}^{K} \frac{1}{m-k} r_{ae}^2(k)$$

Dabei steht a für die vorgeweißte Input-Reihe, *e* für die Residuen des Transferfunktionsmodells; *m* bezeichnet die Anzahl der verwertbaren Elemente der Residualreihe. Sie ist *n-d-w-p*. Dabei entspricht *n* der ursprünglichen Zahl von Meßzeitpunkten, *p* der Zahl der autoregressiven Parameter im Noise-Modell (hier p=0), *d* gibt den Wiederholungsfaktor (den "Grad") der Differenzenbildung an (hier *d=l),* und *w* ist gleich *r* oder gleich *s+b,* je nachdem, welcher Betrag größer ist. In unserem Beispiel ist für Modell A *s* = 5 größer als

r = 1. Somit erhalten wir m = 135-1-5 = 129 effektive Fälle.[10] Klein k indiziert den jeweiligen Lag, groß K die Menge der insgesamt betrachteten Lags (Lagspanne). Die Testgröße Q ist χ^2-verteilt; die Freiheitsgrade sind mit ***df = K + 1*** *- (Zahl der gefitteten δ- und β-Paramter)* zu berechnen.

Somit ergeben sich für verschiedene Lag-Spannen die Signifikanzwerte der Tabelle 6.

Tabelle 6: Q-Statistik für die Residuen aus Modell A mit dem vorgeweißten Input

Lag k	Q	α
3	1.33	0.25
6	2.34	0.68
12	6.83	0.74

Die Nullhypothese (kein Zusammenhang) kann also beibehalten werden; Modell A ist nach diesem Kriterium akzeptabel.

Anders sieht das Kreuzkorrelogramm für Modell B (Abb. 6) aus. Es zeigt eine deutliche Musterung im positiven Lag-Bereich und dort auch signifikante Koeffizienten bei Lag k=5 und k=6. Die Q-Statistik legt ebenfalls eine Ablehnung der Nullhypothese nahe, wenn man bedenkt, daß es hier vor allem um einen möglichst geringen *Beta-Fehler,* also um einen relativ hohen *Alpha-Wert* geht (siehe Tab. 7).

Tabelle 7: Q-Statistik für die Residuen aus Model B mit dem vorgeweißten Input

Lag k	Q	α
3	3.67	0.06
6	12.90	0.01
12	16.85	0.07

[10] Ich habe hier keine Korrektur für die mir unbekannte Zahl der linear interpolierten Fälle vorgenommen.

Damit zeigt sich, daß in Modell B die Transferfunktion **nicht** adäquat spezifiziert worden ist. Modell A schneidet bei der Diagnose eindeutig besser ab.

Wie ist nun das Schätzergebnis für Modell A zu interpretieren? Sowohl die Input-Variable (AWL) als auch die Output-Variable (SPD) sind als Prozentanteile befragter Personengruppen gebildet worden. Da beide Variablen mit dem gleichen Differenzenfilter transformiert wurden, gelten die Parameter auch für die ursprünglichen Skalenwerte (Prozentgrößen) der beiden Reihen. $\beta_0 = 0.133$ besagt: wenn der Prozentanteil von Personen, die die wirtschaftliche Lage als gut oder sehr gut einschätzen um 10 Prozentpunkte ansteigt, wächst innerhalb des gleichen Monats der Anteil der SPD-Präferenzen um 1.3 Prozentpunkte. Die Reaktion auf eine **anhaltende** Niveauverschiebung in AWL setzt sich bis nach dem vierten Monat fort. Die SPD-Präferenzen erreichen bis dahin eine Zuwachsgröße, die 24 % des Niveauanstiegs in AWL entspricht.

$$(43) \quad y_4 = \beta_0(1+\delta+\delta^2+\delta^3+\delta^4) = \beta_0\frac{1-\delta^4}{1-\delta} = 0{,}133\frac{1-0{,}482^4}{1-0{,}482} = 0{,}24$$

Das heißt, ungefähr ein Viertel derjenigen Personen, die ihre Einschätzung der wirtschaftlichen Lage im angegebenen Sinne ändern, ändern zunächst (innerhalb von vier Monaten) auch ihre Parteipräferenz für oder gegen die SPD. Wenn der mit β_5 indizierte Gegeneffekt nicht einträte, ließe sich langfristig ein Äquilibriumszuwachs von

$$(44) \quad G = \frac{0{,}133}{1-0{,}482} = 0{,}26$$

ermitteln. Doch mit dem fünften Monat setzt ein Abschwungeffekt ein: einige Leute korrigieren ihre Präferenzänderung, so daß ab t=5 positive und negative Zuwachsbeträge addiert werden müssen. Für t=5 ergibt sich ein positives Inkrement von $\beta_0\delta^5 = 0.035$ und ein negatives von $\beta_5 = 0.043$: Nach t=4 beginnt die Zuwachskurve also zu fallen. Langfristig nähert sie sich einem Äquilibriumszuwachs von

$$(45) \quad G = \frac{0{,}133-(0{,}043)}{1-0{,}482} = 0{,}17$$

(vergl. Abb. 4 a). Das heißt, von 100 Personen, die ihre Einschätzung der wirtschaftlichen Lage "dauerhaft" ändern, ändern vorübergehend 26 und schließlich 17 dauerhaft auch ihre

Präferenz für oder gegen die SPD.[11] *Kirchgässner* (1985) hat die gleiche Analyse auch für die CDU-Präferenzen durchgeführt und dabei ein fast spiegelbildliches Ergebnis erhalten: Die Zahl derjenigen Personen, die aufgrund einer veränderten Einschätzung der wirtschaftlichen Lage ihre Präferenz von der SPD abziehen, ist nahezu identisch mit der Zahl der Personen, die gleichzeitig zur CDU wandern.

Daß Identifikation und Schätzung eines Transferfunktionsmodells in diesem sozialwissenschaftlichen Beispiel relativ glatt vonstatten ging, ist vor allem dem Umstand zu verdanken, daß zwischen den beiden Variablen AWL und SPD tatsächlich ein starker Zusammenhang besteht. Die Modellidentifikation gelingt nämlich nur, wenn die Varianz der Input-Reihe im Vergleich zu der Varianz der Output-Werte, die durch das "Noise" erzeugt wird, relativ groß ist oder/und eine große Menge von Meßzeitpunkten vorliegt (s. *Box/Jenkins* 1976, S. 387; vergl. *Downing/Pack* 1982, S. 235). Diese Voraussetzungen sind in den Sozialwissenschaften oft nicht erfüllt. Die Problemlage wird durch das Pre-whitening verschärft, weil durch diese Datentransformation unter Umständen der Anteil der Meßfehler an der Gesamtvarianz erhöht wird (siehe *Kirchgässner* 1981). Zusätzliche Probleme können entstehen, wenn Daten vorliegen, die bereits in irgendeiner Weise mit Verfahren außerhalb der *Box/Jenkins-Methode* transformiert wurden. Die Daten können z. B. vorgängig saisonal bereinigt worden sein. Häufig führt das zu einer Verzerrung des strukturellen Zusammenhangs der beiden Variablen.

7. Zusammenfassung

In Regressionsanalysen mit Querschnittsdaten können die geschätzten Modellparameter nur dann sinnvoll als Indikatoren kausaler Beziehungen interpretiert werden, wenn sich die Daten zum Zeitpunkt der Messung (näherungsweise) in einem Gleichgewichtszustand befinden. Zeitreihendaten und die Methoden der Zeitreihenanalyse eröffnen im Prinzip die Möglichkeit, diese restriktive, oft unrealistische Voraussetzung fallenzulassen und den "Zeitpfad" der Wirkung einer Variablen *X* auf eine Variable *Y* zu rekonstruieren. Wie dies geschieht, wurde hier ausführlich am Beispiel des geometrischen Lag-Modells erörtert. Solche "dynamischen" Modelle können im Idealfall aus einer substanzwissenschaftlichen Theorie abgeleitet werden; andererseits stellt die Zeitreihenanalyse auch Instrumente bereit, mit denen bestimmte Modellkomponenten empirisch ("aus den Daten heraus") identifiziert werden können - eine Möglichkeit, die der *"Box/Jenkins"*-Ansatz in besonderer Weise

[11] Nach den Schätzergebnissen von *Kirchgässner* sind es 15 Personen.

betont. Aber auch die empirische Modellidentifikation (die hier an einem Beispiel demonstriert wurde) kommt ohne theoretische Vorgaben nicht aus. Der Forscher muß zumindest eine Vorstellung darüber haben, welche Variablen in einer strukturellen Beziehung zueinander stehen könnten. Das empirisch spezifizierte Noise-Modell beispielsweise löst nicht das Problem der "omitted variables", das aus der Regressionsanalyse mit Querschnittsdaten bekannt ist. Es löst nur das Problem autokorrelierter Fehler, nicht das der ausgelassenen Variablen, die mit den im Modell berücksichtigten Input-Reihen korrelieren.

Wie andere formale Modelle auch, sollen dynamische Regressionsmodelle inhaltliche Theorien präzisieren und den (zweigleisigen) Transfer zwischen Theorie und Daten erleichtern. In diesem Artikel sind nur einfache Modelle mit lediglich einer abhängigen Variablen vorgestellt worden. Bezogen auf sozialwissenschaftliche Theorien sind sie häufig nicht komplex genug. Wie schon in der Querschnittsanalyse, so ist auch in der Zeitreihenanalyse in diesen Fällen von Ein- auf (simultane) Mehrgleichungsmodelle überzugehen. Sie eröffnen insbesondere die Möglichkeit, Kausalketten und Wechselwirkungen zwischen Variablen zu spezifizieren. Derartige Modelle sowie die entsprechenden Verfahrenstechniken und geeignete Software sind sowohl in der ökonometrischen Tradition als auch innerhalb des *Box/Jenkins-Ansatzes* entwickelt, bisher aber in der Soziologie kaum erprobt worden.

Literaturverzeichnis

Box, G. E. P./Jenkins, G. M., 1976: Time series analysis. Forecasting and control, San Francisco u. a.: Holden-Day (revised edition).

Cohen, L. E./Land, K. C, 1987: Age structure and crime: symmetry versus asymmetry and the projection of crime rates through the 1990s.

Devine, J. A./Sheley, J. F./Smith, M. D., 1988: Macroeconomic and social-control policy influences on crime rate changes, 1948 - 1985. American Sociological Review 53: 407 - 420.

Dhrymes, Ph. J., 1971: Distributed lags. Problems of estimation and formulation, San Francisco: Holden-Day und Edinburgh: Oliver & Boyd.

Downing, D. J./Pack, D. J., 1982: The vanishing transfer function, S. 227 - 246 in: O. D. Anderson (Hrsg.), Time series analysis: Theory and practice 1, Amsterdam: North Holland.

Gottman, J.M., 1981: Time-series analysis. A comprehensive introduction for social scientists, Cambridge u. a.: Cambridge University Press.

Hibbs, D. A., 1974: Problems of statistical estimation and causal inference in time-series regression models, S. 252 - 308, in: H. L. Costner (Hrsg.), Sociological methodology 1973-1974, San Francisco u.a.: Jossey-Bass.

Johnston, J., 1972: Econometric methods, New York u. a.: McGraw-Hill (2nd edition).

Judge, G. GJGriffiths, W. E./Hill, R. C./Lee, T.-C, 1980: The theory and practice of econometrics, New York u. a.: Wiley (Neuauflage 1985).

Kirchgässner, G., 1981: Einige neuere Verfahren zur Erfassung kausaler Beziehungen zwischen Zeitreihen, Göttingen: Vandenhoeck und Ruprecht.

Kirchgässner, G., 1985 a: Überprüfung der Hypothese rationaler Erwartungen anhand von Popularitätsdaten. Eine Untersuchung für die Bundesrepublik Deutschland, 1971 - 1982. Vortrag vor dem ökonometrischen Ausschuss des Vereins für Sozialpolitik, Rauischholzhausen, 21. 3. 1985, vorläufige Fassung.

Kirchgässner, G., 1985 b: Causality testing of the popularity function. An empirical investigation for the Federal Republic of Germany, 1971-1982. Public Choice 45: 155 - 173.

Kirchgässner, G., 1986: Economic conditions and the popularity of West German parties: A survey. European Journal of Political Research 14: 421 - 439.

Kmenta, J., 1986: Elements of econometrics, New York u. London: Macmillan (2nd edition).

Liu, L.-M./Hanssens, D. M., 1982: Identification of multiple-input transfer function modeis. Communications in Statistics A 11: 297 - 314.

Norpoth, H., 1984: Economics, politics, and the cycle of presidential popularity. Political Behavior 6: 253 - 273.

Rao, P./Miller, R. L., 1971: Applied Econometrics, Belmont: Wadsworth.

Stack, St., 1987: Publicized executions and homicide, 1950 - 1980. American Sociological Review 52 (1987): 532 -540.

Thome, H., 1988: Probleme des Identifizierens und Testens von Kausalbeziehungen in der statistischen Zeitreihenanalyse, S. 93 - 117 in: F. Meier (Hrsg.), Prozeßforschung in den Sozialwissenschaften, Stuttgart: G. Fischer.

Clusteranalyse und Korrespondenzanalyse in der Lebensstilforschung - ein Vergleich am Beispiel der Wohnungseinrichtung

von Jörg Blasius und Werner Georg[1]

Abstract

In der vorliegenden Arbeit werden zwei Verfahren vorgestellt und verglichen, die häufig in der Lebensstilforschung verwendet werden: die Clusteranalyse und die Korrespondenzanalyse. In einer von Burda (1988) in Auftrag gegebenen und von SINUS bundesweit durchgeführten Studie wurden 2516 Befragten u.a. 27 Fotografien von Wohnungseinrichtungen vorgelegt. Die Zustimmungsäußerungen zu diesen Bildern sind die Datengrundlage der vorliegenden Arbeit. Mit Hilfe multivariater Analysen wurde geprüft, ob es bestimmte Kombinationen gibt, die als "Stile" interpretiert und welchen Personengruppen welche "Stile" zugeordnet werden können. Es stellte sich heraus, daß mit beiden Verfahren ähnliche Ergebnisse ermittelt werden, die Interpretation und Handhabung der Korrespondenzanalyse jedoch einfacher ist.

The aim ofthe paper is to introduce and to compare two multivariate techniques which are both frequently used in life style research: cluster analysis and correspondence analysis. In a nationwide study (N=2516) by Burda (1988) and SINUS 27 pictures of flat interior decorations were shown. The degree of consent is the data base ofthis study. By using both mentioned techniques one wanted to check ifthere are any combinations which can be interpreted as "style", and which characteristics ofthe respondents correspond with which style. It can be shown that both techniques lead to similar Solutions, but correspondence analysis has some advantages in respect to handling and interpretation.

1. Einleitung

Wohnungseinrichtungen haben häufig eine über ihre gegenstandsbezogene Nutzung hinausweisende, auf die Demonstration von Status und Sozialprestige bezogene Funktion. Bereits um die Jahrhundertwende bezog sich Thorstein ***Veblen*** bei seiner Beschreibung der zeitgenössischen "leisure class" auf diese Funktion (vgl. ***Veblen*** 1989: 134). Die erste Skala des sozialen Status, die entwickelt wurde ***(Chapin*** 1947), beruhte auf Beobachtungen von Einrichtungen. Auch in der neueren Literatur wird die Wohnungseinrichtung zur horizontalen Unterscheidung gesellschaftlicher Gruppierungen verwendet, so **z.B.** von ***Laumann*** und ***House*** (1970), ***Pappi*** (1973), ***Pappi*** und ***Pappi*** (1978) und ***Blasius*** (1991).

1 **Dr. *Werner Georg*** ist wissenschaftlicher Mitarbeiter im Fachbereich 2 (Sozialpädagogik) an der Universität-Gesamthochschule Siegen, Adolf-Reichwein-Str. 2, D-5900 Siegen.

Im Rahmen dieser Analyse werden Bilder von Wohnungseinrichtungen verwendet, die 1987 vom SINUS-Institut innerhalb einer groß angelegten Studie zu den "Wohnwelten in Deutschland 2" ***(Burda*** 1988) insgesamt 2516 Befragten vorgelegt wurden. Mit Hilfe dieser Bilder soll geprüft werden, ob es bestimmte Kombinationen gibt, die als "Stile" interpretiert werden können und welchen Personengruppen welche "Stile" zugeordnet werden können. Für die (multivariate) Identifizierung derartiger (zu erwartender) Muster gibt es eine Vielzahl von Möglichkeiten. Wir werden uns auf die zwei beschränken, die wir für am effektivsten halten: die Clusteranalyse und die Korrespondenzanalyse. Beide Verfahren werden nach einem kurzen Überblick über die Anwendungsmöglichkeiten angewandt und die Ergebnisse einander gegenübergestellt.

Die dieser Studie zugrundeliegende Theorie, anhand derer mögliche Cluster von Bildern erwartet werden, basiert auf Arbeiten von *Pierre Bourdieu* (1982, 1983)[2]. Entgegen einer Vielzahl von anderen Theorien der "neuen sozialen Ungleichheit" (zur Übersicht vgl. den Sammelband von *Berger und Hradil* 1991) hat die Theorie von *Bourdieu* den Vorteil, daß die einzelnen Elemente operationalisiert wurden und daß empirische Überprüfungen möglich sind, sie wurden u.a. von ihm **selbst durchgeführt.**

2. Die Theorie von Bourdieu

In der Reproduktionstheorie Bourdieus gibt es explizite Hinweise auf Attribute von Einrichtungen als Distinktionsmerkmal zwischen den Angehörigen unterschiedlicher sozialer Klassen (vgl. *Bourdieu* 1982: 384 ff). *Bourdieus* Theorie, die an dieser Stelle nicht voll entfaltet werden kann, basiert auf einer Soziologie des Geschmacks. Lebensstilausprägungen, wie z.B. Präferenzen für bestimmte Wohnungseinrichtungseigenschaften (z.B. sauber, gepflegt, stilvoll) und Orte des Möbelkaufes (z.B. beim Handwerker, im Möbelhaus) werden im "Raum der Lebensstile" verortet, von dem eine Homologie zu den sozialen Lagen unterschiedlicher gesellschaftlicher Gruppen ("Raum der sozialen Positionen") angenommen wird (vgl. *Bourdieu* 1982: 212 f.). Der "Raum der sozialen Positionen" wird operationalisiert mit Hilfe der Variable "Berufspositionen" (u.a. mit den Ausprägungen "Hochschullehrer", "höhere Handelsangestellte"). Beide Räume sind letztlich identisch: Während die Dimensionen des "Raumes der Lebensstile" - z.B. die Wohnungseinrichtungseigenschaft "sauber" oder als Ort des Möbelkaufs "beim Antiquitätenhändler" - Ausprägungen in den einzelnen Berufspositionen haben - z.B. in "höhere Handelsangestellte" und "Hochschullehrer" -, haben die Dimensionen des "Raumes der sozialen Positionen" (also z.B. die "Hochschullehrer") Ausprägungen in den Lebensstilen (also u.a. in der Wohnungseinrichtungseigenschaft "sauber"). Beide "Räume"

2 Es ist strittig, ob die Arbeiten von *Bourdieu*, ebenso wie die anderer Autoren zur "neuen sozialen Ungleichheit" als "Theorie" oder lediglich als "Modell" bezeichnet werden sollten. Im klassischen (Popperschen) Sinn dürfte eigentlich nicht von einer "Theorie" gesprochen werden.

haben somit die gleichen Komponenten, sie unterscheiden sich lediglich durch die Art der Betrachtungsweise; die Prozentuierung (die Bildung von Profilen) verläuft einmal zeilen- und einmal spaltenweise. Mit Hilfe der Korrespondenzanalyse verbindet *Bourdieu* (1982) die beiden genannten Räume zum "sozialen Raum" (zur Beschreibung dieses Raumes s. *Bourdieu* 1985, zur Zusammenführung beider Räume vgl. ausführlich *Blasius* 1991).

Mittels der horizontalen Achse beschreibt *Bourdieu* (1982), ob und wie stark die Handlungsressourcen eines Akteurs dem ökonomischen und dem kulturellen Bereich zuzurechnen sind; er verwendet hierfür die Begriffe des ökonomischen und kulturellen Kapitals. Die dritte von *Bourdieu* (1983) eingeführte Kapitalart, das soziale Kapital, das im wesentlichen aktivierbare soziale Netzwerke umfaßt, wird weder von ihm empirisch belegt noch von uns in der nachfolgenden Analyse berücksichtigt.

Entsprechend der oben beschriebenen Unterscheidung differenziert *Bourdieu* zwischen oberen, mittleren und unteren Klassen, die er, anhand der Berufszuordnungen, in unterschiedliche Klassenfraktionen subdifferenziert. Diese unterschiedlichen Klassenfraktionen werden, innerhalb der für sie jeweils gültigen Klasse, anhand des kulturellen und ökonomischen Kapitals ihrer Angehörigen unterschieden. Zur groben Unterscheidung ordnet er den oberen Klassen einen "Sinn für Distinktion", den mittleren Klassen eine "Bildungsbeflissenheit" und den unteren Klassen einen "Notwendigkeitsgeschmack" zu. Die, ihn stärker interessierende, Unterscheidung von Klassenfraktionen innerhalb der drei Klassen widmet *Bourdieu* (1982) seinem Buchtitel "Feine Unterschiede". *Blasius* und *Winkler* (1989) stellen in einer empirischen Überprüfung der Bourdieuschen Theorie anhand eines Kölner Datensatzes fest, daß die feinen Unterschiede innerhalb der Klassen nur marginal nachweisbar, die "groben Unterschiede" zwischen den Klassen(fraktionen) jedoch sehr ausgeprägt sind.

Handlungstheoretisch verknüpft *Bourdieu* die Makroebene der Sozialstruktur und die Mikroebene des Akteurs über den Habitus. Mittels des Habitusbegriffes konzipiert er seine Akteure, die weder als blinde Erfüller von Regeln noch als eigenständige, autonome Konstrukteure ihrer sozialen Umgebung angesehen werden können. Dabei wird die spezifische Strukturiertheit der sozialen Umgebung der Individuen modellhaft verinnerlicht und (unbewußt) als ein generalisiertes Muster auf die unterschiedlichen Bereiche des Lebensstils übertragen.

Im Rahmen dieses Habitus-Konzepts kann die Wohnungseinrichtung als ein symbolisches "Kampffeld" der Angehörigen der unterschiedlichen Berufsgruppen (der Klassenfraktionen) um Distinktionsgewinne verstanden werden. Entsprechend der Bourdieuschen Theorie sind die spezifischen Stilausprägungen der Wohnungseinrichtungen auf die Zusammensetzung der Kapitale zurückführbar. Eine postmoderne oder avantgardistische Wohnungseinrichtung hat beispielsweise für die Angehörigen der Neuen Mittelschichten die Funktion der Abgrenzung

gegenüber gutbürgerlichen Haushalten mit ihren auf Seriosität und Gediegenheit achtenden repräsentativen Wohnungseinrichtungen.

Im Rahmen seiner empirischen Befunde zum Bereich der bevorzugten Wohnstile der unterschiedlichen Klassenfraktionen verzeichnet ***Bourdieu*** ein Ansteigen "der Optionen, die genuin ästhetische Eigenschaften herausheben (kunstvoll gestaltet, phantasievoll, harmonisch), mit steigendem Sozialstatus, während in der gleichen Richtung der Anteil der Optionen, die man 'funktionalistisch' nennen könnte, abnimmt (sauber, praktisch, pflegeleicht)" *(Bourdieu* 1982: 383).

3. Daten

Lebensstile und andere Merkmale sozialer Ungleichheit, z.B. in Form von Wohnungseinrichtungen, werden auch in der Marktforschung als Indikatoren alltagsästhetischer Präferenzen verwendet (vgl. *Becker* und *Nowak* 1982). Im Rahmen der Analyse sozialer Milieus als Prädiktoren für Marktsegmente[3] führte das SINUS-Institut eine Vielzahl von "Lebenswelt-Explorationen" durch, die teilweise durch Fotografien von Wohnungen dokumentiert wurden.

Der hier analysierte Datensatz wurde vom Burda-Verlag in Auftrag gegeben und gemeinsam mit dem SINUS-Institut in Heidelberg durchgeführt. Die Studie wurde unter dem Titel 'Wohnwelten in Deutschland 2" veröffentlicht ***(Burda*** 1988). Kernstück dieser Studie, die neben produktnahen Einstellungen auch Wertorientierungen und die Milieuzugehörigkeit (zu dem Milieukonzept des SINUS-Instituts vgl. ***Becker*** und ***Nowak*** 1982; ***Hradil*** 1987) der Befragten erfaßt, sind Wohnstil-Indikatoren, die u.a. 47 Aufnahmen von Wohnungen enthalten und 27 Fotografien von exemplarischen Wohnungseinrichtungen, das sogenannte "Leitbild-Inventar". Beide Instrumente wurden mittels einer vierstufigen Skala eingeschätzt (1: gefällt mir sehr gut; 2: gefällt mir gut; 3: gefällt mir weniger; 4: gefällt mir überhaupt nicht). Grundlage für die hier vorgestellte Analyse war das Leitbild-Inventar, dessen Variablen aus methodischen Gründen (euklidische Distanzen bei ordinalen Variablen können nicht angemessen interpretiert werden) in Zustimmung (Kategorien 1 und 2) und Ablehnung (Kategorien 3 und 4) dichotomisiert wurden.

Zwischen Mai und Juni 1987 wurden im Rahmen der Studie 2516 Personen interviewt. Das Untersuchungsgebiet war die gesamte Bundesrepublik Deutschland (alte Bundesländer einschließlich West-Berlin), befragt wurden deutschsprachige Personen im Alter von 18 bis 64

3 Der Milieubegriff ist etwas irreführend, es handelt sich hier eher um Wertorientierungen, die dem Bereich der Subkulturen zuzuordnen sind.

Jahren in Privathaushalten, die in einer mehrstufig geschichteten Zufallsauswahl (ADM-Mastersample) ermittelt wurden[4].

4. Methoden

Das Problem bei der Erforschung von Lebensstilen und Geschmacksmustern, das hier exemplarisch anhand der Wohnungseinrichtung behandelt wird, liegt zunächst in der - exploratorischen - Konstruktion von Typologien. Die statistischen Hilfsmittel, die hierfür zur Verfügung stehen, sind vor allem Faktorenanalysen, Clusteranalysen, die Korrespondenzanalyse, aber auch unterschiedliche Formen der Multidimensionalen Skaüerung. In diesem Beitrag beschränken wir uns auf die Vergleiche der Ergebnisse einer Clusteranalyse und einer Korrespondenzanalyse.

4.1 Clusteranalysen

Clusteranalysen werden in einer Vielzahl von unterschiedlichen Lebensstiluntersuchungen verwendet, so von *Lüdtke* (1989), *Gluchowski* (1988), ***Burda*** (1988) sowie von ***Conrad*** und ***Burnett*** (1990). Mit Hilfe der Clusteranalyse wird eine Vielzahl von Objekten (Personen) anhand von Merkmalen derart zusammengefaßt, daß "Personen" mit ähnlichen Merkmalen räumlich nahe beieinander liegen. Diese Personen sollen zum gleichen Cluster (zur gleichen Gruppe) gehören. Formal ausgedrückt: Die clusterinterne Varianz soll minimal, die clusterexterne Varianz soll maximal sein. Das Problem bei der Berechnung von Clusteranalysen ist zum einen die Bestimmung der Ähnlichkeit (wie wird diese definiert?) und zum anderen die Art der Zusammenfassung der Objekte zu einem Cluster (nach welchem Algorithmus erfolgt die Zuordnung zu der einen oder anderen Gruppe?).
Aufgrund der Vielzahl von Maßen, mit denen die Ähnlichkeit zweier Merkmalsträger beschrieben werden kann und aufgrund der Vielzahl von Algorithmen, nach denen die Cluster gebildet werden können, gibt es nicht *die* Clusteranalyse, sondern lediglich verschiedene Verfahren der Clusteranalyse. Zum Beispiel umfaßt das relativ häufig verwendete Programm CLUSTAN von ***David Wishart*** 40 Ähnlichkeitsmaße und 27 unterschiedliche Algorithmen (vgl. ***Wishart*** 1984, S. 12). Dabei führt in der Regel die Verwendung verschiedener Ähnlichkeitsmaße bzw. verschiedener Algorithmen auch zu unterschiedlichen Ergebnissen, was häufig den Eindruck der Beliebigkeit der Ergebnisse hinterläßt (was ohne die ausführliche Begründung für den gewählten Algorithmus und für das verwendete Ähnlichkeitsmaß auch durchaus berechtigt ist).

---

4 Der Datensatz wird unter der Studiennummer 1788 vom Zentralarchiv zur Verfügung gestellt.

Clusteranalysen können in die sehr rechenintensiven, partitionierenden und die weniger rechenintensiven, hierarchischen Verfahren unterschieden werden. Bei den hierarchischen Clusteranalysen entspricht am Beginn jedes Objekt einem Cluster. Im ersten Analyseschritt werden die zwei ähnlichsten Objekte zu einem Cluster zusammengefaßt, im zweiten Schritt jene Objekte, die einander am zweitähnlichsten sind usw. Dieser Prozeß wird solange fortgesetzt, bis alle Objekte zu einem Cluster gehören bzw. bis die vorgegebene Anzahl von Clustern erreicht wurde. Ähnlich einem Stammbaum kann dieser Vereinigungsprozeß graphisch in Form eines Dendogrammes zusammengefaßt werden (vgl. *Eckes* und *Roßbach* 1980).

Bei den partitionierenden Clusteranalysen werden die zu clusternden Objekte (Individuen) im ersten Schritt zufällig einer vorgegebenen Anzahl von Clustern zugeordnet. Mit Hilfe eines iterativen Suchprozesses werden die Objekte so lange nach einem Optimierungskriterium umgruppiert, bis ein vorgegebenes Abbruchkriterium (z.B. die clusterinterne Varianz kann nicht weiter reduziert werden) erreicht ist. Innerhalb der partitionierenden Verfahren sind insbesondere zwei Algorithmen von Bedeutung: der k-means *(MacQueen* 1967) und der hill-climbing Algorithmus *(Rubin* 1967). Ausgegangen von einer Startkonfiguration - die z.B. mittels einer wenig rechenintensiven, hierarischen Clustermethode ermittelt werden kann -, wird beim k-means Algorithmus die in der Regel euklidische Distanz eines Objektes zu den Clusterschwerpunkten (den Clustermittelwerten) berechnet. Wenn diese Distanz zu einem anderen Cluster näher ist als zum gegenwärtigen, wird das Objekt in dieses Cluster umgruppiert, die Clusterschwerpunkte werden von neuem berechnet und der iterative Prozeß beginnt von vorne. Im Gegensatz hierzu wird im hill-climbing Algorithmus nicht die euklidische Distanz eines einzelnen Objektes zu den Clusterschwerpunkten minimiert, sondern das sogenannte "Spur W-Kriterium" (vgl. *Eckes* und *Roßbach* 1980). Die Matrix W ist die Summe der Dispersions- oder Streuungsmatrizen innerhalb der Cluster, und die Spur von W ist die Summe der Diagonalelemente dieser Matrix. Eine Minimierung des Spur W-Kriteriums entspricht somit der parallelen Minimierung der clusterinternen Varianz aller Cluster.

Hinsichtlich der bereits vermerkten Vielzahl von Ähnlichkeitsmaßen und Algorithmen ist anzumerken, daß in allen Anwendungsfällen die Wahl des Verfahrens ausführlich begründet werden muß. Wenn diese Begründung nicht substanzwissenschaftlich erfolgen kann, sollten jene Verfahren verwendet werden, die, statistisch gesehen, die stabilsten Ergebnisse liefern.

Insgesamt gesehen wird bei allen Clusteranalysen der Versuch unternommen, sich einem globalen Optimum, einer optimalen Partition der zu analysierenden Daten anzunähern und dabei lokale Optima zu vermeiden. Mit Hilfe von Simulationsstudien, in denen das globale Optimum bekannt ist, wurden systematische Bias der einzelnen Verfahren untersucht. So neigt der hill-climbing Algorithmus zur Bildung gleich großer Cluster, er kommt aber dem globa-

len Optimum am nächsten (vgl. ***Bardeleben*** 1991). Er wurde daher in einer modifizierten Form (s. ***Bardeleben*** 1991) von uns verwendet[5].

4.2 Korrespondenzanalyse

Die Korrespondenzanalyse ist ein exploratives, multivariates Analyseverfahren, dessen wichtigstes Charakteristikum die graphische Darstellung von Zeilen und Spalten von zwei- und höherdimensionalen Kontingenztabellen ist. Sie kann als Skalierungstechnik angesehen werden, welche insbesondere zur Strukturierung nominal skalierter Daten verwendet wird. Das Verfahren kann auf ***Hirschfeld*** (1935) zurückgeführt werden, der als erster eine algebraische Formulierung von Zeilen und Spalten von Kontingenztabellen gab. Die erste detaillierte Beschreibung des Verfahrens, wie es derzeit verwendet wird, gaben ***Benzécri*** et al. (1973). Infolge der in den 70er Jahren nahezu ausschließlich französisch-sprachigen Publikationen blieb das Verfahren außerhalb von Frankreich lange Zeit relativ unbeachtet. Erst mit den Lehrbüchern von ***Greenacre*** (1984) und ***Lebart*** et al. (1984) begann die Diskussion des Verfahrens auch im übrigen Europa und in den Vereinigten Staaten. Mit der Aufnahme entsprechender Prozeduren durch die großen Anbieter von Statistik-Paketen (BMDP und SAS 1988, SPSS 1990) dürfte die Korrespondenzanalyse in naher Zukunft auch in den Sozialwissenschaften wesentlich stärker eingesetzt werden, als es bisher der Fall ist.

Innerhalb des Verfahrens können zwei Ansätze unterschieden werden: die einfache und die multiple Korrespondenzanalyse. Eingabedaten bei der einfachen Korrespondenzanalyse sind Kontingenztabellen, wobei i.d.R. eine zu beschreibende Variable in Abhängigkeit zu einer Vielzahl von beschreibenden gesetzt wird. In der multiplen Form ist die Eingabematrix entweder eine Indikatormatrix oder eine Burt-Matrix. Beide Verfahren basieren auf demselben Algorithmus (vgl. ***Greenacre*** 1984), d.h. mit den Prozeduren der einfachen Korrespondenzanalyse können auch multiple berechnet werden, wobei jedoch anzumerken ist, daß im multiplen Anwendungsfall die Eigenwerte "per Hand" umgerechnet werden müssen (s. hierzu ***Benzécri*** 1979, ***Greenacre*** 1984). Aufgrund dieser Ähnlichkeit der Algorithmen von einfacher und multipler Korrespondenzanalyse bezeichnen ***van*** *derHeijden* und ***de*** *Leeuw* (1989, S. 55) die multiple Korrespondenzanalyse auch als einfache Korrespondenzanalyse einer Indikator- bzw. einer Burt-Matrix.

---

5 Die nachfolgende Clusteranalyse wurde mit dem Programm CONCLUS von *Hans Bardeleben* (1991) durchgeführt. Zusätzlich zu der hier vorzustellenden exploratorischen Clusteranalyse können damit auch konfirmatorische berechnet werden (theoretisch angenommene Cluster werden vorgegeben und ihre Existenz im Datensatz wird überprüft), Variablen können gewichtet und es können simultan nach Gruppen getrennte Analysen berechnet werden.

Mit Hilfe der Korrespondenzanalyse sollen die Zustimmungen zu den einzelnen Bildern der von uns verwendeten Studie näher untersucht werden. D.h. es soll herausgefunden werden, ob es Personengruppen gibt, die, in unserem Beispiel ausschließlich charakterisiert über sozio-demographische Merkmale, bestimmte Bilder bevorzugen. Um eine Zuordnung der einzelnen Bilder zu diesen Personenmerkmalen zu finden, verwenden wir die einfache Korrespondenzanalyse. Da die Zustimmungen und die Ablehnungen zu den 27 ausgewählten Bildern komplementär zueinander sind, und da lediglich die Zustimmungen inhaltlich relevant sind (Untersuchung von Stilausprägungen), werden auch nur diese in der Analyse berücksichtigt. Verwendet wird somit eine multi-response Variable der Befürwortungen der genannten 27 Bilder, diese wurde mit den entsprechenden sozio-demographischen Variablen kreuztabelliert.

4.3 Vergleich der Verfahren

Anders als bei der Clusteranalyse gibt es bei der Korrespondenzanalyse einen eindeutigen Algorithmus, der zum Ergebnis führt. Die Berechnung des optimalen Projektionsraumes (zur Berechnung s. *Greenacre* 1984) kann an keiner Stelle vom Anwender frei gewählt werden, d.h. das Einfügen von numerischen Optionen, die einen Einfluß auf das Ergebnis haben, ist bei der Korrespondenzanalyse nicht möglich. Damit entfällt die bei der Clusteranalyse vorhandene "Willkürlichkeit" der Ergebnisse.

Auf der Ebene der Interpretation muß beachtet werden, daß es sich bei der Korrespondenzanalyse um ein Vektormodell handelt, die Clusteranalyse ist hingegen ein Distanzmodell. Wird für die graphische Darstellung der Korrespondenzanalyse die symmetrische Form verwendet - diese ist die übliche Darstellungsart in der "französischen Schule" -, so ist es nicht zulässig, die Abstände zwischen Zeilen- und Spaltenmerkmalen als Distanzen zu interpretieren, sondern lediglich zwischen den Zeilen- sowie zwischen den Spaltenmerkmalen (vgl. ausführlich ***Greenacre*** und ***Hastie*** 1987, ***Greenacre*** 1989). Sollen die Ähnlichkeiten zwischen Zeilen- und Spaltenmerkmalen dennoch direkt aus der Graphik interpretiert werden, so muß die Interpretation über die Ähnlichkeit der Winkel (ihrer Cosinusse) erfolgen (vgl. ***Blasius*** und ***Rohlinger*** 1989). Einschränkend ist dabei anzumerken, daß es bei geringer Varianzerklärung durch die graphisch dargestellten Achsen zu Fehlspezifikationen kommen kann, so daß die Interpretation mittels der numerischen Ergebnisse erfolgen sollte.

5. Ergebnisse

5.1 Clusteranalyse

In einem ersten Schritt wird die optimale Clusteranzahl bestimmt, wobei inhaltliche und formale Gründe zu beachten sind. Die inhaltliche Begründung beinhaltet das, was *Schulze* (1991) in Anlehnung an physikalische Fragestellungen das "Unschärfeproblem" nennt: Je mehr Cluster extrahiert werden, desto differenziertere Strukturen werden abgebildet und desto größer ist die Gefahr, aufgrund niedriger Clusterbesetzungen Zufallsvariation (inhaltlich) zu interpretieren. Es hängt letztlich vom inhaltlichen Interesse des Forschers ab, wieviel "Unscharfe" er in Kauf nehmen möchte: Sollen möglichst viele Cluster gewählt werden, um Feinstrukturen abzubilden, oder soll eine möglichst starke Informationsreduktion vorgenommen werden, um so wenig Cluster wie möglich unterscheiden zu müssen. Ist eine substanzwissenschaftliche Theorie vorhanden, anhand derer die Anzahl der Cluster bestimmt werden kann - und anhand derer vorhergesagt werden kann, durch welche Variablen die einzelnen Cluster beschrieben werden können -, so entfällt die schwierige Entscheidung über die angemessene Anzahl von Clustern. In vielen empirischen Anwendungen gibt es jedoch keine explizite Theorie, so daß lediglich eine explorative Bestimmung der optimalen Clusterzahl vorgenommen werden kann. Auch im vorliegenden Fall war uns keine Theorie bekannt, anhand derer die Anzahl der Cluster hätte bestimmt werden können.

Um auf dem explorativen Weg eine angemessene Clusterzahl zu ermitteln, wurden in einem ersten Schritt sämtliche Lösungen von zehn bis zwei Clustern berechnet. Bei der Interpretation der einzelnen Ergebnisse wurde deutlich, daß es dominante Muster des Geschmacks gab, deren Unterformen, die bei Lösungen von sechs bis zehn Clustern ermittelt wurden, nur wenig zusätzliche Information beinhalteten. Wir entschieden uns anhand dieser Ergebnisse für die 5-Cluster-Lösung, mit Hilfe dieser konnten die dominanten Geschmacksmuster am klarsten differenziert werden. Im folgenden wird der Prozentsatz an Zustimmung zu den 5 Clustern wiedergegeben.

Tabelle 1: Prozentsatz an Zustimmung in den fünf Clustern

	Zuordnungen zu den fünf Clustern					Anteil der
	C 1	C 3	C 3	C 4	C 5	Zustimmungen
B1	51.3	93.3	85.4	19.9	88.6	70.9
B2	23.1	76.9	81.2	20.4	45.0	49.7
B3	7.8	7.4	75.7	27.2	4.6	18.7
B4	9.8	6.4	63.6	60.6	3.5	19.1
B5	23.5	82.7	64.7	7.3	91.3	59.4
B6	31.8	92.3	75.1	22.0	86.3	65.8
B7	15.5	82.7	71.8	80.8	3.4	31.8
B8	12.0	55.5	59.9	13.5	17.5	31.0
B9	40.3	90.6	64.9	25.6	82.0	65.0
B10	17.9	75.4	71.4	17.4	32.7	43.4
B11	12.6	11.7	64.1	43.6	2.4	21.3
B12	36.3	62.2	71.2	77.7	11.1	47.6
B13	7.5	7.8	71.5	43.8	1.7	20.2
B14	49.2	64.2	73.2	69.9	15.5	50.4
B15	22.2	17.4	64.0	73.8	4.8	30.0
B16	42.3	49.1	66.6	48.8	19.4	42.0
B17	35.2	57.7	70.5	95.5	13.2	48.0
B18	38.2	82.9	72.8	54.4	48.3	59.4
B19	14.4	82.7	65.7	10.5	63.9	53.7
B20	19.3	49.7	51.1	61.0	4.8	37.9
B21	6.7	3.0	58.8	28.3	1.4	13.5
B22	9.2	9.4	64.5	46.0	2.3	19.9
B23	13.1	19.4	77.0	70.0	3.3	28.1
B24	27.4	64.8	75.5	30.9	30.3	43.2
B25	37.1	75.4	69.6	24.8	43.1	53.2
B26	47.2	66.4	70.8	56.2	29.8	54.6
B27	22.7	72.7	49.1	28.4	28.2	44.1
N	454	652	318	418	674	2452 - 2477

Cluster 1 umfaßt 454 Befragte, denen keine eigene Stilpräferenz zugeordnet werden kann. Vielmehr handelt es sich bei diesem Cluster um ein "Ablehnungscluster", alle Bilder werden überdurchschnittlich oft abgelehnt (vgl. die Spalten "Cl" und "Anteil der Zustimmungen"). Diese Ablehnung gilt insbesondere für die Bilder 5, 6 und 19, auf denen bürgerlich-konventionelle Einrichtungsstile (Sitzgruppen mit Tisch, Schrankwand) abgebildet sind. Dem zweiten Cluster können 652 Befragte zugeordnet werden, wobei eine relativ deutliche Struktur von Zustimmung und Ablehnung zu den meisten der vorgegebenen Bilder festgehalten werden kann. Zustimmung finden insbesondere konventionelle, klassische und z.T. ins Repräsentative gehende Wohnungseinrichtungen, wie sie auf den Bildern 1, 5, 6, 7, 9, 18 und 19 wiedergegeben sind[6]. Relativ oft abgelehnt werden dementgegen die Bilder 3, 4, 13, 21 und 22;

6 In den Tabellen und in der graphischen Abbildung als B1, B2,..., B27 gekennzeichnet.

die abgebildeten Einrichtungen können als postmodern, funktional bzw. avantgardistisch bezeichnet werden. Die 318 Personen, die durch das dritte Cluster charakterisiert werden, stimmten nahezu allen der vorgegebenen Bilder überdurchschnittlich oft zu: Sollen dennoch einige Einrichtungsstile hervorgehoben werden, so sind insbesondere die Zustimmungen zu den Bildern 2, 3, 4, 7, 13, 14, 15, 23 und 24 zu nennen. Diese Personen haben demnach Präferenzen für moderne Einrichtungsstile, die z.T. durch nostalgische Elemente ergänzt werden. Das vierte, 418 Personen beinhaltende Cluster, ist durch eine deutliche Struktur von Zustimmungen und Ablehnungen gekennzeichnet. Präferiert wurden insbesondere die Bilder 4, 7, 12, 15, 17, 20 und 23, abgelehnt wurden die Einrichtungen auf den Bildern 1, 5, 6, 9 und 19. Während auf der ersten Gruppe von Bildern vor allem klar strukturierte Räume mit modernen Einrichtungen dargestellt sind, beinhaltet die zweite Gruppe konventionelle, rustikale Einrichtungen (Sitzgruppe aus Leder, Schrankwand aus Holz, rustikale Küche). Das fünfte Cluster umfaßt 674 Befragte, bei lediglich zwei relativ häufigen Zustimmungen (zu den Bildern 5 und 6) dominieren die Ablehnungen (der Bilder 7, 11, 12, 13, 14, 16, 17, 20, 21, 22, 23, 26 und 27). Während eine konventionell gemütliche Wohnungseinrichtung präferiert wird, wird nahezu alles Moderne abgelehnt.

Im theoretischen Teil dieses Beitrages wurden Wohnungseinrichtungen als Indikatoren für soziale Differenzen behandelt. Im folgenden soll die Zusammensetzung der fünf Geschmackstypen nach sozialen Kontextmerkmalen untersucht werden. Da das Hauptanliegen dieser Arbeit im Methodenvergleich und nicht in einem umfassenden Test der Bourdieuschen Theorie liegt, beschränken wir uns auf zwei Variablen, von denen eine direkt aus seiner Theorie abgeleitet werden kann. Zum einen wird die Bildung[7], gemessen über den höchsten Schulabschluß, als Merkmal des kulturellen Kapitals verwendet. Ähnlich wie *Bourdieu* den einzelnen Klassen und Klassenfraktionen unterschiedliche Geschmacksausprägungen zuweist, müßte eine eindeutige Differenzierung auch bei Verwendung der unterschiedlichen Bildungsabschlüsse nachweisbar sein. Zum anderen verwenden wir eine Variable, bei der im Hinblick auf die Stilisierung des Wohnbereiches eine Differenzierung zu erwarten ist: das Geschlecht. Für die Beschreibung der Cluster nach den beiden genannten Variablen wurden Kreuztabellen zwischen den fünf Clustern und dem "Geschlecht", sowie der "Bildung" der Befragten berechnet. Die Ergebnisse sind in Tabelle 2 wiedergegeben.

7 Die stärker interessierende Variable "Berufsposition" wurde lediglich über die - heutzutage nur noch wenig aussagende - Variable "Stellung im Beruf operationalisiert.

Tabelle 2: Beschreibung der 6 Cluster nach Geschlecht und Bildung, Angaben in Prozent, C = (korrigierter) Kontingenz-Koeffizient

	Zuordnungen zu den fünf Clustern C1	C2	C3	C4	C5	insgesamt
Geschlecht:						
männlich	44.7	41.6	49.1	48.3	47.0	1150
weiblich	55.3	58.4	50.9	51.7	53.0	1366
N	454	652	318	418	674	2516
Chi^2=7.8, df=4, n.s.						
Bildung:						
Hauptschule	45.8	60.9	51.6	29.7	66.0	1338
weiterf. Schule	33.5	25.8	28.6	29.4	25.7	706
Abitur	11.0	6.4	8.2	20.6	5.0	239
Hochschule	9.7	6.9	11.6	20.3	3.3	233
N	454	652	318	418	674	2516
Chi^2=248.0, df=12, p.<.001, C=.34						

Auffällig ist, daß die Variable "Geschlecht" keinen Effekt auf die Clusterzugehörigkeit hat, d.h. diesen Ergebnissen zufolge gibt es keine geschlechtsspezifischen Stilausprägungen. Dementgegen hat die "Bildung des Befragten" einen deutlichen Effekt: So bevorzugen die Volksschüler überdurchschnittlich häufig die traditionellen Wohnungseinrichtungen, wie sie durch die Cluster 2 und 5 repräsentiert werden. Hingegen bevorzugen Abiturienten und Hochschulabsolventen relativ oft moderne Varianten der Wohnungseinrichtung, bei gleichzeitiger Ablehnung konventioneller, traditioneller Stile (Cluster 4). Bei den Absolventen der weiterführenden Schulen können keine spezifischen Merkmale hinsichtlich der Präferenzen für die Wohnungseinrichtung nachgewiesen werden.

5.2 Korrespondenzanalyse

Mit Hilfe der Clusteranalyse wurden im ersten Schritt die Bilder zu Clustern zusammengefaßt, und im zweiten Schritt wurde die Variable der Clusterzugehörigkeiten mit "Bildung" und "Geschlecht" kreuztabelliert. Mit diesem Vorgehen konnten lediglich den Bildungsgruppen unterschiedliche Stilausprägungen zugeordnet werden.

Im folgenden sollen mit Hilfe der einfachen Korrespondenzanalyse die "27 Bilder" nach den oben genannten Variablen differenziert werden. Dafür wurde eine aus "Geschlecht" und "Bil-

düng" kombinierte Variable (mit den acht Ausprägungen "Hauptschulabschluß", "weiterführende Schule", "Abitur" und "Universitätsabschluß", kreuztabelliert mit "Geschlecht") mit den dichotomisierten Zustimmungen zu den 27 Bildern kreuztabelliert. Die absoluten Werte der positiven Nennungen in diesen Tabellen sind untereinandergeschrieben das Eingangsmaterial der Korrespondenzanalyse, d.h. die Matrix der Ausgangsdaten hat 27 Zeilen (die Zustimmungen zu den Bildern) und 8 Spalten (die Ausprägungen der sozio-demographischen Variable). Um die Ergebnisse besser nachvollziehen zu können, ist die Matrix dieser Eingabedaten in Tabelle 3 wiedergegeben.

Tabelle 3: Matrix der Eingabedaten

INPUT DATA

A	MHS	MWS	MABI	MUNI	WHS	WWS	WABI	WUNI
B1	449	216	65	59	613	259	44	45
B2	288	145	52	57	391	183	45	61
B3	103	53	26	32	124	72	21	25
B4	100	59	35	34	115	73	28	27
B5	403	173	56	40	544	189	29	22
B6	371	203	65	67	557	252	54	52
B7	145	98	67	61	150	134	60	61
B8	162	90	30	33	276	117	21	34
B9	395	195	59	63	577	231	40	35
B10	250	139	42	46	360	157	36	42
B11	109	75	36	45	111	83	36	30
B12	208	158	74	84	317	203	70	62
B13	101	74	27	33	114	81	36	31
B14	266	174	64	62	362	203	63	53
B15	142	95	56	59	158	117	52	61
B16	224	147	58	52	293	167	46	52
B17	238	163	75	77	305	195	73	63
B18	313	193	71	81	449	234	64	63
B19	336	169	46	47	488	179	35	27
B20	182	126	57	57	257	152	47	57
B21	70	51	19	20	81	55	16	22
B22	94	76	34	42	110	74	28	32
B23	135	105	46	57	156	110	42	43
B24	220	125	43	52	352	178	52	45
B25	298	162	59	54	434	219	45	43
B26	309	175	63	66	407	220	56	50
B27	218	122	49	55	345	191	53	54

Die Korrespondenzanalyse basiert auf einer Chi-Quadrat-Metrik (vgl. ***Greenacre*** 1984, ***Blasius*** und ***Rohlinger*** 1989). Um zu prüfen, ob überhaupt genügend Variation in den Daten ist, so daß eine weitergehende Analyse Sinn macht, oder ob lediglich "weißes Rauschen" analysiert wird, wird der Chi-Quadrat-Wert für die Eingabematrix berechnet[8].

In unserem Beispiel ist $\chi^2=890$ mit 182 Freiheitsgraden. Auch wenn der Wert nicht als exakter Verteilungswert interpretiert werden sollte und er schon aufgrund der hohen Fallzahl signifikant sein dürfte, so kann doch daraus abgelesen werden, daß es Zusammenhänge zwischen Zeilen- und Spaltenmerkmalen gibt, die nicht als Randomvariation interpretiert werden können. Die graphische Darstellung der Ergebnisse der Korrespondenzanalyse ist in Abbildung 1 wiedergegeben.

Wie bereits erwähnt, ist es nicht zulässig, die Abstände zwischen Zeilen- und Spaltenmerkmalen als Distanzen zu interpretieren (vgl. ***Greenacre*** und ***Hastie*** 1987, ***Greenacre*** 1989). Möglich ist es jedoch, sowohl die Winkel zum Achsenkreuz als auch die Projektionen der Vektorendpunkte auf die Achsen zu interpretieren. Bei der Projektion der Spaltenmerkmale auf die erste Achse, mit der 89,4% der Gesamtvariation erklärt werden, fällt auf, daß diese durch unterschiedliche Bildungsabschlüsse gekennzeichnet ist. Die zweite Achse, mit der weitere 4,3% der Gesamtvariation erklärt werden und die aufgrund ihres geringen Varianzanteiles als nahezu bedeutungslos angesehen werden muß, kann mit Hilfe von geschlechtsspezifischen Geschmacksausprägungen beschrieben werden. Dies entspricht auch den Ergebnissen der Clusteranalyse, wo bereits festgestellt wurde, daß lediglich die Variable "Bildung" einen signifikanten Effekt auf die Clusterzugehörigkeit hatte.

Wird die erste Achse interpretiert, so wird der Gegensatz von 'Volksschulabsolventen" (MHS, WHS - diese Variablenausprägungen laden im positiven Bereich) und den "Abiturienten" und "Hochschulabsolventen" (MABI, MUNI, WABI, WUNI - diese laden im negativen Bereich) deutlich. Bei einer Zuordnung der Projektionen der "Bilder" auf die erste Achse können den Hochschulabsolventen und Abiturienten insbesondere jene modernen und avantgardistischen Einrichtungen zugeschrieben werden, die auf den Bildern 7, 11, 15, 22 und 23 (helle, sparsam möblierte Räume mit "klaren Formen") abgebildet sind. Dementgegen äußerten sich die Hauptschulabsolventen überdurchschnittlich oft zu den rustikalen, konventionellen Einrichtungen auf den Bildern 1, 5, 9 und 19 (Schrankwände mit Sitzgruppen, rustikale Küche) zustimmend. Die Absolventen weiterführender Schulen (MWS, WWS) sind nicht eindeutig zuzuordnen - ihr Geschmack liegt zwischen beiden Gruppierungen und kann nicht eindeutig beschrieben werden.

---

8 Dieser Chi-Quadrat-Wert, dividiert durch das Gesamt-N der Eingabedaten, i.d.F. also die Summe aller Zustimmungen aller Personen, wird als Gesamtträgheitsgewicht bezeichnet - er ist gleich der Summe der Eigenwerte, die zu den Eigenvektoren des Projektionsraumes gehören (vgl. *Blasius* und *Rohlinger* 1989, S. 100).

Abbildung 1: Graphische Darstellung der Ergebnisse

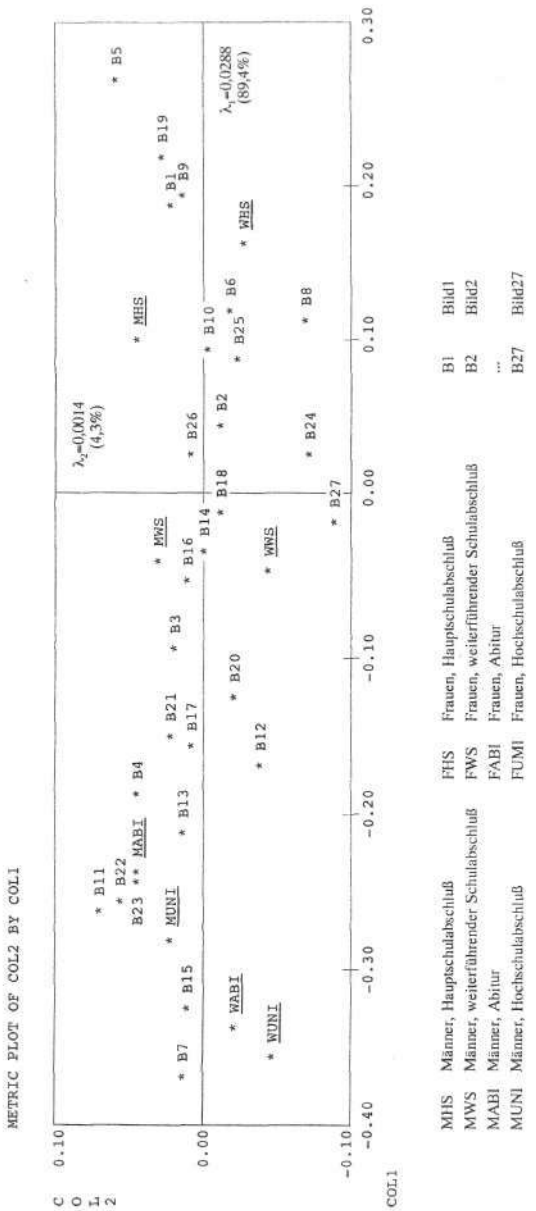


Tabelle 4: Numerische Darstellung der Ergebnisse der Korrespondenzanalyse

GENERAL STATISTIC

GENSTAT	MASS	SQCOR	INR	LOC1	QCOR1	INR1	LOC2	QCOR2	INR2
MHS	0.222	0.947	0.087	0.099	0.774	0.075	0.047	0.173	0.349
MWS	0.129	0.487	0.024	-0.044	0.323	0.009	0.031	0.164	0.091
MABI	0.050	0.932	0.104	-0.246	0.898	0.105	0.048	0.034	0.083
MUNI	0.052	0.938	0.146	-0.290	0.933	0.152	0.023	0.006	0.019
WHS	0.305	0.991	0.254	0.161	0.965	0.274	-0.027	0.026	0.157
WWS	0.156	0.793	0.026	-0.049	0.443	0.013	-0.043	0.350	0.214
WABI	0.043	0.948	0.167	-0.343	0.945	0.176	-0.021	0.004	0.014
WUNI	0.043	0.926	0.192	-0.362	0.910	0.196	-0.048	0.016	0.072
B1	0.063	0.988	0.072	0.189	0.974	0.079	0.023	0.014	0.023
B2	0.044	0.332	0.009	0.044	0.312	0.003	-0.011	0.020	0.004
B3	0.016	0.648	0.008	-0.099	0.620	0.006	0.021	0.028	0.005
B4	0.017	0.929	0.023	-0.196	0.877	0.023	0.048	0.052	0.028
B5	0.053	0.985	0.129	0.272	0.935	0.135	0.063	0.051	0.152
B6	0.059	0.991	0.026	0.117	0.966	0.028	-0.019	0.026	0.015
B7	0.028	0.964	0.127	-0.374	0.963	0.137	0.014	0.001	0.004
B8	0.028	0.802	0.018	0.110	0.574	0.012	-0.069	0.228	0.096
B9	0.058	0.984	0.069	0.195	0.979	0.076	0.014	0.005	0.008
B10	0.039	0.886	0.013	0.096	0.885	0.013	-0.003	0.001	0.000
B11	0.019	0.961	0.047	-0.268	0.896	0.047	0.072	0.065	0.072
B12	0.043	0.905	0.045	-0.172	0.864	0.044	-0.037	0.041	0.043
B13	0.018	0.868	0.032	-0.222	0.864	0.031	0.014	0.003	0.003
B14	0.045	0.359	0.005	-0.036	0.358	0.002	-0.001	0.000	0.000
B15	0.027	0.957	0.094	-0.329	0.956	0.101	0.011	0.001	0.002
B16	0.038	0.560	0.006	-0.053	0.532	0.004	0.012	0.028	0.004
B17	0.043	0.960	0.037	-0.163	0.958	0.040	0.008	0.003	0.002
B18	0.053	0.359	0.001	-0.010	0.156	0.000	-0.012	0.202	0.005
B19	0.048	0.982	0.072	0.216	0.966	0.078	0.029	0.017	0.028
B20	0.034	0.918	0.021	-0.132	0.895	0.021	-0.021	0.023	0.011
B21	0.012	0.760	0.012	-0.158	0.745	0.010	0.023	0.015	0.004
B22	0.018	0.934	0.042	-0.262	0.889	0.042	0.059	0.045	0.044
B23	0.025	0.963	0.052	-0.250	0.932	0.054	0.046	0.032	0.038
B24	0.039	0.824	0.009	0.024	0.076	0.001	-0.075	0.748	0.157
B25	0.048	0.863	0.015	0.091	0.814	0.014	-0.022	0.049	0.017
B26	0.049	0.356	0.003	0.024	0.320	0.001	0.008	0.036	0.002
B27	0.039	0.920	0.011	-0.019	0.038	0.000	-0.090	0.882	0.230

Die zweite Achse kann, wenn überhaupt, durch geschlechtsspezifische Merkmale beschrieben werden. Dabei kann den Männern keines der 27 Bilder als für sie typisch zugeschrieben werden, bei den Frauen sind nostalgische Stilelemente zu nennen, wie sie auf den Bildern 24 und 27 dargestellt sind (die dazugehörigen Ausprägungen liegen, wie auch die Ausprägungen der "weiblichen" Bildungsvariablen, im negativen Abschnitt der zweiten Achse). Um zu einer detaillierteren Auswertung der Ergebnisse zu kommen, werden die numerischen Lösungen verwendet. Diese sind in Tabelle 4 wiedergegeben.

Wie bereits aus der graphischen Abbildung ersichtlich wurde, werden auf der ersten Achse die unterschiedlichen Bildungsabschlüsse abgebildet. Während die Variablenausprägungen "Absolventen weiterführender Schulen: männlich, weiblich" nur schwach auf der ersten Achse laden (vgl. die Spalte QCOR1), besteht ein Gegensatz zwischen den Hauptschulabsolventen sowie den Abiturienten und Hochschulabsolventen (vgl. die Spalten LOC1 und QCOR1, zur Interpretation der einzelnen Koeffizienten siehe ausführlich *Blasius* 1988). Mit Hilfe der Zuordnung über die quadrierten Faktorladungen (QCOR1) können die meisten der auf den Fotografien abgebildeten Wohnungseinrichtungen einer der beiden "(Bildungs-)Gruppierungen", nicht aber anhand des Geschlechtes zugeordnet werden.

Verwenden wir als (willkürlich gesetzten) Schwellenwert zur Zuordnung eines Bildes zur ersten Achse den Wert QCOR1=.50, d.h. durch die erste Achse werden 50% der Gesamtvariation dieser Zeilenvariablen(ausprägung) erklärt, so können dem positiven Abschnitt, und damit der Gruppe der Volksschulabsolventen, folgende Bilder zugeordnet werden: 1, 5, 6, 8, 9, 10, 19 und 25. Alle diese Bilder beinhalten eher konventionelle, rustikale, aber auch gut bürgerliche Wohnungseinrichtungen (vom Ledersofa über Sitzgarnituren mit Wohnzimmertisch bis hin zu englischen Möbeln). Im Gegensatz hierzu können den Abiturienten und Hochschulabsolventen folgende Bilder zugewiesen werden: 3, 4, 7, 11, 12, 13, 15, 16, 17, 20, 21, 22 und 23. Diese Einrichtungsbeispiele können als modern, avantgardistisch bzw. postmodern beschrieben werden (von sparsam möblierten bis hin zu avantgardistisch eingerichteten Zimmern).

Wird zusätzlich der Trägheitskoeffizient interpretiert (INR1), so kann festgestellt werden, daß letztlich nur wenige Zeilenvariablenausprägungen einen deutlichen Einfluß auf die geometrische Ausrichtung der ersten Achse haben. Wird als (wiederum willkürlich gesetzter) Schwellenwert für die Zuordnung INR 1=0.075 gewählt, d.h. durch diese Variablenausprägungen werden mindestens 7.5% der Variation der ersten Achse erklärt, so sind die Bilder 1, 5, 9 und 19 den Hauptschulabsolventen und die Bilder 7 und 15 den Abiturienten und Hochschulabsolventen zuzuweisen (zu den Bildern 5, 7, 9 und 15 siehe Abbildung 2).

Auf der zweiten Achse werden, wenn auch längst nicht so deutlich wie für die erste Achse geschildert, die geschlechtsspezifischen Differenzen bezüglich der Wohnungseinrichtung ersichtlich (vgl. die Spalte LOC2), wobei der einzig interpretierbare Gegensatz bei den Haupt- und Realschulabsolventen aufgezeigt werden kann (vgl. die Spalten QCOR2 und INR2). Wird eine Zuordnung zu den Bildern vorgenommen, so haben die Männer anscheinend überdurchschnittlich oft eine Präferenz für Bild 5, während die Frauen die Bilder 8, 24 und 27 überdurchschnittlich oft angaben (vgl. die Spalten LOC2 und INR2, die Bilder 5, 24 und 27 sind in Abbildung 2 wiedergegeben).

6. Diskussion

Mit der vorliegenden Arbeit wurden zwei Verfahren vorgestellt und verglichen, die häufig in der Lebensstilforschung verwendet werden: die Clusteranalyse und die Korrespondenzanalyse. Auch wenn die durchgeführten Analysen nur in einem sehr eingeschränkten Bereich, bezogen auf lediglich zwei differenzierende Variablen (höchster Schulabschluß und Geschlecht), erfolgten, so sollte die generelle Logik der Verfahren und deren unterschiedliche Interpretationsmöglichkeiten deutlich werden.

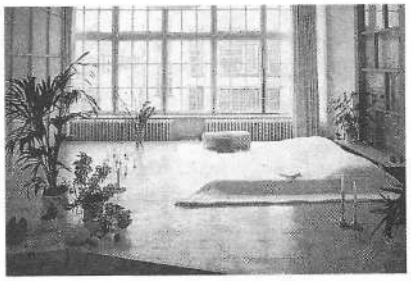
Mit der Clusteranalyse wurden fünf Cluster extrahiert, von denen insbesondere das vierte Cluster die unterschiedlichen Bildungsgruppen differenzierte. Die mit diesem Verfahren nachgewiesene Zuordnung der Abiturienten und Hochschulabsolventen zu postmodernen bis avantgardistischen Wohnungseinrichtungsstilen konnte mittels der Korrespondenzanalyse bestätigt werden. Im einzelnen konnten alle Bilder, die zur Beschreibung des vierten Cluster verwendet wurden (4, 7, 12, 15, 17, 20, 23), auch mit Hilfe der Korrespondenzanalyse den Bessergebildeten zugeordnet werden. Nicht ganz so eindeutig konnten die "Hauptschulabsolventen" einem der fünf Cluster zugeordnet werden. Sie werden, verglichen mit der Differenzierung beim vierten Cluster, lediglich leicht überdurchschnittlich oft durch die Cluster 2 und 5 repräsentiert. Werden die Merkmale dieser Cluster mit den Ergebnissen der Korrespondenzanalyse verglichen, so kann festgehalten werden, daß fünf der sieben Bilder, die zur Beschreibung des zweiten Cluster verwendet werden (1, 5, 6, 9,19), und die beiden Bilder, die für Cluster 5 charakteristisch sind (5 und 6), (auch) mit Hilfe der Korrespondenzanalyse den Hauptschulabsolventen zugeordnet wurden. Grob gesehen wurden mit beiden Verfahren den Angehörigen der gleichen Bildungsgruppierungen ähnliche Stilausprägungen zugeordnet. Bezüglich des Geschlechtes konnte mit der Korrespondenzanalyse lediglich bei den Frauen ein eigener - und auch nur tendenziell vorhandener - Einrichtungsstil identifiziert werden, mittels der Clusteranalyse konnten keine geschlechtsspezifischen Stilrichtungen angegeben werden.

Abbildung 2: Exemplarische Wohnungseinrichtungen

W5

W7

W9



W15

W24

W27

Werden die Ergebnisse auf die Theorie von ***Bourdieu*** zurückbezogen, so kann ein Gegensatz der Wohnungseinrichtungsstile nachgewiesen werden, und zwar ein moderner, avantgardistischer (bei den Bessergebildeten) sowie ein konservativ, traditioneller (bei den Hauptschulabsolventen). Die Stilrichtungen der klassischen Moderne (Bauhaus) und der Postmoderne der höheren Klassen - diese wurden in unserem sehr stark verkürzten Beispiel ausschließlich über die Bessergebildeten operationalisiert - können als "Distinktionsgewinne" im Sinn von Bourdieu angesehen werden. Weniger eindeutig ist die Identifikation des von ***Bourdieu*** postulierten "Notwendigkeitsgeschmacks" der unteren Klassen - jene wurden in unserem Beispiel ausschließlich über den Hauptschulabschluß operationalisiert -, deren Angehörige traditionelle, konservative Einrichtungen präferieren. Beachtet werden muß bei der Interpretation auch, daß bei den verwendeten multivariaten Verfahren relative Unterschiede in den Nennungen abgebildet werden - und die Hauptschulabsolventen haben überdurchschnittlich oft konservative, traditionelle Wohnungseinrichtungsstile präferiert. Dies bedeutet nicht, daß die Mitglieder dieser Gruppierung diese Einrichtungen gegenüber allen anderen möglichen Einrichtungen präferieren, sondern lediglich, daß sie diese Einrichtungen gegenüber den anderen zur Auswahl stehenden überdurchschnittlich oft nannten. Des weiteren gehören zu den von ihnen überdurchschnittlich oft genannten Stilen gerade jene Einrichtungsgegenstände, die - nach Bourdieu - am ehesten als "praktisch" und "funktional" angesehen werden können, z.B. im Sinn von "Ordnung zu machen" und "Stauraum" zu haben (z.B. die Schrankwand).

Weder mit der Clusteranalyse noch mit der Korrespondenzanalyse konnte den Absolventen weiterführender Schulen ein eigener Stil zugewiesen werden, ihre Präferenzen liegen zwischen beiden Gruppierungen. Zu einem ähnlichen Ergebnis kamen auch *Blasius* und *Winkler* (1989) bei ihrer - auf Köln bezogenen - Beschreibung von unterschiedlichen Klassenfraktionen mit Hilfe von Lebensstilmerkmalen: auch sie konnten in ihrer Analyse den "mittleren Klassen" keinen eigenen Stil zuweisen.

Während mit Hilfe der Korrespondenzanalyse die abgebildeten Wohnungseinrichtungen den Angehörigen der unterschiedlichen Bildungsgruppen direkt zugeordnet und auf diese Weise "geclustert" werden konnten, mußte bei der Clusteranalyse der Umweg über die Zusammenfassung der Variablen (der Bilder) zu Clustern erfolgen. Der Nachteil der Clusteranalyse hegt, zusätzlich zu der Problematik, die mit der Festlegung der Clusterzahl sowie der Wahl einer geeigneten Metrik und eines geeigneten Algorithmus verbunden ist, darin, daß die Interpretation über die einzelnen Cluster erfolgt und somit umständlicher und fehleranfälliger ist.

Literatur

Bardeleben, H. (1991), Soziale Determinanten sexuellen Verhaltens. Eine empirisch-methodische Studie. Gießen, unveröffentlichte Dissertation.

Becker U. und *H. Nowak* (1982), Lebensweltanalyse als neue Perspektive der Meinungs- und Marketingforschung. In E.S.O.M.A.R. Kongreß, Band 2,247-267.

Benze'cri, J.P. et collaborateurs (1973), L' Analyse des Données. L'Analyse de Correspondence. Paris.

Benzécri, J.P. (1979), Sur le Calcul des Taux d'Inertie dans 1'Analyse d'un Questionnaire. In: Cahiers de 1'Analyse des Données, 4,377-378.

Berger PA. und *St. Hradil,* Hrsg. (1991), Lebenslagen, Lebensläufe, Lebensstile. Sonderband 7 der Soziale Welt. Göttingen.

Blasius, J. (1988), Zur Stabilität von Ergebnissen bei der Korrespondenzanalyse. In: ZA-Information, Heft 23, 46-61.

Blasius, J. (1991), Gentrification und Lebensstile. Eine empirische Untersuchung in einem Kölner Stadtteil. Köln, unveröffentlichte Dissertation.

Blasius, J. and *H. Rohlinger* (1989), "KORRES - A Program for Multivariate Analysis of Categorical Data frorn Contingency Tables". In: SAS Institute Inc. (eds.): SEUGI'89, Proceedings of the SAS Users Group International Conference. Gary, N.C., 98-117.

Blasius, J. und *J. Winkler* (1989), "Gibt es die "feinen Unterschiede"? Eine empirische Überprüfung der Bourdieuschen Theorie". In: Kölner Zeitschrift für Soziologie und Sozialpsychologie, 72-94.

Bourdieu, P. (1982), Die feinen Unterschiede. Frankfurt/Main.

Bourdieu, P. (1983), Ökonomisches Kapital, kulturelles Kapital und soziales Kapital. In: *R. Kreckel* (Hrsg.): Soziale Ungleichheiten, Sonderband 2 der Soziale Welt, 183-198.

Bourdieu, P. (1985), Sozialer Raum und Klassen. Leçon sur Leçon. Frankfurt/Main.

Burda (1988), Wohnwelten in Deutschland 2. Offenburg.

Chapin, &F. (1947), Experimental Designs in Sociological Research. New York.

Conrad M. and *L. Burnett* (1991), Life Style Research 1990. Frankfurt/Main und Heidelberg, unveröffentlichter Forschungsbericht, Band 1.

Eckes, T. und *H. Rosbach* (1990), Clusteranalysen. Stuttgart.

Gluchowski, P. (1988), Freizeit und Lebensstile. Erkrath.

Greenacre, M. (1984), Theory and Applications of Correspondence Analysis. London.

Greenacre, M. (1989), "The Carroll-Green-Schaffer Scaling in Correspondence Analysis: A Theoretical and Empirical Appraisal", Journal of Marketing Research, 26,358-365.

Greenacre, M. and *T. Hastie* (1987), "The Geometrie Interpretation of Correspondence Analysis", Journal of American Statistical Association, 82,437-447.

Hirschfeld, H.O. (1935), "A Connection between Correlation and Contingency", Proceedings of the Cambridge Philosophical Society, 31,520-524.

Hradil, *St.* (1987), Sozialstrukturanalyse in einer fortgeschrittenen Gesellschaft. Leverkusen.

Laumann, E.O. and *J.S. House* (1970), Living Room Styles and Social Attributes: The Patterning of Material Artifacts in a Modern Urban Community. In: Sociology and Social Research, 54, 321-342.

Lebart, L., A. Morineau, wAK. M. Warwick (1984), Multivariate Descriptive Statistical Analysis: Correspondence Analysis and Related Techniques for Large Matrices. New York.

Lüdtke H. (1989), Expressive Ungleichheit. Opladen.

MacQueen, J. (1967), Some Methods for Classification and Analysis of Multivariate Observations. In: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, 1,281-297.

Pappi,F.U. (1973), Sozialstruktur und soziale Schichtung in einer Kleinstadt mit heterogener Bevölkerung. In: Kölner Zeitschrift für Soziologie und Sozialpsychologie, 25,23-74.

Pappi,F.U. und /. *Pappi* (1978), Sozialer Status und Konsumstil. Eine Fallstudie zur Wohnzimmereinrichtung. In: Kölner Zeitschrift für Soziologie und Sozialpsychologie, 27,293-311.

Rubin, J. (1967), Optimal Classification into Groups: An Approach for Solving the Taxomony Problem. In: Journal of Theoretical Biology, 15,103-144.

Schulze, G. (1991), Die Transferfunktion sozialer Milieus in der Bundesrepublik Deutschland. In: *P.A. Berger* und *St. Hradil* (Hrsg.), 409-432.

van der Heijden,P.G.M. and J, *de Leeuw* (1989), "Correspondence Analysis, with Special Attention to the Analysis of Panel Data and Event History Data". In: C.C. Clogg (ed.): Sociological Methodology 1989. Oxford, 43-87.

Veblen, T. (1989), Theorie der feinen Leute. Frankfurt/Main.

Wishart, D. (1984), CLUSTAN Benutzerhandbuch. Stuttgart.

Workshop: Einführung in die Korrespondenzanalyse vom 3.-6.11.1992

Gemeinsam bieten das Zentrum für Umfragen, Methoden und Analysen und das Zentralarchiv für empirische Sozialforschung vom 3. bis 6. November 1992 in Mannheim eine Einführung in die Korrespondenzanalyse an. Die Veranstaltung soll Anfängern über Demonstrationen und praktischen Übungen am PC einen Einstieg in dieses Analyseverfahren ermöglichen. An drei der insgesamt vier Veranstaltungstagen wird das von ***M. J. Greenacre*** entwickelte Programm SimCA verwendet. Am vierten Tag kommt das an der Universität Leiden entwickelte SPSS-Modul "Categories" zum Einsatz. Nach einer kurzen Einführung in den Algorithmus wird anhand inhaltlicher Beispiele die Anwendung der Korrespondenzanalyse diskutiert.

Organisiert und betreut wird der Workshop von ***Jürgen H.P. Hoffmeyer-Zlotnik*** (ZUMA) und *Jörg Blasius* (ZA) unter Mitarbeit von ***Herbert Matschinger*** vom Zentralinstitut für seelische Gesundheit in Mannheim. Der Veranstaltungsort ist Mannheim. Für die Teilnahme wird ein Beitrag von DM 80.- erhoben. Die Teilnehmerzahl ist auf 20 Personen begrenzt. Interessenten werden gebeten, sich bis zum 25. August 1992 bei
ZUMA, TagungsSekretariat, Postfach 122155, D-6800 Mannheim 1, anzumelden.

Tagung der Arbeitsgruppe: Prognoseverfahren

Die Deutsche Gesellschaft für Operations Research e.V. Arbeitsgruppe: Prognoseverfahren wird in Verbindung mit dem Zentralarchiv für empirische Sozialforschung an der Universität zu Köln ihre 28. Sitzung am 22. und 23. Oktober 1992 im Zentralarchiv für empirische Sozialforschung in Köln abhalten. Ein zentrales Thema dieser Sitzung wird sein: **Structural** Time Series Models. Dazu wird Dr. *Siem Koopman* von der London School of Economics and Political Science am 22. Oktober 1992,14.00 Uhr, einen Überblicksvortrag halten. *Dr. Koopman* ist Mitarbeiter (Research Officer) von ***Prof. Andrew Harvey*** und Koautor des PC-Programms "Structural Time Series Analyser, Modeller and Predictor (STAMP)".

Zusätzlich zu den Mitgliedern der Arbeitsgruppe wendet sich die Tagung an empirisch und methodisch orientierte Sozialwissenschaftler, Ökonomen und Wirtschaftshistoriker.

Interessenten werden gebeten sich mit **Dr. *Rainer* *Metz***, Zentralarchiv für empirische Sozialforschung, Abt. ZHSF, Bachemerstr. 40, 5000 Köln 41 (Tel. 0221-4769436; FAX 0221-4769444) in Verbindung zu setzen.

Erleichterung computergestützter Inhaltsanalysen durch verbesserte automatische Texterkennung

von Peter Hauptmanns und Rainer Schnell

Abstract

Ein Hauptproblem bei EDV-gestützter Inhaltsanalyse war bisher der Aufwand, der mit der Dateneingabe verbunden war. Neue Texterkennungssoftware zusammen mit preiswerten und leistungsfähigen Personal Computern erlauben nun automatische Texterfassung mit sehr kleinen Fehlerquoten auch für sozialwissenschaftliche Forschungsinstitute mit kleinem Etat.

Data input was one ofthe main problems of computer-assisted content analysis. New character recognition programs and low-priced powerful personal computers now permit high quality automated text recognition even for low budget research institutes.

Die aufwendige Dateneingabe für EDV-gestützte Inhaltsanalysen war bisher das technische Haupthindernis für einen breiteren Einsatz dieser Methode. Auch die in der Regel sehr kleinen Stichproben an untersuchten Texten sind eine Konsequenz des Eingabeaufwandes. Zumeist wird die Dateneingabe manuell durchgeführt, obwohl entsprechende maschinelle Techniken verfügbar sind. So gab *Weber* (1985:44) an, daß 100.000 Worte in 4 1/2 Stunden bei 0.1% Fehlern maschinell erfaßt werden könnten, *Züll/Mohler/Geiss* (1991:45) geben 50.000-200.000 Zeichen pro Stunde an. Allerdings sind die Preise für solche Geräte prohibitiv: *Züll/Mohler/Geiss* (1991:44) geben Preise von 30.000-40.000 DM an. Diese Kosten liegen aber außerhalb der finanziellen Möglichkeiten fast aller sozialwissenschaftlicher Forschungsinstitute.

Die technische Weiterentwicklung und der Preisverfall bei leistungsfähigen Personal Computern hat in Verbindung mit neuer Software diese Situation grundlegend geändert: Maschinelle Datenerfassung ist für Texte nunmehr mit weit geringeren Investitionen realisierbar. Möglich ist dies z.B. mit Programmen wie WORD-SCAN-PLUS, RECOGNITA oder OMNIPAGE. Um den praktischen Einsatz für die Dateneingabe bei Inhaltsanalysen beurteilen zu können, haben wir mit dem Programm WORDSCAN-PLUS der US-Firma CALERA einige Tests durchgeführt.

Neben dem Programm (ca. 700 US-$) und einem Flachbettscanner (ca. 2000 DM) wird nur ein dem derzeitigen Stand der Technik entsprechender PC (80486 CPU, 4 MB-RAM, 200 MB-Platte, VGA) samt Systemsoftware (DOS, Windows) benötigt: Die vollständige Ausrü-

stung liegt bei ca. 10.000 DM. Dies ist für fast jeden Institutsetat im Rahmen des Erträglichen, zumal die Ausrüstung darüberhinaus auch für die sonst üblichen PC-Anwendungen (Textverarbeitung, Datenanalyse, Datenverwaltung, Desk-Top-Publishing) einsetzbar ist.

Die Fehlerrate bei Texten von drucktechnisch guten Vorlagen lag bei den Versuchen im Mittel bei 1,7 Fehlern pro 100 Worte. Die meisten dieser Fehler (ca. 75%) ließen sich durch den Spelling-Checker des verwendeten Textverarbeitungsprogramms finden und beseitigen (WORDSCAN schreibt neben ASCII direkt auch eine Reihe von Textsystemformaten, u.a. MS-WORD und WORDPERFECT). Die Schrifttype der Textvorlage erwies sich dabei als weitgehend irrelevant für die Güte der Ergebnisse; eher von Bedeutung ist die Schriftgröße - bei Schriften kleiner als 8 Pkt. nimmt die Zahl der Fehler leicht zu. Das Programm muß nicht für jeden Buchstaben trainiert werden (wie beispielsweise das ältere SPOT von Flagstaff Engineering); z.B. konnte die englische Abstract-Seite der Kölner Zeitschrift für Soziologie und Sozialpsychologie mit nur zwei Fehlern pro Seite erfaßt werden. Die Fehler traten nur bei den kursiv gesetzten Namen der Autoren auf. Neben Zeitschriften, Zeitungsartikeln, Büchern, Faxausdrucken und Schreibmaschinentexten wurde auch ein Matrix-Drucker-Output (für diesen existiert ebenfalls ein spezieller Erfassungsmodus) gescannt und erfaßt [1]. In allen Fällen waren die Ergebnisse mehr als akzeptabel.

In einer "Testreihe" wurden acht Seiten aus verschiedenen, für Sozialwissenschaftler relevanten Quellen gescannt und vom Programm bearbeitet. Die Textvorlagen waren von unterschiedlicher drucktechnischer Qualität; weiterhin wiesen alle getesteten Seiten verschiedene Textformatierungsmerkmale (unterschiedliche Schriftgrößen, Fettschrift, Kursivschrift, Unterstreichungen o.a.) auf. Bei den Vorlagen (1),(4) und (6) wurden mehrere (zwei bis drei) Versuche unternommen, um die optimale Scannereinstellung zu ermitteln (Helligkeit, Kontrast). Bei allen anderen Vorlagen wurde mit den Default-Einstellungen der Software gearbeitet. In der Regel erbringt ein längeres Experimentieren mit den Einstellungsmöglichkeiten Ergebnisverbesserungen.

Bei den Tests zeigten sich durchweg sehr gute Ergebnisse (siehe Tabelle 1); nur bei dem Versuch, eine Tageszeitungsseite mit schlechter Druckqualität zu erfassen, war die Fehlerrate relativ hoch. Allerdings wurden auch in diesem Fall 84% der Fehler durch den Spelling Checker gefunden. Nennenswerte Probleme gab es allein bei dem Versuch, eine Seite aus einem ZfS-Aufsatz von *Diekmann* (1990:269) einzulesen. Diese Seite enthält eine Reihe von Formeln mit griechischen Buchstaben, die vom Programm nicht korrekt erkannt wurden[2]. Außerdem wurden die Bruchstriche in den Formeln nicht als Textbestandteil interpre-

1 Nur der Vollständigkeit halber sei erwähnt, daß kein Programm existiert, das Handschriften erkennen kann. Das Programm eignet sich ebenfalls nicht zur Erfassung von Antworten in Fragebögen.

2 Allerdings waren diese Fehler systematisch: das "alpha" wurde immer als "a", das "lambda" immer als "k" erkannt.

tiert[3], so daß in diesem Fall Nachkorrekturen notwendig waren. Dieser Text wurde daher aus dem Test genommen. Texte dieser Art sind zur Zeit anscheinend nur mit größerem Aufwand maschinell erfaßbar.

Bei den übrigen Texten sind fast alle aufgetretenen Fehler auf Unsauberkeiten im Druck zurückzuführen[4]. Mehrspaltige Texte (wie z.B. bei der ZfS-Seite) werden vom Programm problemlos gelesen. Bei der Speicherung im ASCII-Format erfolgt die Trennung der Spalten über eine gleiche Anzahl von Leerzeichen. Bei der Speicherung im Format eines Textverarbeitungsprogrammes, daß zur Mehrspaltenverarbeitung fähig ist, erfolgt die Trennung der Spalten durch die entsprechende Formatierung.

Tabelle 1: Ergebnisse des Scan-Tests

Text	Worte	Fehler *	Fehleranteil	Bearbeitungs-zeit (sec.)
Zeitschrift für Soziologie	748	2 (2)	0,25 %	195
Kölner Zeitschrift für Soziologie und Sozialpsychologie	420	6 (2)	1,42 %	160
Buchvorlage (Sans Serif)	218	(-)	0 %	91
Frankfurter Rundschau	269	25 (21)	9,29 %	375
Fotokopie (Times Roman)	260	1 (1)	0,38 %	86
Offset-Druck (Times Roman)	258	7 (7)	2,71 %	244
ZUMA-Nachrichten	269	3 (2)	1,11 %	134
ZA-Information	308	3 (1)	0,97 %	99

* Durch den Spelling Checker (MS-WORD) erkannte Fehler

3 Senkrechte oder waagerechte Linien in der Vorlage werden bei der Texterfassung vom Programm als "Grafik" interpretiert und ignoriert.

4 Häufige Fehler waren z.B. die Erkennung des Buchstaben "m" als "rn" oder des Buchstaben "1" als "1".

Neben der Eingabe von Texten für die Inhaltsanalyse ist ein solches System vielseitig einsetzbar. So lassen sich ohne größere Mühe veröffentlichte Programme einlesen und dann kompilieren, wie Versuche mit abgedruckten FORTRAN-Programmen (die Vorlage bestand aus den Programmen von Lee 1980 und Loehlin 1987) und PASCAL-Programmen zeigten. Das Hauptproblem bei PASCAL-Programmen lag im Pointer-Symbol ("^"). Bei den drucktechnisch meist schlechteren FORTRAN-Programmen lag das Hauptproblem bei der Verwechslung von "J" mit")" und "I" mit "J". Programmtexte lassen sich mit den heutigen Compilern rasch von solchen Fehlern säubern. Die Zahl der Fehler in Programmtexten lag mit 2-3 Fehlern pro Seite deutlich unter der Fehlerzahl, die durch programmierunkundige Datentypisten bei Programmen erzielt wird.

Ebenso von praktischem Interesse ist das Einlesen veröffentlichter Daten. Für rein numerisches Material existiert ein spezieller Erfassungsmodus, der hier für publizierte Korrelationsmatrizen und Datentabellen verwendet wurde. Eine Korrelationsmatrix (15*15) läßt sich nach dem Scannen in ca. 30 Sekunden in einen ASCII-File umwandeln, die Fehlerquote liegt bei ca. 1%. Dabei sind die meisten Fehler doppelte Vorzeichen oder Dezimalpunkte. Tabellendaten (ohne Dezimalzeichen, keine Vorzeichen) wurden nahezu fehlerfrei gelesen. Bei Daten mit Vorzeichen und Dezimalpunkt liegt die Fehlerquote ebenfalls bei ca. 1%. Der häufigste Fehler liegt in der Verwechslung eines Dezimalpunkts mit einem Dezimalkomma. In keinem einzigen Fall wurden zwei Zahlen miteinander verwechselt. Durch zweifaches Scannen desselben Materials können die Fehler rasch gefunden werden. In der Regel lassen sich mehr als 400 Dezimalzahlen pro Minute (ca. 1 Tabellenseite) so erfassen, mit Korrektur in ca. 2 Minuten.

Die Fehlerrate hängt sehr stark vom verwendeten Scanner und der Güte der Vorlage ab. So sind die Graphik-Fileformate "PCX" und "TIF", wie sie z.B. von Handscannern geschrieben werden, als Eingabe möglich. Allerdings konnten bei Versuchen mit Handscannern (eingesetzt: LOGITECH SCANMAN) nur für Zahlen bei jeder Vorlagengüte und für Texte auf außergewöhnlich gutem Material akzeptable Resultate erzielt werden. Mit einem Flachbett-Scanner (eingesetzt: HEWLETT PACKARD SCANJET+) lassen sich auch bei schlechten Vorlagen (Fotokopien, Faxe) gute Ergebnisse erzielen. Gegebenenfalls kann man durch einen modernen Kopierer vor dem Scannen die Qualität der Vorlage aufbessern. Die Geschwindigkeit des Einscannens lag bei ca. 20-30 Sekunden pro Seite, die Dauer der Texterkennung lag im Mittel bei ca. 2,5 Minuten pro Seite (auf einem PC 486-33). Dieser Wert variiert allerdings stark in Abhängigkeit von der Qualität der Vorlage (zwischen 1:26 Min. und 6:06 Min.). Da das Programm die Möglichkeit bietet, eine Liste von gescannten Files ohne interaktiven Eingriff abzuarbeiten, spricht nichts gegen das Scannen durch Hilfskräfte (so einfach wie Fotokopieren) und dem Erfassen als Batch. Die zusätzlich zur Hard- und Software entstehenden Personalkosten sind also sehr gering.

Literatur

Diekmann,A. (1990):
Der Einfluß schulischer Bildung und die Auswirkungen der Bildungsexpansion auf das Heiratsverhalten. In: ZfS, Jg. 19, H. 4, S. 265-277.
Lee, E.T. (1980):
Statistical Methods for Survival Data Analysis. Belmont/California (Wadsworth).
Loehlin, J.C. (1987):
Latent Variable Models: An Introduction to Factor, Path, and Structural Analysis. Hillsdale (Lawrence Erlbaum).
Weber, R.P. (1985):
Basic Content Analysis. Beverly Hills (Sage).
Züll, C./Mohler, P.Ph./Geis, A. (1991):
Computerunterstützte Inhaltsanalyse mit TEXTPACK PC. Stuttgart (Gustav Fischer)

Politischer Umbruch und Kriminalitätsentwicklung: Kongreß in Budapest

Die International Society for Criminology mit Sitz in Paris veranstaltet in Budapest vom 22. bis 27. August den 11. Internationalen Kriminologischen Kongreß. Kongreßsprachen sind Deutsch, Englisch, Französisch, Spanisch und Ungarisch. Der Kongreß wird organisatorisch in Plenarveranstaltungen, Vertiefungssitzungen und kleineren Arbeitsgruppen sowie mit ergänzenden Exkursionen durchgeführt werden. Das Generalthema lautet: Sozialer und politischer Umbruch und Kriminalitätsentwicklung - Eine Herausforderung auf dem Weg ins Dritte Jahrtausend -.

Die Plenarveranstaltungen widmen sich folgenden zentralen Unterthemen: (1) Grenzüberschreitende Strukturen der Wirtschaftskriminalität; (2) Umweltkriminalität; (3) Staats- bzw. Regierungskriminalatät und Korruption; (4) Terrorismus und Widerstandsbewegungen.

Für Vertiefungsveranstaltungen und Kleingruppen sind bisher rund 50 aktuelle Einzelthemen vorgeschlagen worden. Interessenten wenden sich bitte an:

NKG-Verbindungsstelle, Corrensstr. 34,
7400 Tübingen, Tel. 07071-292931.

Buchhinweise

Datenverarbeitung und statistische Auswertung mit SAS
Band I: Einführung in das Programmsystem, Datenmanagement und Auswertung

SAS-Version 6 (Großrechner: Version 6.06; PC Versionen 6.03 und 6.04)

Johannes Gogolok, Dr. Rudolf Schuemer, Dr. Gerhard Ströhlein, Fernuniversität Hagen

1992, XH, 787 Seiten, 17 x 24 cm, DM 86,-, Fischer Verlag, ISBN 3-437-40223-4

Inhalt: Überblick - Grundlagen - Einrichten, Speichern und Weiterverarbeitung von S AS-Dateien - Formate, Informate, Variablen und Werteetiketten - Datenmodifikation und -Selektion - Funktionen in SAS - Weitere Anweisungen zum Programmieren in SAS - Aufruf von Prozeduren und Anweisungen im PROC-Step - Ausgabesteuerung und Prozeduren zum Drucken sowie zur grafischen Gestaltung - Statistikprozeduren - Hinweise auf Änderungen bei den Statistikprozeduren in Band II

SAS ist ein bekanntes Programmsystem zur statistischen Datenanalyse, das die Bearbeitung sehr unterschiedlicher Aufgabenstellungen in vergleichsweise einfacher Form erlaubt. Es bietet umfangreiche Möglichkeiten zur Datenerfassung und Datenverwaltung, zur Datenanalyse, zur Berichterstattung und zur präsentationsreifen Aufbereitung von Ergebnissen.

Das vorliegende Buch ist der erste Band einer zweibändigen Beschreibung des SAS-Systems. Dieser Band bezieht sich im wesentlichen auf das Leistungsangebot des SAS Base Moduls, der zweite Band auf das Angebot des SAS/STATS Moduls. Das Base Modul stellt den Grundbaustein des SAS-Systems dar. Entsprechend umfaßt es die grundlegenden Anweisungen und Prozeduren für die Arbeit an den Daten und mit dem SAS-System selbst. Darüber hinaus enthält das Base Modul Prozeduren zur Durchführung einfacher deskriptiver statistischer Analysen sowie Prozeduren für die Erstellung von Berichten und Tabellen. Für kompliziertere multivariate statistische Auswertungen stellt das SAS/STAT Modul die erforderlichen Prozeduren zur Verfügung.

Über das Angebot des Base Moduls hinaus werden im vorliegenden ersten Band verschiedene Prozeduren aus dem SAS/STAT Modul (ausnahmslos aus dem Bereich Varianzanalyse) behandelt. Hinweise auf aktuelle Änderungen der übrigen Statistikprozeduren des SAS/STAT Moduls ergänzen den zweiten, bereits 1990 erschienenen Band II. Weitere Hinweise betreffen das Editieren von Daten und die Erstellung von interaktiven Dateneingabemasken mit dem Modul SAS/FSP sowie die Erstellung von Graphiken mit dem Modul SAS/GRAPH.

Der Aufbau der Darstellung ist im wesentlichen an technischen Aufgabenstellungen orientiert. Die verschiedenen Kapitel werden (meist) durch eine kurze Inhaltsangabe eingeleitet.

Die Darstellung der einzelnen Anweisungen beinhaltet in der Regel eine kurze Erläuterung ihrer Funktion, Angaben zur Syntax, Übersichten der verfügbaren Kommandooptionen, illustrierende Beispiele sowie Hinweise zur Anwendung. Seitengenaue Verweise auf die umfangreiche Originaldokumentation zum SAS-System erleichtern die gezielte Beschaffung weiterführender Informationen.

Für Umsteiger aus anderen Programmsystemen und etwas weiter fortgeschrittene Anwender bietet das Buch viele Arbeitshilfen. Dennoch läßt das Buch - wie bei der Komplexität des SAS-Systems angesichts dauernder Programmänderungen auch nicht anders zu erwarten ist - einige Wünsche offen: Die Gliederung könnte noch konsequenter aufgabenbezogen sein und sich stärker an der für das SAS-System elementaren Trennung von PROC und DATA Steps orientieren. Das durchaus umfangreiche Stichwortverzeichnis könnte noch weiter vervollständigt werden. Verbesserungsfähig wäre auch das Layout, insbesondere die Textgestaltung wünscht man sich etwas übersichtlicher.

Für Einsteiger, die noch nie mit SAS oder einem vergleichbaren Programmsystem gearbeitet haben, ist das Buch allerdings nur bedingt empfehlenswert. Die für den Einstieg relevanten Abschnitte müssen aus dem ganzen Buch zusammengesucht werden. Dabei sind die Tips der Autoren, welche Abschnitte unbedingt zu lesen und welche gefahrlos anfangs überschlagen werden können, zwar hilfreich, aber dennoch bleibt es ein eher mühsames Unterfangen.

Ungeachtet mancher Kritikpunkte handelt es sich um ein ausgesprochen nützliches Buch, das dem Anwender die Verwendung der bekanntermaßen extrem umfangreichen Originaldokumentation zum SAS-System erleichtert und oft auch ganz erspart. Es liefert kurz und kompakt, aber trotzdem ausführlich alle Informationen, die man normalerweise bei der praktischen Arbeit mit dem SAS-System braucht.

Petra Hartmann

A help for new friends of SAS:

Kurzeinführung in SAS-PC für alte SPSS-PC-Freunde
von Petra Hartmann,
Zentralarchiv für empirische Sozialforschung
1992

Für fortgeschrittene SPSS-PC-Benutzer, die gern auch mit dem SAS-System arbeiten möchten, ist im Kontext des diesjährigen Frühjahrsseminars eine Kurzeinführung ins SAS-System erstellt worden. In dieser Kurzeinführung werden auf lediglich 14 Seiten die wichtigsten Kommandos zum Umgang mit dem System dargestellt und Äquivalente für die wichtigsten SPSS-PC-Kommandos zum Dateimanagement, zur Datenbeschreibung und Datenmodifikation benannt. Diese Kurzeinführung ist auf Anfrage beim Zentralarchiv für empirische Sozialforschung erhältlich.

Weitere Buchhinweise

Handbuch für computerunterstützte Datenanalyse
Band 1: Grundlagen computerunterstützter Datenanalyse

Reinhard Wittenberg, Sozialwissenschaftliches Institut, Universität Erlangen-Nürnberg

1991, XTV, 223 Seiten, DM 36,80 (incl. Begleitdiskette 3,5")
ISBN 3-437-50341-3 (UTB-Nr. 1603)

Das neue praxisorientierte Handbuch führt in statistische Analyseverfahren ein, die in der empirisch orientierten Sozialforschung verwendet werden. Es leitet systematisch und leicht verständlich zur selbständigen, kompetenten Arbeit mit Datenanalysesystemen an.

Der erste Band des Handbuches dient als theoretisch-methodische Grundlage sowie als Nachschlagewerk für die Folgebände, die den Umgang mit bekannten Programmsystemen zur Datenanalyse vermitteln. Der dem ersten Band auf Diskette beigefügte Datensatz erlaubt es, die im Grundlagenband und den Folgebänden vorgeführten Beispiele am Computer nachzuvollziehen. Eigene Fragestellungen können vom Benutzer anhand des realistischen Forschungsprogramms ALLBUS 88 (Allgemeine Bevölkerungsumfrage der Sozialwissenschaften) selbständig bearbeitet werden.

Inhalt: Grundlagen computerunterstützter Datenanalyse - Der Beispieldatensatz ALLBUS 88 - Datenanalyse am Beispiel ALLBUS 88 - Zusammenstellung der mathematisch-statistischen Voraussetzungen der vorgestellten Verfahren, Tests und Koeffizienten im Überblick - Anhang

Band 2: Datenanalyse mit SPSS

Reinhard Wittenberg, Sozialwissenschaftliches Institut, Universität Erlangen-Nürnberg, und *Hans Cramer,* Erlangen

1992, 200 Seiten, DM 24,80, ISBN 3-437-50342-1 (UTB Nr. 1602)

Das mehrbändige praxisorientierte "Handbuch für computerunterstützte Datenanalyse" führt in statistische Analyseverfahren ein, die in allen empirisch ausgerichteten Wissenschaften verwendet werden. In den Einzelbänden werden die am meisten verbreiteten Softwarepakete beschrieben, mit denen diese Verfahren ausgeführt werden können.
Der als Lehr- und Arbeitsbuch konzipierte Band 2 gibt eine Einführung in die Datenanalyse mit SPSS 4.0 und SPSS/PC+4.0, dem meistgenutzten Softwaresystem für angewandte Statistik. Anhand eines vorher erwähnten Datensatzes, entwickelt aus der "Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften" (ALLBUS), kann der Leser nachvollziehen, wie typische Auswertungsprobleme einer empirischen Untersuchung mit Hilfe von SPSS gelöst werden.

Der auch hierfür verwendete Datensatz liegt der Diskette zu Band 1 (UTB 1603) bei, der den theoretisch-methodischen Hintergrund für die in SPSS verfügbaren Statistikverfahren liefert und zudem als nützliches Nachschlagewerk dient.

In Vorbereitung:
Band 3: Küffner/Gogolok - Datenanalyse mit P-Stat. 1991.
Etwa DM 19,80; (Best.-Nr. 40266, UTB 1628)
Band 4: Küffner/Gogolok - Datenanalyse mit SAS. 1991.
Etwa DM 19,80; (Best.-Nr. 50343, UTB 1601)

Empirische Sozialforschung im vereinten Deutschland.
Bestandsaufnahme und Perspektiven.

Dieter ***Jaufmann****, Ernst* ***Kistler,*** *Klaus Meier,* ***Karl-Heinz Strech:***

1992, 445 Seiten, DM 78,-; Campus-Verlag Frankfurt/New York, ISBN 3-593-34512-9

Die Entwicklung der Empirischen Sozialforschung und der Umfrageforschung ist während der letzten 40 Jahre in den zwei Teilen Deutschlands unterschiedlich verlaufen. Beide Forschungsrichtungen waren in der DDR vor der Wende eigentlich keine öffentlichen Wissenschaften. Ihre Geschichte muß somit erst noch geschrieben werden. Die dazu nötige Bestandsaufnahme wird exemplarisch mit diesem Buch angegangen. Sie wird auch im Kontext der Entwicklungen in der früheren Bundesrepublik vorgenommen und kritisch kommentiert. Bei der Sicherung, Analyse und Darstellung dieses Materials geht es über die Gewinnung von Orientierungswissen hinaus um die Zurverfügungstellung von Interpretationswissen. Darauf aufbauend werden dann Perspektiven für Wissenschaft und Praxis aufgezeigt.

Im Vorwort zu diesem Buch schreibt ***Harald Uhl*** vom BMFT:
Dieses Buch schließt an ein historisches Datum an. Vor 40 Jahren hat unweit von Ladenburg, im benachbarten Weinheim a.d. Bergstraße, die erste repräsentative Tagung der deutschen empirischen Sozialforschung nach den äußeren und inneren Verwüstungen des NS-Regimes und des 2. Weltkriegs stattgefunden. Die theoretische, empirische und methodische Arbeit der deutschen Sozialforschung und ihre internationale Öffnung haben durch grundlegende Referate von *Leopold von Wiese,* **Theodor** ***Adorno*** und zahlreichen anderen Männern und Frauen der deutschen Nachkriegssoziologie damals entscheidende Impulse erhalten.

Die Tagung in der ***Gottlieb Daimler-*** und ***Karl Benz***-Stiftung am 19. und 20. Februar 1991 setzt ein neues historisches Datum. Zum ersten Mal seit der jahrzehntelangen, politisch verursachten Trennung der deutschen Sozialwissenschaften durch Mauer und Stacheldraht ist es möglich gewesen, Sozialforscher aus allen Teilen des wiedervereinigten Deutschlands zur Bestandsaufnahme und zum Meinungsaustausch über Perspektiven der Einstellungs-, Markt- und Meinungsforschung zusammenzubringen. Frau **Prof.** ***Elisabeth Noelle-Neumann*** und

Prof. ***Ludwig von Friedeburg*** stellten eine unmittelbare persönliche Brücke zur historischen Weinheimer Tagung von 1951 her. Gemeinsam mit allen anderen Mitwirkenden stellten sie nicht nur die historische Situation, sondern vor allem die historische Dimension der Herausforderung dar, der sich die deutschen Sozialwissenschaften im vereinigten Deutschland heute gegenübersehen.

Noch wird die Situation durch Fremdheit und Unkenntnis in vielen Bereichen der gesellschaftlichen und wissenschaftlichen Situation bestimmt. Allein die gegenseitige Information, der Abbau von Informationsdefiziten und auch von Distanz im gegenseitigen Umgang haben diese Tagung als wichtig erwiesen. Die Entdeckung gemeinsamer methodischer Zugänge, die Wiederaufnahme verschütteter oder nicht zugänglicher Traditionen und vor allem die gemeinsame Überzeugung, daß die deutsche Sozialforschung in gemeinsamer Anstrengung wesentliche Beiträge zur Analyse und Lösung der einzigartigen gesellschaftlichen Problembreite im wiedervereinigten Deutschland leisten kann, haben die Tagung über den Informationsaustausch und das Knüpfen neuer wissenschaftlicher Kontakt zu tragfähigen Ansätzen für die gemeinsame Bewältigung dieser wissenschaftlichen Herausforderung geführt.

Auch für den guten Kenner der sozialwissenschaftlichen Landschaft in der Bundesrepublik in den letzten Jahrzehnten sind daher die Referate dieses Bandes mehr als eine aktuelle Situationsbeschreibung. Sie bieten Anregungen und Hinweise auf eine Fülle neuer Themen, die auf die Sozialforschung zugekommen sind, und sie bieten ermutigende Einblicke in ein hohes Maß an Kooperationsbereitschaft zwischen der Markt- und Meinungsforschung, universitären und außeruniversitären Forschungseinrichtungen sowie die in der Bundesrepublik inzwischen fest etablierten, die neuen Länder in naher Zukunft einbeziehenden Infrastruktureinrichtungen der deutschen Sozialforschung. Wichtig sind zudem die zahlreichen Hinweise, nicht in eine neue Zahlen- oder Befragungseuphorie hineinzugeraten, sondern im Interesse der soliden und seriösen Sozialforschung die qualitativen Ansprüche gerade angesichts der neuen und erfreulichen Möglichkeiten in keinem Falle zu reduzieren. Der rasche Neuaufbau der Sozialforschung in den neuen Bundesländern, der durch die Empfehlungen des Wissenschaftsrates für die Wirtschafts- und Sozialwissenschaften, durch den Bundesminister für Forschung und Technologie, durch die Länderministerien und die Markt- und Meinungsforschungsinstitute zügig vorangetrieben wird, stand verständlicherweise im Mittelpunkt aller struktureller Überlegungen.

Es wäre sicherlich verfrüht, bereits jetzt eine Bewertung der Wirkung dieser Tagung vornehmen zu wollen. Der historische Zeitpunkt, die aufgezeigten Perspektiven und die demonstrative Kooperationsbereitschaft aller Zweige der deutschen Sozialforschung lassen aber erwarten, daß diese Tagung und der sie dokumentierende Band eine dauerhafte Wirkung auf die Problemdefinitionen, das methodische Instrumentarium und die strukturelle Zusammenarbeit der Sozialforschung im vereinten Deutschland ausüben werden.

Einführung in die Verlaufsdatenanalyse,
Statistische Grundlagen und Anwendungsbeispiele zur
Längsschnittanalyse kategorialer Daten

Hans-Jürgen Andreß

Köln (Zentrum für Historische Sozialforschung) 1992, 328 S.
(Historical Social Research / Historische Sozialforschung, Supplementreihe, Heft 5).

Dieses Buch befaßt sich mit den Möglichkeiten und Problemen der Analyse sozialer Prozesse, wenn Art und Zeitpunkt von Veränderungen für eine Stichprobe von Untersuchungseinheiten bekannt sind. Beispiele solcher Verlaufsprozesse sind die geographische und berufliche Mobilität von Personen, Mortalitätsdaten, Bildungskarrieren, Veränderungen der Familienzusammensetzung, aber auch Änderungen der politischen Verfassung von Regierungssystemen oder Entwicklungsprozesse formaler Organisationen. Verlaufsdaten sind ganz allgemein Informationen über die zeitliche Abfolge von Ereignissen.

Im Abschnitt 1.1 wird anhand von Beispielen erläutert, worum es sich dabei konkret handelt, welche neuen Erkenntnisse man mit Verlaufsdaten gewinnen kann und welche Fragestellungen im Rahmen einer Verlaufsanalyse untersucht werden können. Im Abschnitt 1.2 werden Verlaufsdaten dann genauer definiert und von anderen Datenstrukturen (z.B. von Zeitreihen) abgegrenzt. Abschnitt 3 beschäftigt sich schließlich mit Erhebungsproblemen.

Nach dieser Einführung konzentrieren sich dann alle folgenden Kapitel auf die eigentliche Auswertung. Abschnitt 2 erläutert einige statistische Grundbegriffe, die für das Verständnis der Methode notwendig sind. Im Abschnitt 3 werden zwei Datensätze vorgestellt, an denen beispielhaft die folgende Auswertung demonstriert werden soll. Dabei werden auch die praktischen Probleme des Datenmanagements und der Auswertung mit Hilfe von EDV-Programmen angesprochen. Die Auswertung selbst wird in mehr explorative und mehr konfirmatorische Verfahren unterteilt. Im Abschnitt 4 geht es darum, Einzelfragen auf einer mehr deskriptiven Ebene zu untersuchen, während im Abschnitt 5 umfassende Erklärungsmodelle für die Daten getestet werden. Das Buch schließt mit dem Abschnitt 6 über weiterführende Fragestellungen der Verlaufsdatenanalyse.

Alle wesentlichen Methoden werden in einer Form behandelt, daß sie erstens anhand eines einfachen Beispiels nachgerechnet werden können und daß zweitens ihre spezifische Anwendung in den Sozialwissenschaften deutlich wird. Für die praktische Auswertung stehen dann EDV-Programme zur Verfügung, jedoch sollte man nach der Lektüre dieses Buches in der Lage sein, einfache Probleme auch mit dem Taschenrechner zu lösen. Wer darüber hinaus an

Spezialfragen oder den allgemeinen statistischen Eigenschaften der verwendeten Methode interessiert ist, erhält auf die Weise einen Einstieg in die statistische Fachliteratur.

Der Textrichtet sich an Personen, die bereits Grundkenntnisse statistischer (insbesonders multivariater) Methoden haben. Eine gewisse Vertrautheit mit der einfachen linearen Regression ist Voraussetzung und erleichtert das Verständnis der hier zu besprechenden Methoden, da sie häufig in Analogie zu diesem klassischen Modell sozialwissenschaftlicher Datenanalyse entwickelt werden. Vorausgesetzt werden auch Grundkenntnisse der Matrizenrechnung, die jedoch in der Regel nicht über die Multiplikation von Matrizen hinausgeht.

Methoden der Verlaufsdatenanalyse beruhen im wesentlichen auf Anwendungen der Wahrscheinlichkeitsrechnung (stochastische Prozesse). Begriffe wie Zufallsvariable, Erwartungswert, Dichte und Verteilungsfunktion sollten daher bekannt sein. Für den interessierten Leser findet sich dazu im Anhang A noch einmal eine Wiederholung der zentralen Begriffe. Außerdem werden die mathematischen Hilfsmittel erläutert, die in diesem Zusammenhang wichtig sind (Differential- und Integralrechnung). Ein Glossar im Anhang C erfaßt alle im Text verwendeten mathematischen Symbole.

Der Text entstand als Skript im Rahmen des Aufbaukurses des ZHSF-Herbstseminars 1991 zum Thema "Event History Analysis" und stellt eine grundlegende Überarbeitung des Bandes 1 der ZUMA-Methodentexte (1985) dar. Das Buch kann zum Preis von 18.-DM direkt bezogen werden beim: Zentralarchiv für empirische Sozialforschung, Abteilung Zentrum für Historische Sozialforschung, Bachemer Str. 40, D-5000 Köln 41 (Fax: 0221 - 4769444).

ZHSF-Herbstseminare 1992

"Methodik der Historischen Sozialforschung"
- Aufbaukurs I -

"Klassifikationsverfahren"

6. bis 11. September 1992

Kursangebot:
6-tägiger Aufbaukurs (6 Tage Kernkurs) für Fortgeschrittene zur Einführung in die Verfahren der computergestützten Klassifikation.
Inhalte:
Aufgabe von Klassifikationsverfahren ist es allgemein, eine Menge von Objekten (z.B. Personen) aufgrund bestimmter Merkmale in (homogene) Klassen zusammenzufassen. Zunächst wird deutlich gemacht, daß die Anwendung eines bestimmten Klassifikationsverfahrens (z.B. "quick-cluster" in SPSS) i.d.R. nur einen Schritt in der Lösung einer Klassifikationsaufgabe darstellt. Folgende Verfahren werden u.a. behandelt: deskriptive Klassifikationsverfahren (Aufsuchen von Extremtypen, Bestimmen aller empirischen Konfigurationen); hierarchische Clusteranalyse; allokative Klassifikationsverfahren; Klassifikationsverfahren, die Meßfehler berücksichtigen (wie z.B. Analyse latenter Klassen, Korrespondenzanalyse). Die Darstellung jedes Verfahrens wird in die Lösung einer Klassifikationsaufgabe eingebunden. Dabei werden u.a. auch Verfahren der Stabilitiätsprüfung (wie stabil ist die gefundene Klassifikationslösung?) behandelt.
Vermittlung:
Die Lerninhalte werden zunächst vom Dozenten in täglich vierstündigen Vorlesungen (vormittags) vermittelt und in nachfolgenden Übungen (nachmittags) computergestützt angewandt. Während der gesamten Kurslaufzeit stehen der Dozent und andere ZHSF-Mitarbeiter für die Beratung der Teilnehmer bei der Durchführung eigener Forschungsarbeiten zur Verfügung. Darüber hinaus besteht die Möglichkeit, in Teilnehmervorträgen solche Forschungsarbeiten vor dem Plenum zu diskutieren.
Datenbasis:
Für den Aufbaukurs ist vor allem ein Übungs-Datensatz mit familialen Haushaltsdaten einer Osttiroler Gemeinde am Ende des 18. Jahrhunderts vorgesehen. Die Teilnehmer können ggf. auch Datensätze aus eigenen Forschungsprojekten analysieren.
Vorkenntnisse:
Es ist die vorhergehende Teilnahme an den Grundkursen I **und** II früherer ZHSF-Herbstseminare erforderlich bzw. werden zumindest vergleichbare Kenntnisse in der Methodik der Histo-

rischen Sozialforschung, der deskriptiven und schließenden Statistik sowie der EDV-Anwendung vorausgesetzt.

Dozent:	**Dr. *Johann Bacher*** (Universität Linz).
Assistent:	***Karl Pierau*** (ZA-ZHSF, Universität Köln)
Ort/Termin:	6.9.-11.9.1992 in den Räumen des ZA und des Philosophikums der Universität Köln.
Kursgebühren:	DM 50,- (Kosten für die Kurs unterlagen).
Kursleitung:	Priv.-Doz. Dr. ***Wilhelm H. Schröder.***
Sekretariat:	***Beate Reinemuth.***
EDV-Organisation:	***Ralph Ponemereo.***
Anmeldung:	Bitte fordern Sie die Anmeldeunterlagen beim Zentralarchiv für Empirische Sozialforschung, Abteilung Zentrum für Historische Sozialforschung (Bachemer Str. 40, D-5000 Köln 41) an.
Anmeldeschluß:	15. Juni 1992.

"Methodik der Historischen Sozialforschung"
- Aufbaukurs II -

"Pfad-Analyse und Strukturgleichungsmodelle"

12. bis 19. September 1992

Kursangebot:
8-tägiger Aufbaukurs (6 Tage Kernkurs, 2 Tage QUANTUM-Workshop) für Fortgeschrittene zur Einführung in die Pfad-Analyse und Strukturgleichungsmodelle.
Inhalte:
Historiker arbeiten häufig mit Kausalhypothesen, die als solche oft nicht explizit formuliert und empirisch überprüft werden. Das in der Forschungspraxis noch am ehesten benutzte Instrument ist die multiple Regressionsanalyse. Dabei werden nur die direkten Einflüsse spezifiziert, die verschiedene "unabhängige" Variablen auf eine "abhängige" Variable ausüben; eventuell vorliegende kausale Beziehungen zwischen den unabhängigen Variablen werden nicht untersucht. Wegen ihres geringen Komplexitätsgrades stellen derartige Regressionsmodelle (Eingleichungsmodelle) im vielen Fällen keine adäquate Übersetzung der jeweils interessierenden theoretischen Aussagen dar. Eine angemessenere Formalisierung mag dann mit Hilfe mehrerer, aufeinander bezogener Gleichungen gelingen, die insgesamt eine Beziehungsstruk-

tur ausdrücken, die in sog. Pfad-Diagrammen graphisch dargestellt werden können. Falls die Variablen nicht direkt beobachtbar sind ("theoretische Konstrukte"), sondern durch mehrere empirische Indikatoren indirekt gemessen werden, stellt sich als weitere Aufgabe die Integration von Theorie- und Meßmodell (beispielsweise in Form von sog. LISREL-Modellen). Bis zu welcher Komplexitätsstufe die Strukturgleichungsmodelle in diesem Aufbaukurs behandelt werden können, hängt von den Vorkenntnissen und dem Interesse der Teilnehmer ab.- QUANTUM-Workshop: wahlweise Teilnahme an inhaltlich-methodisch orientierten Sektionen oder an softwareorientierten Einführungen/Demonstrationen ("Entwurfstechniken von relationalen Datenbanken", "TUSTEP", "Datenbanken auf dem PC"; "Nichtlineare Regressionsanalyse", "Computerunterstütztes wissenschaftliches Edieren mit TUSTEP"; "KLEIO für Statistikanwender").

Vermittlung:

Die Lerninhalte werden zunächst vom Dozenten in täglich vierstündigen Vorlesungen (vormittags) vermittelt und in nachfolgenden Übungen (nachmittags) computergestützt angewandt. Während der gesamten Kurslaufzeit stehen der Dozent und andere ZHSF-Mitarbeiter für die Beratung der Teilnehmer bei der Durchführung eigener Forschungsarbeiten zur Verfügung. Darüberhinaus besteht die Möglichkeit, in Teilnehmervorträgen solche Forschungsarbeiten vor dem Plenum zu diskutieren.

Datenbasis:

Für den Aufbaukurs ist ein Übungs-Datensatz mit Daten der Historischen Sozialforschung im 19. und 20. Jahrhundert vorgesehen.

Vorkenntnisse:

Es ist die vorhergehende Teilnahme an den Grundkursen I **und II** früherer ZHSF-Herbstseminare erforderlich bzw. werden zumindest vergleichbare Kenntnisse in der Methodik der Historischen Sozialforschung, der deskriptiven und schließenden Statistik sowie der EDV-Anwendung vorausgesetzt.

Dozent:	Prof. Dr. *Helmut Giegler* (Universität Erlangen-Nürnberg).
Assistent:	*Karl Pierau* (ZA-ZHSF, Universität Köln)
Ort/Termin:	12.9.-19.9.1992 in den Räumen des ZA und des Philosophikums der Universität Köln.
Kursgebühren:	DM 50,- (Kosten für die Kursunterlagen).
Kursleitung:	Priv.-Doz. Dr. *Wilhelm H. Schröder.*
Sekretariat:	*Beate Reinemuth.*
EDV-Organisation:	*Ralph Ponemereo.*
Anmeldung:	Bitte fordern Sie die Anmeldeunterlagen beim Zentralarchiv für Empirische Sozialforschung, Abteilung Zentrum für Historische Sozialforschung an.
Anmeldeschluß:	15. Juni 1992.

Informationen zum 26. Deutschen Soziologentag "Lebensverhältnisse und soziale Konflikte im neuen Europa" 28. September bis 2. Oktober 1992 in der Heinrich-Heine-Universität Düsseldorf

1. Soziologentag als Bildungsurlaub

Ein Besuch des 26. Deutschen Soziologentages, der vom 28. September bis 2. Oktober 1992 zum Thema "Lebensverhältnisse und soziale Konflikte im neuen Europa" in der Heinrich-Heine-Universität Düsseldorf stattfindet, ist in Nordrhein-Westfalen als Bildungsurlaub nach dem Arbeitnehmer-Weiterbildungsgesetz anerkannt. Der Antrag auf Anerkennung des Kongresses als Bildungsurlaub wurde auch in anderen Bundesländern mit gesetzlicher Freistellungsregelung gestellt. Die Bescheide stehen jedoch noch aus.

2. Poster-Sessions auf dem Soziologentag 1992

Der Vorbereitungsausschuß zur Planung des 26. Deutschen Soziologentages hat beschlossen, eine für den Soziologentag neue Form der Präsentation von Forschungsergebnissen zu erproben. Auf sog. Poster-Sessions können 4-5 Aussteller Projektergebnisse für 2 Stunden graphisch präsentieren; die Vertreter des Projekts stehen während dieser Zeit für Diskussionen mit den Besuchern zur Verfügung.

Der Vorbereitungsausschuß lädt ein, sich für eine entsprechende Präsentation zu bewerben.

Folgende Vorgaben sind zu beachten:

- es sollte sich (i.d.R.) um ein empirisches Projekt handeln,
- die Forschungsergebnisse sollten noch nicht (bzw. noch nicht vollständig) der wissenschaftlichen Öffentlichkeit bekannt sein;
- es sollen vor allem jüngere Wissenschaftlerinnen und Wissenschaftler die Chance der Präsentation erhalten;
- ein inhaltlicher Bezug zum Tagungsthema muß nicht gegeben sein.

Die Aussteller müssen ihre Präsentation vor Beginn der Sitzung selber anbringen und anschließend wieder mitnehmen: für jede Präsentation stehen 2 qm Stellwand zur Verfügung. Anmeldungen werden mit folgenden Angaben erbeten: Autor(en), Institution; Titel des Projekts.

Das Auswahlgremium wird sich bemühen, inhaltlich verwandte Projekte unter einem Obertitel zu einer Poster-Session zusammenzubringen.

Weitere Auskünfte durch:

Heinrich-Heine-Universität Düsseldorf
Sozialwissenschaftliches Institut
Kongreßbüro des 26. Deutschen Soziologentages
Universitätsstr. 1
4000 Düsseldorf 1
Tel.: 02 11/311-47 11

3. Das Zentralarchiv auf den Soziologentag

Auf dem diesjährigen Soziologentag wird das Zentralarchiv wieder am GESIS-Stand vertreten sein. Zum Thema des Soziologentags werden wir einige Datensätze aus unseren Archivbeständen besonders herausstellen und präsentieren. Interessenten und Benutzer des Zentralarchivs laden wir schon jetzt zu einem Besuch an unseren Informationsstand ein.

Steffen Kühnel
Zentralarchiv für
empirische Sozialforschung
Universität zu Köln
Bachemer Str. 40

5000 Köln 41

Tel. 0221/47694-32
Fax 0221/47694-44

Ich interessiere mich für das Frühjahrsseminar 1993 zum Thema "lineare Strukturgleichungsmodelle".

Name _____

(Institut) _____

Anschrift _____

_____

Telefon _____

Telefax _____

Senden Sie mir bitte Informationen über das Seminar und das Anmeldeformular zu, sobald die Planung des Seminars abgeschlossen ist.