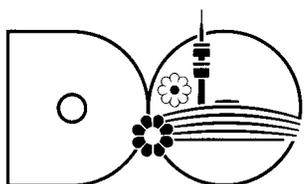


Dortmunder Statistik



**Amt für Statistik
und Wahlen
Stadt Dortmund**

Die Volkszählung 1987 als Basis zur
Interpretation der Kommunalwahl in NW 1989
- Am Beispiel Dortmund -

Volker Nürnberg
Fachbereich Statistik
Universität Dortmund

November 1990
Sonderheft 128

**Die Volkszählung 1987 als Basis zur
Interpretation der Kommunalwahl in
NW 1989
- Am Beispiel Dortmund -**

Volker Nürnberg

Fachbereich Statistik
Universität Dortmund

In Zusammenarbeit mit:

Amt für Statistik und Wahlen der Stadt Dortmund

Bereich Anwendungssysteme
Hochschulrechenzentrum Dortmund

November 1990

Vorwort

Die vorliegende Studie ist ein weiteres Ergebnis der Zusammenarbeit des Amtes für Statistik und Wahlen der Stadt Dortmund mit der Universität Dortmund, hier insbesondere dem Fachbereich Statistik und dem Bereich Anwendungssysteme des Hochschulrechenzentrums.

Diese Zusammenarbeit ging aus einer Initiative des Leiters des Dortmunder Amtes für Statistik und Wahlen, Werner Noeske, hervor, die Volkszählung 1987 für die Analyse der Ratswahl in Dortmund 1989 mit statistischen Methoden auszuwerten.

Zu diesem Zweck nahm er vor der Kommunalwahl 1989 Kontakt zum Fachbereich Statistik der Universität auf. Als Lehrstuhlinhaber für Wirtschafts- und Sozialstatistik vergab ich das Thema als Diplomarbeit an den Autor dieser Studie. Diese erscheint noch vor der Diplomarbeit, um zur kommenden gesamtdeutschen Wahl am 2. Dezember der interessierten Öffentlichkeit zugänglich zu sein.

Diese Studie verwendet nur eine einzige statistische Methode, nämlich die einer multiplen Regression. Den Schwerpunkt hat der Autor, Volker Nürnberg, weniger auf die Methoden als auf die Ergebnisse und deren Interpretation gelegt. In seiner voraussichtlich im kommenden Frühjahr fertigen Diplomarbeit wird er sich intensiver den theoretischen Aspekten der angewandten Methodik zuwenden und auch andere Modelle betrachten.

Neben den bereits erwähnten Personen ist der Autor zu Dank verpflichtet: Rainer Döhla, H.-G. Felinks, Harald Gneiß und Werner Kötter im Dortmunder Amt für Statistik und Wahlen, die bei der Aufbereitung der Volkszählungsdaten halfen, Francisco Fernández-Marín, Sibylle Rister-Mende, Elisabeth Schach und Dr. Wolfgang Thiboud im Bereich Anwendungssysteme des Hochschulrechenzentrums Dortmund, die die Koordinaten der Statistischen Unterbezirke Dortmund in geeigneter Form zur Verfügung stellten und sich beim graphischen Teil dieser Arbeit sehr hilfreich zeigten, sowie Heike Feldmann, Wilhelm Heckmann, Bernd Hupfeld, Franz-Horst Möllers und Dr. Horst-Jürgen Wienen, seinen Kollegen im Amt für Statistik, Stadtforschung und Wahlen der Stadt Bochum, wo er als studentische Hilfskraft beschäftigt ist, für zahlreiche Anregungen im Entstehungsprozeß dieser Schrift.

Prof. Dr. Walter Krämer, Fachbereich Statistik, Universität Dortmund

Geleitwort

Mit der hiesigen Universität gemeinsam Projekte abzuwickeln hat in Dortmund inzwischen Tradition - Technologiezentrum und Technologiepark sind sichtbar gewordener Ausdruck des erfolgreichen Bemühens, Probleme der Praxis mit dem Wissen und den Möglichkeiten der Universität anzugehen. Warum also nicht auch auf dem wichtigen Felde der Wahlstatistik einmal den Versuch unternehmen, kooperativ bisher nicht vorgelegte Informationen zu erarbeiten und der Öffentlichkeit zu präsentieren, zumal sich die Chance, aktuelle Volkszählungsdaten in ein solches Projekt mit einzubeziehen, bestenfalls erst in 10 Jahren wieder bietet.

Dies war die Ausgangssituation, als sich rd. 2 Jahre nach der Durchführung der Volkszählung 1987 und somit unmittelbar vor der Kommunalwahl 1989 Mitarbeiter des Amtes für Statistik und Wahlen und des Fachbereichs Statistik der Universität Dortmund zusammenfanden und auf das Vorhaben verständigten, dessen Ergebnisse nunmehr hiermit vorgelegt werden. Dabei beschränkte sich der Beitrag meiner Mitarbeiter auf thematische Festlegungen und die Zurverfügungstellung von Daten, während die Hauptarbeit der methodischen Aufbereitung und textlichen Darstellung bei der Universität Dortmund lag. Ich danke deshalb vor allem Herrn Volker Nürnberg, daß das gemeinsame Wahlanalyseprojekt erfolgreich abgeschlossen werden konnte und nunmehr im Rahmen der "Dortmunder Statistik" erscheinen kann.

Zu hoffen bleibt, daß die hier vorgelegten Ergebnisse über die kommende Bundestagswahl hinaus für die Praxis von Bedeutung sein werden.

Koch, Stadtrat

Inhaltsverzeichnis

1 Ziel und Gegenstand der Studie	1
2 Die wesentlichen Ergebnisse	1
3 Einführung in die Problemstellung	5
4 Detaillierte Resultate	8
4.1 Für die CDU	8
4.2 Für die SPD	11
5 Methodische Strategien und Modelle	14
5.1 Das Regressionsmodell	14
5.1.1 Grundsätzliches über Korrelations- und Regressionsrechnung	14
5.1.2 Das einfache lineare Regressionsmodell	15
5.1.3 Der Fall mehrerer Einflußvariablen	18
5.2 Das Problem des "ökologischen Fehlschlusses" bei Aggregatdaten	18
5.3 Ein multiples Regressionsmodell zur Erklärung der Stimmenanteile der einzelnen Parteien	22
5.3.1 Vorüberlegungen	22
5.3.2 Unterschiedliche Modellansätze	25
5.3.3 Der Aussagewert der Residuen	29
5.4 Modellbildung mittels Residuen	31
5.4.1 Sukzessive Definition der Einflußvariablen	31
5.4.2 Vorteile der Modellbildung durch Residualanalyse und Unterschiede zu anderen Modellbildungstechniken	57
6 Schlußkommentar	60
A Detaillierte Einzelergebnisdarstellung	64

Tabellenverzeichnis

1	Beispiel ökologischer Beziehungen zwischen SPD-Wählern und Arbeiterbevölkerung	19
2	Parteipräferenzen der Selbständigen als theoretische Anteile .	23
3	Korrelationen verschiedener Prädikatoren zu den Residuen der einfachen Regression	32
4	Korrelationen wichtiger Variablen der Berufs- und Familienstruktur untereinander	32
5	Stimmbezirke, in denen der CDU-Stimmenanteil bei der einfachen Regression um über 9 Prozentpunkte überschätzt wurde	35
6	Stimmbezirke, in denen der CDU-Stimmenanteil bei der einfachen Regression um über 9 Prozentpunkte unterschätzt wurde	36
7	Stimmbezirke, die Altersheime enthalten (1. Teil)	40
8	Stimmbezirke, die Altersheime enthalten (2. Teil)	41
9	Stimmbezirke ohne verfügbare Vorwahlergebnisse und Residuen des für diese Bezirke möglichen Modells für die CDU	48
10	Liste der Residuen der endgültigen Modelle für SPD- und CDU-Stimmenanteile auf Stimmbezirksebene	64

Abbildungsverzeichnis

1	Graphische Darstellung der CDU-Ergebnisse bei der KW 1989	10
2	Graphische Darstellung der SPD-Ergebnisse bei der KW 1989	13
3	Beispiele für Stichproben mit verschiedenen Werten des Korrelationskoeffizienten r	15
4	Darstellung der extremen Residuen bei der einfachen Regression	38
5	Hochburgen und Diaspora der SPD bei der KW 1984	50
6	Hochburgen und Diaspora der CDU bei der KW 1984	51
7	Hochburgen und Diaspora der Nichtwähler bei der KW 1984 .	52
8	Häufigkeitsstruktur von Hochburgen bzw. Diaspora der Parteien bei der KW 1984	55

1 Ziel und Gegenstand der Studie

Die vorliegende Studie bemüht sich um eine Objektivierung des Begriffs der "Wahlerwartung" und - damit zusammenhängend - des Wahlerfolges. Eine Wahlerwartung ergibt sich aus festen Annahmen über sozio-ökonomisch bedingtes Wahlverhalten.

Das "erwartete Wahlergebnis" wird hier funktional zurückgeführt auf mehrere Einflußvariablen (Regressoren), die aus der Volkszählung bzw. aus den Ergebnissen der Vorwahl verfügbar sind. Aus der Volkszählung sind dies die Variablen: Selbständigenanteil, Arbeiteranteil und Anteil der ledigen Männer jeweils an der wahlberechtigten Bevölkerung.

Die Ergebnisse dieser "multiplen Regression" werden mit den tatsächlichen Wahlergebnissen verglichen und damit der Wahlerfolg bewertet.

Die verwendeten Daten stammen aus der VZ 1987 und den Kommunalwahlen in Dortmund 1984 und 1989.

Noch ein lesetechnischer Hinweis: Da den geneigten Leser die Ergebnisse der Studie natürlich mehr als die Methoden interessieren, stehen die Ergebnisse im Text an Anfang. Gleichwohl sind sie zeitlich zuletzt, nach umfangreichen methodischen Vorüberlegungen, die in Kapitel 5 zusammengefaßt sind, erzielt worden. Sollten im ersten Teil daher einige Details unklar bleiben sei hiermit auf die ausführliche Diskussion im methodischen Teil verwiesen, was hoffentlich der allgemeinen Verständlichkeit der ersten 4 Kapitel keinen Abbruch tut.

2 Die wesentlichen Ergebnisse

Die Anteile der Wähler der beiden "großen" Parteien, der SPD und der CDU bei der Kommunalwahl 1989 werden in dieser Analyse auf 3 Variablen der Berufs-, Bildungs- bzw. Sozialstruktur auf Stimmbezirksebene und auf Variablen des politischen Kontextes, soweit diese aus der Vorwahl bekannt sind, zurückgeführt.

Das schließlich benutzte mathematisch-statistische Modell einer multiplen linearen Regression erklärt über 78 % der Gesamtvariabilität der CDU-Stim-

men und 72 % der Gesamtvariabilität der SPD-Stimmen¹. Die übrigen 22 % bzw. 28 % der Gesamtvariabilität lassen sich nach einer umfassenden Analyse aller für das gesamte Stadtgebiet verfügbaren Variablen nicht auf globale Einflußgrößen zurückführen. Mithin ist diese durch das Modell nicht erklärbare Variabilität ein Ausdruck des Wahlerfolges oder -mißerfolges der Parteien auf Stimmbezirksebene.

Wird der Stimmenanteil einer Partei in einem Stimmbezirk durch das Modell überschätzt, d.h. der berechnete Erwartungswert liegt über dem tatsächlichen Resultat, so hat die Partei vor diesem Hintergrund schlecht abgeschnitten und zwar um so schlechter, je höher die Differenz zwischen dem vorhergesagten und tatsächlich eingetretenen Stimmenanteil² ist.

Wenn das Modell umgekehrt den Stimmenanteil einer Partei in einem Stimmbezirk unterschätzt, so deutet dies an, daß diese Partei dort "besser als erwartet" abgeschnitten hat und zwar wiederum um so besser, je mehr das Modell den tatsächlichen Stimmenanteil unterschätzt hat.

Die Abbildungen 1 und 2 geben diese "objektivierte Erfolgsbilanz" der Parteien in anschaulicher Form wieder. Dabei ist zweierlei zu beachten. Zum einen ließen sich für 32 der 620 Dortmunder Stimmbezirke keine Vorhersagen machen, da keine Aussagen über die politische Ausgangslage gemacht werden konnten. Der Grund hierfür sind die seit der letzten Kommunalwahl im Jahr 1984 veränderten Stimmbezirksgrenzen.

Für 2 Stimmbezirke lagen keine Volkszählungsdaten vor. Stimmbezirk 590 ist ein Altersheim, das weniger als eine Blockseite umfaßt, Stimmbezirk 4290 wurde, was die VZ-Daten betrifft zu 4207 hinzugerechnet. Sechs Stimmbezirke hatten derart vom allgemeinen Trend unerklärbar abweichende Ergebnisse, daß deren Werte in der Analyse nicht berücksichtigt wurden. Tabelle 9 gibt Zwischenresultate für diese Stimmbezirke.

Zum anderen stellen die Karten die Ergebnisse auf der Basis der 170 Statistischen Unterbezirke der Stadt Dortmund dar.

Die verbliebenen 580 Stimmbezirke wurden nach ihrem Bevölkerungsschwerpunkt auf die Statistischen Unterbezirke verteilt. Die angezeigte Erfolgsbilanz stellt also das Mittel aller dem entsprechenden Unterbezirk bevölkerungsmäßig zugeordneten Stimmbezirke dar.

¹d.h. die unterschiedlichen Anteile der Stimmen für die CDU bzw. die SPD lassen sich in ihrer Höhe zu 78 % bzw. 72 % auf 3 Variablen, die aus der kleinräumigen Analyse der Volkszählung 1987 zur Verfügung stehen und auf die Hochburgen- bzw. Diasporaverhältnisse der Parteien, wie sie aus der Vorwahl bekannt sind, zurückführen.

Für die Stimmenanteile der GRÜNEN, der FDP und der Republikaner hat das verwendete Modell einen weit geringeren Anteil an erklärter Varianz (unter 55 %).

²d.h. den "Residuen" der Regression.

Detaillierten Aufschluß über die Abweichungen der tatsächlichen Ergebnisse von denen durch das Modell geschätzten gibt Tabelle 10 im Anhang, allerdings auch nur für die 580 Stimmbezirke, bei denen alle im Modell enthaltenen Variablen verfügbar waren.

Im wesentlichen aus der Notwendigkeit und dem Wunsch, die Nichtwähler gleichberechtigt in die Analyse einfließen zu lassen, wurde der Stimmenanteil der Parteien nicht als der Anteil der gültigen Stimmen für eine bestimmte Partei an der Gesamtzahl der gültigen Stimmen in einem Stimmbezirk berechnet, wie dies für die Verteilung der politischen Mandate notwendig ist, sondern es wurde der Anteil der gültigen Stimmen für eine Partei innerhalb jedes Stimmbezirkes³ an denen im *Wählerverzeichnis ohne Sperrvermerk "W" aufgeführten Wahlberechtigten*, also an allen Wahlberechtigten abzüglich der Zahl derjenigen, die hier einen Wahlschein beantragt haben⁴, berechnet. In gleicher Weise kann dann nämlich auch der Anteil der Nichtwähler bestimmt werden.

Die CDU hatte ihre "relativ Besten" Ergebnisse in:⁵

Unterbezirk	Lage	Abweichung vom Erwartungswert
512	Benninghofen-Loh	10,5 Prozentpunkte
084	Westfalendamm (Gartenstadt)	8,2 Prozentpunkte
672	Löttringhausen-Nord	7,2 Prozentpunkte
342	Wickede-Dorf	6,5 Prozentpunkte
671	Großholthausen	3,8 Prozentpunkte
212	Derne-Ost	3,7 Prozentpunkte
620	Bittermark	3,5 Prozentpunkte
323	Brackel-Reichshof	3,5 Prozentpunkte
663	Schnee	3,4 Prozentpunkte
002	City-West	3,2 Prozentpunkte
244	Husen-Süd	3,0 Prozentpunkte
251	Lanstrop-West	3,0 Prozentpunkte
824	Huckarde-Mitte	2,9 Prozentpunkte
341	Wickede-Nord	2,9 Prozentpunkte
641	Eichlinghofen	2,9 Prozentpunkte

³Eine Berücksichtigung der Briefwahl ist über die Variable WAHLSCHE möglich, die aber die Ergebnisse kaum beeinflusst, wie die Analyse ergab, sondern nur die Bewertung der CDU etwas nach oben hin korrigiert.

⁴Der Anteil derjenigen Wahlscheininhaber, die in diesen Wahllokalen direkt wählen und dadurch die Analyse verfälschen, ist so gering, daß er hier vernachlässigt werden kann.

⁵Angaben über Statistische Unterbezirke. Die Bewertungen sind hier Mittelwerte aller Stimmbezirke, die nach dem Bevölkerungsschwerpunkt den jeweiligen Unterbezirken zuzuordnen sind.

2 DIE WESENTLICHEN ERGEBNISSE

4

und ihre "relativ schlechtesten" Ergebnisse in:

Unterbezirk	Lage	Abweichung vom Erwartungswert
922	Mengeder Heide	-4,7 Prozentpunkte
416	Aplerbecker-Mark II	-4,7 Prozentpunkte
324	Brackel-Dorf	-4,0 Prozentpunkte
415	Aplerbeck-Mark I	-3,9 Prozentpunkte
023	Dorstfelder Brücke (Hoesch)	-3,4 Prozentpunkte
924	Mengede-Alte-Kolonie	-3,3 Prozentpunkte
314	Asseln-Kolonie-Holstein	-3,0 Prozentpunkte
542	Holzen	-3,0 Prozentpunkte
511	Benninghofen	-3,0 Prozentpunkte
422	Berghofen II	-2,9 Prozentpunkte
261	Scharnhorst-Nord	-2,6 Prozentpunkte
841	Alt-Kirchlinde	-2,6 Prozentpunkte
211	Derne-West	-2,5 Prozentpunkte
570	Wichlinghofen	-2,4 Prozentpunkte
130	Holthausen	-2,3 Prozentpunkte

Die SPD hatte ihre "relativ besten" Ergebnisse in:

Unterbezirk	Lage	Abweichung vom Erwartungswert
612	Barop-Nordwest	7,7 Prozentpunkte
243	Husen-Nord	7,4 Prozentpunkte
823	Wischlingen	6,7 Prozentpunkte
112	Brechten (Süd)	6,5 Prozentpunkte
342	Wickede-Dorf	6,2 Prozentpunkte
441	Sölde I	5,9 Prozentpunkte
674	Kirchhörde-Ost	5,5 Prozentpunkte
222	Hostedde	5,3 Prozentpunkte
241	Kurl-Nord	4,9 Prozentpunkte
511	Benninghofen (Nord)	4,9 Prozentpunkte
642	Universität	4,8 Prozentpunkte
741	Marten-Germania	4,5 Prozentpunkte
620	Bittermark	4,4 Prozentpunkte
760	Westrich	3,3 Prozentpunkte
732	Lütgendortmund-Deipenbeck	4,3 Prozentpunkte

und ihre "relativ schlechtesten" Ergebnisse in:

Unterbezirk	Lage	Abweichung vom Erwartungswert
923	Mengede	-8,9 Prozentpunkte
653	Kleinholthausen	-6,6 Prozentpunkte
661	Persebeck	-6,2 Prozentpunkte
562	Wellinghofen-Süd	-6,0 Prozentpunkte
002	City-West	-5,9 Prozentpunkte
842	Kirchlinde-West	-5,3 Prozentpunkte
041	Hafen (Fredenbenbaum)	-5,2 Prozentpunkte
681	Rombergpark	-5,0 Prozentpunkte
052	Nordmarkt (östl. Bornstraße)	-4,6 Prozentpunkte
332	Wambel-Süd	-4,5 Prozentpunkte
924	Mengede-Alte-Kolonie	-4,2 Prozentpunkte
263	Scharnhorst-Süd	-3,8 Prozentpunkte
570	Wichlinghofen	-3,8 Prozentpunkte
532	Hörde-Mitte	-3,6 Prozentpunkte
841	Alt-Kirchlinde	-3,2 Prozentpunkte

3 Einführung in die Problemstellung

Wie kann man eine subjektive, geheime Entscheidung, wie die Stimmabgabe bei einer Wahl, objektiv beurteilen?

Die Wahlforschung sucht nach Antworten auf diese Frage. Sie benutzt im Gegensatz zur nur tabellarischen Wahlstatistik deren Ergebnisse zur Erforschung und Entwicklung von Theorien zu individuellem oder auch gruppentypischen Wahlverhalten. Sie ist ein Versuch, zur Beurteilung der subjektiven Wahlentscheidung "objektive" Aussagesysteme zu entwickeln, die zur Beschreibung, Erklärung und Prognose von Wahlverhalten geeignet sind.

Das statistische Wahlergebnis setzt sich aus Teilergebnissen von Gruppen zusammen, deren Stimmabgabe in der Regel häufig voneinander abweicht. Stimmengewinne oder Verluste einer Partei sind der Saldo häufig gegenläufiger Bewegungen in den Teilgruppen.

Es erscheint sinnvoll, das Wahlergebnis hinsichtlich der Struktur der wahlberechtigten Bevölkerung zu beurteilen.

Dazu muß man zunächst versuchen, die Wähler gemäß bestimmter Eigenschaften Gruppen zuzuordnen. Mit Eigenschaften sind hier demographische Merkmale wie auch sozial-psychologische Einstellungen gemeint.

Eine Unterscheidung nach demographischen Gesichtspunkten erfolgt, wenn die Gruppen nach Alter, Geschlecht, Schulbildung, Beruf und vergleichbaren Merkmalen gegeneinander abgegrenzt werden.

Der Gruppierung liegen Einstellungen zugrunde, wenn man Unterscheidungs-

kriterien verwendet, wie z.B. Parteilichkeit, Orientierung an politischen Normen und ähnliches mehr. Durch Strukturierung nach solchen Merkmalen wird die Wählerschaft in abgrenzbare, in sich möglichst homogene Teilgruppen eingeteilt. Hierdurch wird eine Beschreibung des Wahlergebnisses möglich, die um so aussagefähiger wird, je homogener die Gruppen gebildet werden können. Dadurch können Strukturen politischen Verhaltens offengelegt werden, die das statistische Wahlergebnis zunächst verhüllt.

So ist es das Ziel dieser Untersuchung, unter Benutzung von Hypothesen über das quantitative Ausmaß gruppentypischen Wahlverhaltens die Wahlanalyse objektiver zu gestalten und das Wahlergebnis zu interpretieren.

Das verfügbare Datenmaterial begrenzt die Möglichkeiten der Forschung.

Durch die Volkszählung 1987 liegen zwar individuelle Kennzahlen für die gesamte Bevölkerung innerhalb eines Wahlgebietes vor, aber die Kennzahlen sind nur bis auf Blocksebene kleinräumig verfügbar. In der Wahlstatistik sind die Stimmbezirke die kleinste Ebene, für die Wahlergebnisse verfügbar sind.

Um nun die Wahlergebnisse in Relation zu den verfügbaren Zählungsdaten zu setzen, ist die kleinste verfügbare Untersuchungseinheit daher der Stimmbezirk. Hier liegen die Ergebnisse der geheimen Wahl als Summen vor. Ebenso müssen die Zählungsdaten auf Stimmbezirksebene "aggregiert" werden. Aggregatdaten sind also Informationen über die Verteilung von Merkmalen in einem bestimmten Gebiet, hier dem Stimmbezirk, z. B. der Anteil oder die absolute Zahl der Selbständigen, Frauen, Alleinlebenden, Arbeiter u.s.w., oder Durchschnittswerte von bestimmten Merkmalen in einem Stimmbezirk, z.B. die durchschnittliche Wohnfläche pro Person. Die Datenaggregate wie Verteilungen verändern sich, wenn man die Teilgebiete zu größeren Einheiten zusammenfaßt, z.B. Stimmbezirke zu Wahlbezirken oder nur über Teile der Gesamtbevölkerung innerhalb einer Gebietseinheit aggregiert. Um die Vergleichbarkeit zu gewährleisten, darf im Falle der Datenaggregation zum Zwecke der Wahlforschung nur für die "Wahlberechtigte Bevölkerung" aggregiert werden, d.h. nur für die deutsche Bevölkerung über 18 Jahre am Ort der Hauptwohnung.

Die Hypothese der Wahlforschung auf der Basis von Aggregatdaten lautet nun: Das politische Verhalten der Wähler wird geprägt durch das soziale (und weitere) Umfeld, in dem diese Wähler leben. Da die Aggregatdatenanalyse in diesem Sinne Beziehungen zwischen individuellem Verhalten von Menschen und deren "Umwelt" beschreibt, wird sie auch als "politische Ökologie"⁶ be-

⁶Der Begriff "Ökologie" bezeichnet alle Wechselwirkungen zwischen Individuum und seiner Umwelt.

zeichnet.

Neben der Wahl einer geeigneten Aggregatebene ist die Möglichkeit eines "ökologischen Fehlschlusses" (ecological fallacy) eines der Hauptprobleme bei der Analyse von Aggregatdaten. Diese Bezeichnung wurde vor 40 Jahren von Robinson⁷ vergeben für die unlogischen und irreführenden Schlußfolgerungen, die im allgemeinen häufig beim Schluß von (ökologischen) Aggregatbeziehungen auf individuelles Verhalten gemacht werden.

Wenn in einem Stimmbezirk mit hohem Arbeiteranteil gleichzeitig auch hohe Stimmenanteile für die SPD auftreten, dann ist die Aussage "die Arbeiter haben SPD gewählt" zwar naheliegend aber methodisch nicht haltbar. Im Extremfall ist es trotz Parallelität der hohen Werte möglich, daß kein einziger Arbeiter sich für die SPD entschieden hat. Robinson zeigte, daß Korrelationen auf Aggregatebene fast immer stärker sind als für die selben Variablen auf individueller oder einer anderen niedrigen Aggregatebene. Durch den Übergang auf eine höhere Aggregatebene kann sich im Extremfall sogar das Vorzeichen der Korrelation verändern, d.h. ein positiver Zusammenhang sich in einen negativen umkehren.

Die meisten der vorliegenden Wahlstudien erkennen zwar das Problem des ökologischen Fehlschlusses, ignorieren es aber dann in der Analyse. Wird dieses Vorgehen überhaupt gerechtfertigt, so durch den Hinweis, man wolle vorhersagende und nicht erklärende Modelle entwickeln.⁸

Die hiermit vorliegende Studie wendet ein Regressionsmodell im genannten ökologischen Kontext an und versucht, die Probleme des möglichen ökologischen Fehlschlusses durch Berücksichtigung der Kontextmerkmale eines Stimmbezirkes innerhalb nominalskaliertter Variablen zu lösen.

Ausgangspunkt ist der Selbständigenanteil als stetige Vorhersagevariable, der um nominale Prädikatoren erweitert wird.

Die theoretischen Begründungen und Rechtfertigungen für die benutzte Modellbildungstechnik sollen im Rahmen dieser Analyse nur kurz angeführt und diskutiert werden. In diesem Zusammenhang sei auf die in Kürze fertige Diplomarbeit des Verfassers verwiesen. Diese beschäftigt sich zusätzlich zum Regressionsansatz auch mit weiteren statistischen Methoden zur Charakterisierung und Quantifizierung von Unterschieden zwischen Stimmbezirken und faßt Literatur über ähnliche Untersuchungen (insbesondere aus dem Bereich der englischen Wahlforschung) zusammen.

⁷Robinson, W.S.(1950)

⁸z.B. bei Heere(1976)

4 Detaillierte Resultate

4.1 Für die CDU

Der erwartete Stimmenanteil der CDU innerhalb der Gruppe der wahlberechtigten Bevölkerung abzüglich der Wahlscheininhaber ist:
(Zum Vergleich: Stadtweit wurde die CDU bei der KW 1989 von 14,3 % der im Wählerverzeichnis ohne Sperrvermerk aufgeführten Personen gewählt)

innerhalb der Selbständigen:	13,04 %
innerhalb der Nicht-Selbständigen:	12,65 %

Darüber hinaus steigt und fällt dieser Erwartungswert abhängig von den nachfolgend aufgelisteten Kontextmerkmalen:

wenn im Stimmbezirk sehr viele ^a ledige Männer leben:	-1,50 %
wenn der Stimmbezirk sehr niedrige Arbeiteranteile aufweist:	+1,17 %
wenn der Stimmbezirk 1984 nur SPD-Hochburg war:	-1,52 %
wenn der Stimmbezirk 1984 nur CDU-Hochburg war:	+3,77 %
wenn der Stimmbezirk 1984 nur Nichtwähler-Hochburg war:	-1,60 %
wenn der Stimmbezirk 1984 SPD- und CDU-Hochburg war:	+3,20 %
wenn der Stimmbezirk 1984 SPD- und Nichtwähler-Hochburg war:	-2,36 %
wenn der Stimmbezirk 1984 nur SPD-Diaspora war:	+1,22 %
wenn der Stimmbezirk 1984 nur CDU-Diaspora war:	-3,22 %
wenn der Stimmbezirk 1984 nur Nichtwähler-Diaspora war:	+1,94 %
wenn der Stimmbezirk 1984 SPD- und Nichtwähler-Diaspora war:	+6,09 %
wenn der Stimmbezirk 1984 CDU- und Nichtwähler-Diaspora war:	-3,06 %

^aDie hier verwendeten Begriffe (sehr viel / sehr niedrig / Hochburg und Diaspora) sind in Kapitel 5.4.1 definiert. *Diaspora* bedeutet hier nur "relativ niedrig", im Gegensatz zur Hochburg, wo die entsprechenden Stimmenanteile "relativ hoch" sind.

Dazu zwei Beispiele:

Von den Nicht-Selbständigen in einem Stimmbezirk mit sehr wenigen ledigen Männern, sehr vielen Arbeitern, 1984 SPD-Hochburg und 1984 CDU und Nichtwähler Diaspora	12,65 % ± 0,00 % ^a ± 0,00 % ^b -1,52 % -3,06 %	Von den Selbständigen in einem Stimmbezirk mit sehr vielen ledigen Männern, sehr wenigen Arbeitern, 1984 CDU-Hochburg und 1984 SPD und Nichtwähler Diaspora	13,04 % -1,50 % +1,17 % +3,77 % +6,09 %
wählen CDU:	8,07 %	wählen CDU:	22,57 %

^aIm Modell ist der Faktor "sehr wenige ledige Männer" nicht enthalten, da er die Varianz der Stimmenanteile nicht signifikant beeinflusst

^bs.o.

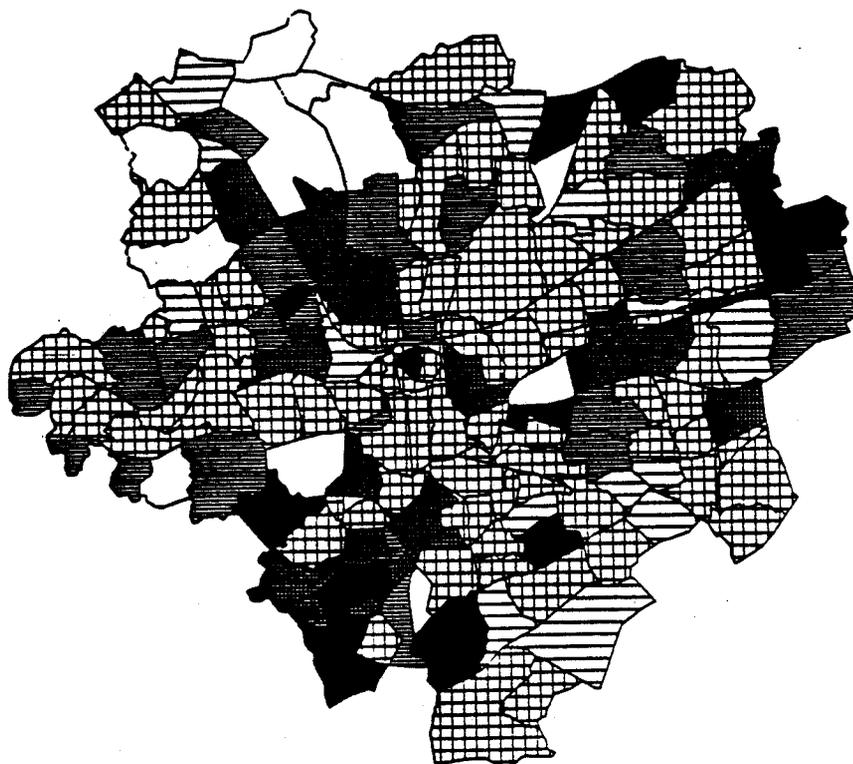
Für 580 Stimmbezirke in Dortmund sind die Anteile der Selbständigen sowie die übrigen Klassifikationsmerkmale (Arbeiteranteile, Anteile lediger Männer, Hochburgen- und Diasporasituation der Vorwahl) bekannt und somit konnten 580 Erwartungswerte für die CDU-Stimmenanteile berechnet werden. Die Abweichungen, mithin eine **Beurteilung des Wahlerfolges der CDU**, gibt die nachfolgende Karte wieder:

Detaillierte Einzelangaben zum Abschneiden von SPD und CDU in jedem der untersuchten 580 Stimmbezirke finden sich in Tabelle 10 im Anhang!!

ERGEBNIS DER CDU BEI DER KW 1989

IN DEN STATISTISCHEN UNTERBEZIRKEN

BEURTEILUNG DER CDU-ERGEBNISSE DURCH DIE BERECHNETEN ERWARTUNGSWERTE



ERGEBNIS CDU:  SEHR SCHLECHT
 GUT  SCHLECHT
 SEHR GUT  ERWARTET
 K.A. MOEGLICH

QUELLE: STÄDT DORTMUND, AMT FUER STATISTIK UND WAHLEN,
 UNIVERSITAET DORTMUND, FACHBEREICH STATISTIK

sehr schlecht: Abweichung vom Erwartungswert
 (ABWEI) kleiner
 -2,28 Prozentpunkte (PP),

schlecht: ABWEI zwischen
 -2,28 PP und -1,31 PP,

erwartet: ABWEI zwischen -1,31 PP und 1,10 PP

gut: ABWEI zwischen 1,10 PP und 2,77 PP,

sehr gut: ABWEI größer 2,77 PP

Abbildung 1: Graphische Darstellung der CDU-Ergebnisse bei der KW 1989

4.2 Für die SPD

Der erwartete Stimmenanteil der SPD innerhalb der Gruppe der wahlberechtigten Bevölkerung abzüglich der Wahlscheininhaber ist:

(Zum Vergleich: Stadtweit wurde die SPD bei der KW 1989 von 31,07 % der im Wählerverzeichnis ohne Sperrvermerk aufgeführten Personen gewählt)

innerhalb der Selbständigen:	32,35 %
innerhalb der Nicht-Selbständigen	32,77 %

Darüberhinaus steigt und fällt dieser Erwartungswert abhängig von den nachfolgend aufgelisteten Kontextmerkmalen:

wenn im Stimmbezirk sehr viele ^a ledige Männer leben:	-2,54 %
ledige Männer leben:	-2,54 %
wenn der Stimmbezirk sehr niedrige Arbeiteranteile aufweist:	-2,13 %
wenn der Stimmbezirk 1984 nur SPD-Hochburg war:	+6,51 %
wenn der Stimmbezirk 1984 nur CDU-Hochburg war:	-2,32 %
wenn der Stimmbezirk 1984 nur Nichtwähler-Hochburg war:	-4,32 %
wenn der Stimmbezirk 1984 SPD- und CDU-Hochburg war:	+4,27 %
wenn der Stimmbezirk 1984 nur SPD-Diaspora war:	-3,28 %
wenn der Stimmbezirk 1984 nur CDU-Diaspora war:	+1,90 %
wenn der Stimmbezirk 1984 nur Nichtwähler-Diaspora war:	+3,03 %
wenn der Stimmbezirk 1984 SPD- und Nichtwähler-Diaspora war:	-2,51 %
wenn der Stimmbezirk 1984 CDU- und Nichtwähler-Diaspora war:	+9,80 %

^aDie hier verwendeten Begriffe (sehr viel / sehr niedrig / Hochburg und Diaspora) sind in Kapitel 5.4.1 definiert. *Diaspora* bedeutet hier nur "relativ niedrig", im Gegensatz zur Hochburg, wo die entsprechenden Stimmenanteile "relativ hoch" sind.

Dazu wieder die gleichen Beispiele wie für die CDU:

Von den Arbeitern	32,77 %	Von den Selbständigen	32,35 %
in einem Stimmbezirk mit		in einem Stimmbezirk mit	
sehr wenigen ledigen Männern,	$\pm 0,00 \%^a$	sehr vielen ledigen Männern,	-2,54 %
sehr vielen Arbeitern,	$\pm 0,00 \%^b$	sehr wenigen Arbeitern,	-2,13 %
1984 SPD-Hochburg und	+6,51 %	1984 CDU-Hochburg und	-2,32 %
1984 CDU und Nichtwähler Diaspora	+9,80 %	1984 SPD und Nichtwähler Diaspora	-2,51 %
 wählen SPD:	 49,08 %	 wählen SPD:	 22,85 %

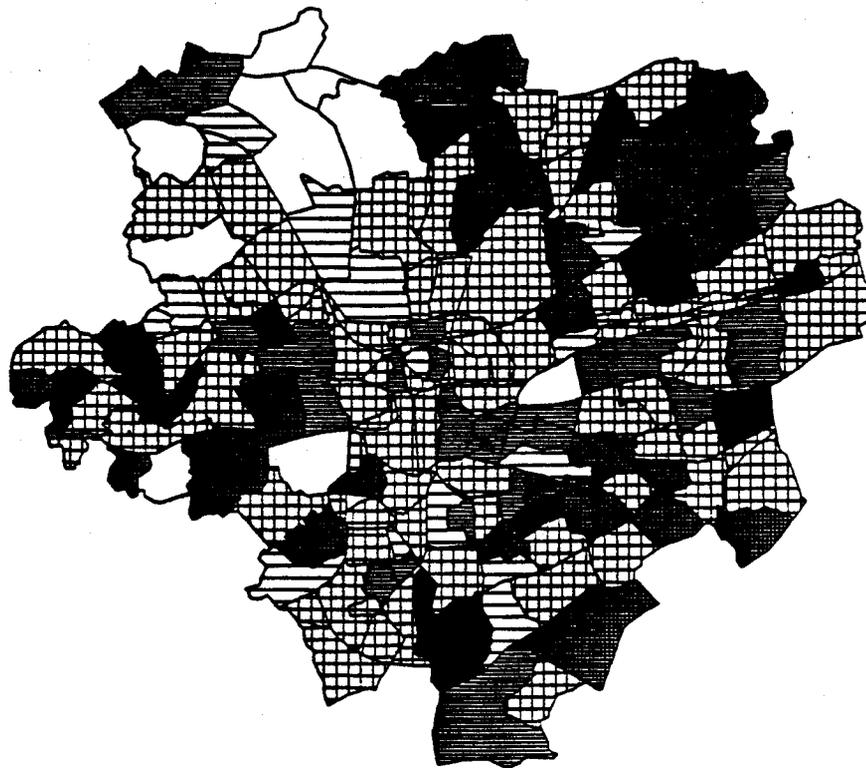
^aIm Modell ist der Faktor "sehr wenige ledige Männer" nicht enthalten, da er die Varianz der Stimmenanteile nicht signifikant beeinflusst
^bs.o.

Für 580 Direktwahlbezirke in Dortmund sind die Anteile der Selbständigen sowie die übrigen Klassifikationsmerkmale (Arbeiteranteile, Anteile lediger Männer, Hochburgen- und Diasporasituation der Vorwahl) bekannt und somit konnten 580 Erwartungswerte für die SPD-Stimmenanteile berechnet werden. Die Abweichungen, mithin eine Beurteilung des Wahlerfolges der SPD, gibt die nachfolgende Karte wieder:

Detaillierte Einzelangaben zum Abschneiden von SPD und CDU in jedem der untersuchten 580 Stimmbezirke finden sich in Tabelle 10 im Anhang!!

ERGEBNIS DER SPD BEI DER KW 1989

IN DEN STATISTISCHEN UNTERBEZIRKEN
BEURTELLUNG DER SPD - ERGEBNISSE DURCH DIE BERECHNETEN ERWARTUNGSWERTE



ERGEBNIS SPD:  SEHR SCHLECHT  SCHLECHT  ERWARTET
 GUT  SEHR GUT  K.A. MOEGLICH

QUELLE: STADT DORTMUND, AMT FUER STATISTIK UND WAHLEN,
UNIVERSITAET DORTMUND, FACHBEREICH STATISTIK

sehr schlecht: Abweichung vom Erwartungswert
(ABWEI) kleiner
-3,06 Prozentpunkte (PP),

erwartet: ABWEI zwischen -1,97 PP und 2,38 PP

gut: ABWEI zwischen 2,38 PP und 4,08 PP,

schlecht: ABWEI zwischen
-3,06 PP und -1,97 PP,

sehr gut: ABWEI größer 4,08 PP

Abbildung 2: Graphische Darstellung der SPD-Ergebnisse bei der KW 1989

5 Methodische Strategien und Modelle

5.1 Das Regressionsmodell

5.1.1 Grundsätzliches über Korrelations- und Regressionsrechnung

Bei der empirischen Datenanalyse stellt sich oft die zentrale Aufgabe, eine Fülle von Beobachtungen, Daten, Erscheinungen in eine Ursache-Wirkungs-Beziehung zu bringen, um sich ein Bild über die Funktionsweise des untersuchten Gegenstandes zu machen.

Dazu müssen zunächst Zusammenhänge zwischen Beobachtungen untersucht werden. In der Wahlforschung der Bundesrepublik ist die Korrelationsrechnung das am weitesten verbreitete statistische Instrument zur Aufdeckung und Verifikation bivariater linearer Zusammenhänge, z.B. zwischen Stimmenanteilen zweier Parteien oder auch zwischen soziodemographischen Bestimmungsgrößen und dem Stimmenanteil einer Partei.

Wurde eine Folge von Paaren zweier Merkmalsausprägungen erhoben, z.B. das Paar (Arbeiteranteil / SPD-Stimmenanteil) in jedem der 620 Dortmunder Stimmbezirke, so drückt ein Korrelationskoeffizient die Stärke des linearen Zusammenhanges zwischen diesen beiden Merkmalen aus. Sind beide Merkmale metrisch skaliert, d.h. Abstände zwischen zwei Merkmalsausprägungen quantitativ interpretierbar, so ist der Korrelationskoeffizient nach Bravais-Pearson⁹ ein geeignetes Maß zur Beurteilung des linearen Zusammenhanges zwischen den beiden Merkmalen. Der Koeffizient r liegt immer zwischen -1 und 1 , und zwar je näher bei 1 , desto gleichgerichteter (positiver) der zwischen beiden Merkmalen bestehende lineare Zusammenhang ist, und je näher bei -1 , desto entgegengerichteter (negativer) dieser Zusammenhang ist. Ein Wert von r nahe der 0 indiziert, daß keine lineare Beziehung zwischen den beiden Merkmalen besteht. Man kann auch alle erhobenen Merkmale in ein Koordinatenkreuz eintragen. Auf der x -Achse trägt man den Wert für das eine Merkmal und auf der y -Achse den Wert für das andere Merkmal auf. Ergebnis ist eine zweidimensionale "Punktwolke", die den Zusammenhang zwischen den beiden Merkmalen anschaulich macht. Dieser ist um so linearer, je deutlicher die Punkte auf einer Geraden liegen.

In Abbildung 3 sind einige Paare von Merkmalsausprägungen mit dem zugehörigen Wert des Korrelationskoeffizienten r graphisch dargestellt:

⁹Siehe Hartung/Elpelt (1984)

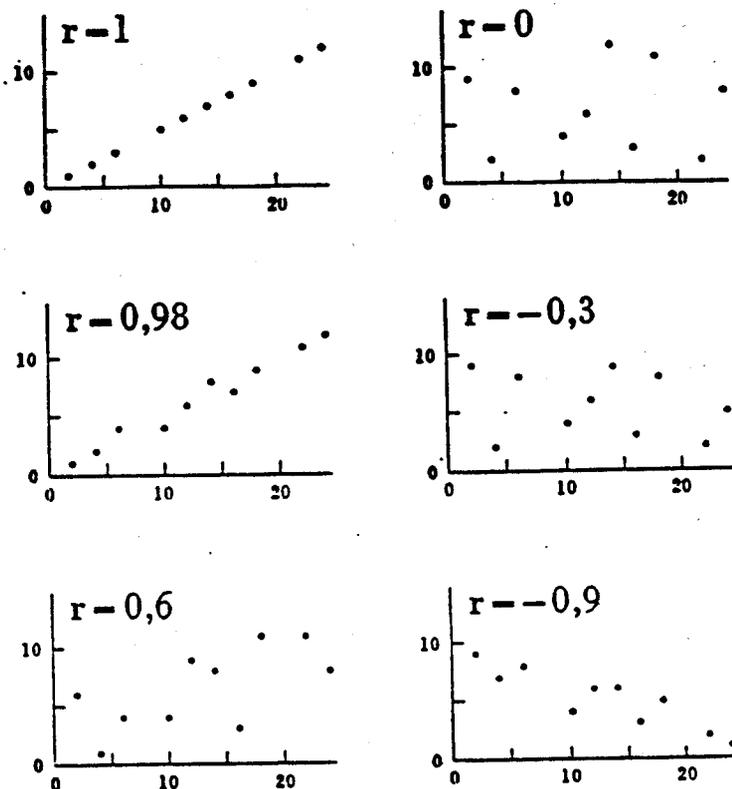


Abbildung 3: Beispiele für Stichproben mit verschiedenen Werten des Korrelationskoeffizienten r

Hat man nun, indiziert durch den Wert von r , einen linearen Zusammenhang zwischen zwei Merkmalen festgestellt, erlaubt die Regressionsanalyse, diesen in Form einer Ursache-Wirkung-Beziehung zu quantifizieren.

5.1.2 Das einfache lineare Regressionsmodell

Ausgehend von einer plausiblen Arbeitshypothese, z.B. der SPD-Stimmenanteil wird durch den Arbeiteranteil "erklärt", wird eine Variable als unabhängig (exogen) und eine andere als abhängig (endogen) und in ihrem Wert durch den Wert der unabhängigen Variable funktional bedingt angesehen.

Hat man einen solchen linear-funktionalen Zusammenhang unterstellt, so versucht die Regressionsanalyse diejenige lineare Funktion (Ausgleichsgerade) zu bestimmen, die sich der Lage der Punktwolke im oben erwähnten Koordinatenkreuz möglichst genau anpaßt.

Diese Anpassung geschieht nach einer objektiven Methode, dem gauß'schen Prinzip der kleinsten Quadrate. Für unser gegenwärtiges Problem besagt dieses Prinzip folgendes:

Die Gerade ist so zu legen, daß die Summe der Quadrate aller Abstände der Punkte des Koordinatenkreuzes von der Geraden möglichst klein wird.

Man benutzt die Quadrate der Abstände, um die Analyse rechentechnisch zu vereinfachen.

Als Abstandsmaß dient ebenfalls aus rechentechnischen Gründen der vertikale Abstand von Punkt und Gerade.

Zu jedem Wert x_i der unabhängigen Variable stellt die einfache lineare Regression einen linearen Zusammenhang her zu dem Wert \hat{y}_i , dem geschätzten Wert der abhängigen Variable. Die Punkte \hat{y}_i liegen für alle i auf der Ausgleichsgeraden.

Es gilt die Beziehung:

$$\hat{y}_i = a + b \cdot x_i$$

Zu minimieren ist die Summe der quadrierten Abstände zwischen den vertikalen Koordinaten der Punktwolke (das sind die y_i) und den vertikalen Koordinaten der korrespondierenden geschätzten Werte auf der Geraden (das sind die \hat{y}_i).

Liegen n Beobachtungspaare von abhängiger und Vorhersagevariable vor, so ist zu minimieren:

$$\sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n [y_i - (a + b \times x_i)]^2 \quad (1)$$

Der Term (1) nimmt sein Minimum an, für:

$$b = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad a = \bar{y} - b \cdot \bar{x}$$

wobei

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \text{ der Mittelwert der Vorhersagevariable und}$$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i \text{ der Mittelwert der abhängigen Variable ist.}$$

a und b heißen auch Kleinste-Quadrate-Schätzer (KQ-Schätzer) für die "wahren" Parameter α und β derjenigen Funktion, die auch die zufälligen Fehler ϵ_i der Regression berücksichtigt.

(Diese ϵ_i sind gerade die Abstände zwischen \hat{y}_i und y_i , die es zu minimieren galt).

Spricht man von einem einfachen linearen Regressionsmodell der abhängigen Variable y auf die unabhängige Vorhersagevariable x , so heißt dies, daß für die abhängige Variable y ein linearer Zusammenhang der Form: $y_i = \alpha + \beta \cdot x_i + \epsilon_i$, wobei die $\epsilon_1, \dots, \epsilon_n$ zufällige Fehler sind, unterstellt wird und die Parameter α und β nach der KQ-Methode durch a und b wie oben geschätzt werden. Dadurch werden die Fehler ϵ_i minimiert.

Um die Genauigkeit der Parameter a und b wie auch die Qualität der Modellanpassung bestimmen zu können, sind für die Fehlervariable ϵ die nachfolgenden Eigenschaften zu fordern:

1. Der Erwartungswert der Größe ϵ ist Null (d.h. die Fehlerterme addieren sich insgesamt zu Null auf).
2. Betrachtet man ϵ als Komponente von y und y als bedingte Verteilung der x_i , so besitzt ϵ für alle vorkommenden Ausprägungen der x_i dieselbe Varianz σ_ϵ^2 .
3. Die Verteilung von ϵ ist eine Normalverteilung der Form $N(0, \sigma_\epsilon^2)$.

Sind diese drei postulierten Eigenschaften für den Fehlerterm ϵ nicht nachzuweisen, so ist zu vermuten, daß die einfache lineare Regression auf x die Variable y nicht hinreichend erklärt.

Neben vielen anderen Gründen ist es möglich, daß die Eigenschaften (1) bis (3) deshalb nicht von ϵ erfüllt werden, weil die Werte von einer oder mehreren anderen weiteren Vorhersagevariablen nicht berücksichtigt wurden, obwohl sie für die Vorhersage von y von Wichtigkeit sind.

In einem solchen Fall ist es notwendig, ein multiples Regressionsmodell zur Erklärung der y zu spezifizieren.

5.1.3 Der Fall mehrerer Einflußvariablen

Die "Erklärung" der Variation einer abhängigen Variablen durch eine Linearkombination mehrerer anderer Variablen (Prädikatoren, Vorhersagevariablen) wird mit dem Begriff "multiple lineare Regression" umschrieben. Wenn Y die abhängige Variable¹⁰ bezeichnet und X_1, X_2, \dots, X_n die p Vorhersagevariablen, so lautet der Regressionsansatz:

$$Y = \alpha_0 + \alpha_1 \cdot X_1 + \alpha_2 \cdot X_2 + \dots + \alpha_i \cdot X_i \dots + \alpha_p \cdot X_p + \epsilon \quad (2)$$

Unter schon für den einfachen Regressionsansatz genannten Annahmen lassen sich durch Anwendung der KQ-Methode erwartungstreue und unverzerrte Schätzwerte für die "wahren" Parameter α_i der Grundgesamtheit bestimmen.

Eine Schätzung der Parameter nach der KQ-Methode impliziert, daß die Varianz der Fehlergröße ϵ minimal ist.

Ist \hat{Y} der geschätzte Wert von Y , d.h. der bis auf den Fehler ϵ aus den X_1, X_2, \dots, X_n reproduzierbare Anteil von Y , so ist die Korrelation zwischen \hat{Y} und Y gleichzeitig die maximale Korrelation zwischen der abhängigen Variablen und der Linearkombination von Vorhersagevariablen.

Das Quadrat dieser Korrelation $R_{\hat{Y}Y}^2$ wird als sog. *Bestimmtheitsmaß* der multiplen Regression bezeichnet. Dieses Maß drückt aus, bis zu welchem Ausmaß die Varianz der abhängigen Variablen durch die Varianz der Linearkombination von Vorhersagevariablen reproduzierbar ist.

5.2 Das Problem des "ökologischen Fehlschlusses" bei Aggregatdaten

Wie schon in Kapitel 2 kurz erwähnt, fallen unter den Begriff des "ökologische Fehlschlusses" alle falschen Schlüsse auf individuelles Verhalten aufgrund ökologischer Parallelitäten. Diese sind naheliegend in dem Sinn, daß Korrelationen auf hoher Aggregatebene fast immer stärker sind, als für dieselben Variablen auf individueller oder anderer niedrigerer Aggregatebene. Der Fehler wird um so wahrscheinlicher, je weniger Aggregateinheiten gebildet werden und je geringer der Anteil ist, zu dem eine Variable in einer Einheit

¹⁰Die Variablen sind in diesem Abschnitt groß geschrieben, um sie als Zufallsvariablen zu identifizieren. Das ist notwendig, um die Annahmen zu erfüllen, die für "statistische Genauigkeitsaussagen" notwendig sind.

beiträgt. Durch fachliche Überlegungen kann man sicherlich einige irreführende Inferenzen verwerfen, dennoch bleibt die Frage, ob herkömmliches statistisches Instrumentarium überhaupt geeignet ist, aus Kontextmerkmalen auf individuelles Verhalten zu schließen.

Seit der ersten Erwähnung 1950 gab es zahlreiche Arbeiten, die sich mit der Problematik beschäftigten. So zeigte Goodmann(1953) und (1959), daß unter bestimmten Voraussetzungen die Technik der Regressionsanalyse als Mittel benutzt werden darf, um von Aggregatdaten Schlüsse über "Individualverhalten" zu machen.

Ein Beispiel soll die Problematik verdeutlichen. In Tabelle 1 ist eine Vierfeldertafel dargestellt. Sie zeigt die Verteilung der SPD-Wähler und der Nicht-SPD-Wähler innerhalb der "Arbeiterbevölkerung":

	Arbeiteranteil	Anteil der Nicht-Arbeiter	innerhalb aller Wähler
Anteil der SPD-Wähler	p_1	r_1	S
Anteil der anderen Wähler	p_2	r_2	S
innerhalb aller Wähler	A ^a	$1 - A$	$p_1 + p_2 = 1$ $r_1 + r_2 = 1$

^a A und S sind als Aggregatdaten bekannt, die p 's und r 's sind unbekannt.

Es gibt zwar Beziehungen zwischen A , S und den p 's und r 's, dem Autor dieser Studie ist aber keine einzige Studie aus der Bundesrepublik bekannt, die diesen Beziehungen Rechnung trägt und ihre Methodik aus ihnen elaboriert.

Tabelle 1: Beispiel ökologischer Beziehungen zwischen SPD-Wählern und Arbeiterbevölkerung

In der 1977 erscheinenden Promotionsschrift von *K.H. Heere* unter dem Titel "Anwendung Statistischer Modelle für wahlanalytische Untersuchungen auf der Basis von Aggregatdaten", einer Diskussion unterschiedlicher statistischer Ansätze zur Wahlanalyse auf der Basis der Ergebnisse der Volkszählung und der Unterhauswahlen in England 1966 auf Wahlkreisebene, wird das Problem am deutlichsten angesprochen. So heißt es:

"Hypothesen über Parteipräferenzen basieren also auf Attributen eines Sozialaggregats. Die Aussageeinheit ist immer der Wahlkreis und nicht der Wähler als Individuum.

Statistisch betrachtet wäre eine individuell interpretierte, aber

ökologisch gewonnene Beziehung ein Rückschluß von den auf Gebietebene gemessenen Randverteilungen zweier oder mehrerer Merkmale auf die gemeinsame Verteilung der Merkmale innerhalb einer Erhebungseinheit." ¹¹

Heere bemerkt dann allerdings sofort, daß extreme Unterschiede zwischen ökologischer und individueller Beziehung im allgemeinen nicht zu erwarten seien. Da Wahlkreise hinsichtlich ihrer sozialökonomischen Merkmale oft sehr homogen sind, würde die Dominanz einer Schicht letztlich bewirken, "daß ein Wahlkreis sich quasi als 'ökologisches Individuum' darbietet." ¹² Da *Heere* im weiteren an vorherrschenden und typisierenden Modellen gelegen ist, widmet er sich dem Problem nicht weiter.

Bei sorgsamer Interpretation der gewonnenen Ergebnisse ist dies durchaus ein begehbarer Weg. Es zeigt sich allerdings immer wieder, daß die genannte Fehlschlußproblematik immer wieder *nicht* beachtet wird.

Ein jüngeres Beispiel dafür ist die kürzlich wieder aufgenommene Debatte darüber, ob die Massenarbeitslosigkeit in der Weimarer Republik tatsächlich den Aufstieg des Nationalsozialismus bewirkt hat, ob Arbeitslose also nationalsozialistisch gewählt haben.

Diese These wurde auf der Basis von Aggregatdaten von *Frey und Weck (1981)* in Übereinstimmung mit vielen der heutigen Historiker aufgestellt. Korreliert man die Anteile der NSDAP und der Arbeitslosen zu den Zeitpunkten der 5 Reichstagswahlen 1928 und im März 1930, so erhält man einen Korrelationskoeffizienten von $r = 0.98$.

Von *Frey und Weck* wird daher die Arbeitslosigkeit als ausschlaggebender Faktor für nationalsozialistisches Wahlverhalten interpretiert. Dies muß jedoch im wesentlichen als ökologischer Fehlschluß gesehen werden. *Falter u.a. (1983)* wiesen nach, daß schon die Korrelationen auf Ebene der 13 Arbeitsamtbezirke wesentlich schwächer sind. Die Untersuchung des Zusammenhangs auf Basis der 865 Stadt- und Landkreiseinheiten des Deutschen Reiches führte sogar zu einem deutlich negativen Ergebnis in dem Sinne, daß im Durchschnitt der Kreise, in denen die Arbeitslosenquote überdurchschnittlich hoch war, der NSDAP-Stimmenanteil deutlich unter dem Durchschnitt lag.

Die vorliegende Studie erörtert die Problematik der "ökologischen Daten"

¹¹ *Heere, K.H. (1977)*, Seite 13

¹² *ebenda*, Seite 13

wesentlich und versucht, sie modelltechnisch zu operationalisieren. Mit Hilfe von multiplen Regressionsmodellen wird versucht, die Erwartungswerte der unbekanntenen Zellproportionen in Abhängigkeit zu den vorliegenden Daten aus der Volkszählung zu bringen, um dann aus den bekannten Beziehungen der Randwerte zu den Zellproportionen eine Abhängigkeit zu den Stimmenanteilen bestimmter Parteien bzw. der Nichtwähler herzustellen.

5.3 Ein multiples Regessionsmodell zur Erklärung der Stimmenanteile der einzelnen Parteien

5.3.1 Vorüberlegungen

Im Vorfeld einer Regressionsanalyse steht eine "plausible Arbeitshypothese" (Siehe Kapitel 4.1.2). In der vorliegenden Analyse wurden dazu Korrelationen zwischen allen erhobenen Merkmalen aus der Volkszählung und den Stimmenanteilen der einzelnen Parteien berechnet.

Am höchsten mit dem Wahlergebnis der einzelnen Parteien korreliert erwiesen sich die Indikatoren hoher Bildung (Anteil der Universitäts- bzw. Fachhochschulabsolventen), der Selbständigenanteil, sowie der Anteil der großen Wohnungen mit 81qm und mehr an Fläche.

Ingesamt wurden 117 Variable aus der Volkszählung betrachtet und die Korrelationen zu den Stimmenanteilen der Parteien berechnet. Abhängig von der Höhe dieser Korrelationen ließen sich Gruppen von Variablen bilden, die in gleicher Weise mit den Stimmenanteilen der einzelnen Parteien korreliert sind¹³.

Aus diesen Gruppen sind die o.a. Variablen diejenigen, die zu verschiedenen Gruppen gehören und die höchsten Einzelkorrelationen zu den Stimmenanteilen von CDU und SPD aufweisen.

Nimmt man den Anteil der Universitäts- und Fachhochschulabsolventen zusammen (diese sind mit $r = 0.8$ miteinander korreliert), so verringern sich die Korrelationen der so neu gebildeten Variablen mit den Stimmenergebnissen der einzelnen Parteien. Der Selbständigenanteil und der Anteil der großen Wohnungen ist hoch negativ mit dem SPD-Stimmenanteil und hoch positiv mit dem CDU-Stimmenanteil korreliert.

Es bietet sich daher als Arbeitshypothese an, den CDU-Stimmenanteil durch eine Regression auf den Selbständigenanteil zu erklären.

Variablen, die in anderen Analysen eine wichtige Rolle spielen, wie z.B. der Katholikenanteil, der Anteil der Beschäftigten in der Landwirtschaft und der Migrationssaldo, die bei Mielke(1987) die wichtigsten Erklärungsfaktoren sind, oder der Arbeiteranteil in den Analysen von Crewe und Payne(1976) scheinen im Falle der Dortmunder Daten für die Wahlentscheidung weniger wichtig zu sein bzw. sind nicht direkt zu ermitteln.

¹³Dazu wurde eine Clusteranalyse nach dem *single linkage*-Verfahren durchgeführt.

So wurde der Anteil der Selbständigen an der wahlberechtigten Bevölkerung zunächst als einzige exogene Variable zur Erklärung der Stimmenanteile der einzelnen Parteien gewählt.

Im folgenden soll ausgehend von dieser Wahl untersucht werden, wie hoch unter den Selbständigen der Anteil derjenigen ist, die SPD, CDU, GRÜNE, F.D.P. oder REP oder gar nicht gewählt haben¹⁴.

Es sollen die Zelleneinträge p_i und r_i in der nachfolgenden Tabelle anhand der bekannten Randhäufigkeiten geschätzt werden.

	Selbständigen anteil	Anteil der Nicht- Selbständigen	innerhalb aller Wähler
Anteil der SPD-Wähler	p_1	r_1	S
Anteil der CDU-Wähler	p_2	r_2	C
Anteil der GRÜNE-Wähler	p_3	r_3	G
Anteil der FDP-Wähler	p_4	r_4	F
Anteil der REP-Wähler	p_5	r_5	R
Anteil der sonstig. Wähler	p_6	r_6	SO
Anteil der Nichtwähler	p_7	r_7	N
innerhalb aller Wähler	SE	$1 - SE$	$\sum_{i=1}^7 p_i = 1$ $\sum_{i=1}^7 r_i r_1 = 1$

Tabelle 2: Parteipräferenzen der Selbständigen als theoretische Anteile

Nun gibt es Beziehungen zwischen den bekannten Randhäufigkeiten SE und S bis N zu den Zellwerten. So läßt sich der CDU-Stimmenanteil für jeden Stimmbezirk in der folgenden Form schreiben:

$$\begin{aligned} C &= p_2 \cdot SE + r_2 \cdot (1 - SE) \\ &= r_2 + (p_2 - r_2) \cdot SE \end{aligned} \quad (3)$$

¹⁴Durchgeführt wurde diese Analyse schließlich nur für die Anteile der SPD und der CDU.

Man kann nun innerhalb einer gegebenen Menge von Stimmbezirken (z.B. den 620 in Dortmund) den erwarteten oder "durchschnittlichen" Anteil von CDU-Stimmen innerhalb der wahlberechtigten Bevölkerung bei gegebenem Selbständigenanteil $E(C/SE)$ betrachten.

In gleicher Weise definiert man den Erwartungswert des Anteils der Selbständigen, die CDU wählen, bei gegebenem Selbständigenanteil als $E(p_2/SE)$ und den erwarteten Anteil der Nicht-Selbständigen, die CDU wählen, als $E(r_2/SE)$.

Unter der Annahme, daß die Erwartungswerte $E(p_2/SE)$ und $E(r_2/SE)$ innerhalb aller Stimmbezirke, auf denen die Regression aufbaut, konstant sind, hat (1) die Form eines einfachen Regressionsmodells der Form

$S = a_2 + b_2 \cdot SE$. a_2 und b_2 könnten dann nach der KQ-Methode geschätzt werden und damit auch die Zellproportionen p_2 und r_2 .

Allerdings kann nur dann so verfahren werden, wenn die genannten Proportionen p_2 und r_2 in allen 620 Dortmunder Stimmbezirken dieselben sind (bzw. zufällig um den geschätzten Mittelwert herum verteilt sind), also der Anteil der Selbständigen, die CDU wählen innerhalb Dortmunds in keiner Weise systematisch von Stimmbezirk zu Stimmbezirk differiert.

Aber eine solche Annahme wäre sehr unrealistisch. Es ist vielmehr anzunehmen, daß dieser Anteil abhängig ist von weiteren Charakteristika der Stimmbezirke.

Ein einfaches Regressionsmodell scheidet daher als Analyseinstrument aus. Es bieten sich zwei unterschiedliche Wege an, die Problematik zu lösen:

- (a) Man unterteilt die 620 Stimmbezirke in Klassen, innerhalb derer der Anteil der Selbständigen, die CDU wählen, annähernd konstant ist und schätzt die Parameter p_2 und r_2 in jeder Klasse extra.
- (b) Man formuliert ein multiples Regressionsmodell, daß den Stimmenanteil der einzelnen Parteien (zunächst der CDU) nicht nur durch den Selbständigenanteil erklärt, sondern zudem noch die Variablen als Erklärungsfaktoren berücksichtigt, die das Wahlverhalten der Selbständigen beeinflussen.

Das größte Problem im Fall (a) ist, daß die Größe der Untergruppen von Stimmbezirken über die Aussagen gemacht werden können, stark abnimmt, je genauer das Wahlverhalten der Selbständigen untersucht und nach anderen Kriterien differenziert wird. Da hier nur insgesamt 620 Stimmbezirke untersucht werden, sind die Möglichkeiten zur Bildung von Untergruppen daher nicht sehr groß.

Die Ergebnisse sind bei Untergliederung der Stimmbezirke und separaten Untersuchungen auch schwieriger zu interpretieren.

Daher bietet sich die zweite Möglichkeit eines multiplen Regressionsmodells eher zur Lösung des Problems an.

5.3.2 Unterschiedliche Modellansätze

Wir können nun den einfachen ökologischen Regressionsansatz (3) erweitern zu einem multivariaten Modell.

Ausgangspunkt dafür ist die nachfolgende Generalisierung von (3):

$$E(CDU/Selbst.) \\ E(C/SE) = E(r_2/SE) + (E(p_2/SE) - E(r_2/SE)) \cdot SE \quad (4)$$

welcher sich auf die nachfolgenden Größen bezieht:

$E(C/SE)$ ist der erwartete Anteil der CDU-Wähler bei gegebenem Selbständigenanteil

$E(r_2/SE)$ ist der Anteil der Nicht-Selbständigen, die CDU wählen, bei gegebenem Selbständigenanteil

$E(p_2/SE)$ ist der Anteil der Selbständigen, die CDU wählen, bei gegebenem Selbständigenanteil

Die Gleichung (2) hat die beiden grundlegenden Komponenten $E(r_2/SE)$ und $E(p_2/SE) - E(r_2/SE)$, von denen im folgenden angenommen werden soll, daß sie abhängig sind von einer Linearkombination anderer Variablen. Für $E(r_2/SE)$ werden die beiden nachfolgenden Abhängigkeiten betrachtet:

$$(a) \quad E(r_2/SE) = a_0$$

$$(b) \quad E(r_2/SE) = a_0 + a_1X_1 + a_2X_2 + \dots + a_nX_n$$

(a) besagt, daß der Anteil r_2 der Nicht-Selbständigen, die CDU wählen, statistisch, d.h. im Mittel, konstant in allen 620 Dortmunder Stimmbezirken ist.

Die Form (b) unterstellt eine Abhängigkeit von r_2 von einer Linearkombination einer Menge $\{X_1, X_2, \dots, X_n\}$ von Vorhersagevariablen.

Ebenso werden für die Komponente $E(p_2/SE) - E(r_2/SE)$ in (2) zwei Formen untersucht:

$$(i) E(p_2/SE) - E(r_2/SE) = b_0$$

$$(ii) E(p_2/SE) - E(r_2/SE) = b_0 + b_1 Y_1 + b_2 Y_2 + \dots + b_n Y_n$$

wobei $\{Y_1, Y_2, \dots, Y_n\}$ eine Menge von Vorhersagevariablen ist, die auch jede der in (b) benutzten Variablen X_1, X_2, \dots, X_n enthalten können.

Die Form (i) unterstellt, daß die *Differenz* der CDU-Wähler bei Selbständigen und Nicht-Selbständigen statistisch konstant in ganz Dortmund ist.

In (ii) ist eine Abhängigkeit dieser Differenz von einer Menge von Vorhersagevariablen $\{Y_1, Y_2, \dots, Y_n\}$ unterstellt.

Vier Paare von Annahmen lassen sich mit den genannten Ansätzen untersuchen:

$$(a)(i): \quad \begin{aligned} E(r_2/SE) &= a_0 \\ E(p_2/SE) - E(r_2/SE) &= b_0 \end{aligned}$$

damit ergibt sich das allgemeine Modell (2) zu:

$$E(C/SE) = a_0 + b_0 \cdot SE,$$

also ein einfaches Regressionsmodell, welches den CDU-Stimmenanteil in lineare Abhängigkeit zum Selbständigenanteil setzt.

Berechnet man die Parameter a_0 und b_0 für dieses Modell auf der Basis der 620 Dortmunder Stimmbezirke, so erhält man die nachfolgenden Parameterschätzungen:

$$\begin{aligned} a_0 &= 8,95 = E(r_2/SE) && \text{und} \\ b_0 &= 1,79 = E(p_2/SE) - E(r_2/SE) \end{aligned}$$

und damit wegen $E(p_2/S) = a_0 + b_0$ ergibt sich $E(p_2/S) = 10,74$

Unter der Annahme, daß das obige Modell richtig ist, sind die Schätzungen dahingehend zu interpretieren, daß innerhalb der wahlberechtigten Bevölkerung Dortmunds abzüglich der Wahlscheininhaber 9,0% der Nicht-Selbständigen CDU wählen, währenddessen der Anteil der CDU-Wähler innerhalb der Selbständigen 10,7% beträgt.

Dieses Modell erklärt aber nur 35% der gesamten Varianz und muß daher als unvollständig verworfen werden.

Begründetere Schätzungen sind zu erwarten, wenn man systematische Variationen des Anteils der Selbständigen und Nicht-Selbständigen, die CDU wählen, im Modellansatz zuläßt, wie dies in den drei anderen Paaren von Annahmen der Fall ist.

$$(b)(i): \quad \begin{aligned} E(r_2/SE) &= a_0 + a_1X_1 + a_2X_2 + \dots + a_nX_n \\ E(p_2/SE) - E(r_2/SE) &= b_0 \end{aligned}$$

damit ergibt sich:

$$E(C/SE) = a_0 + a_1X_1 + a_2X_2 + \dots + a_nX_n + b_0 \cdot SE,$$

also die Annahme, daß r_2 und p_2 in der selben Weise linear abhängig von einer Menge $\{X_1, X_2, \dots, X_n\}$ von Vorhersagevariablen sind, mit einer konstanten Differenz zwischen den beiden Proportionen.

Diese Annahme entspricht einer Regression des CDU-Stimmenanteils auf die Vorhersagevariablen X_1, X_2, \dots, X_n und SE .

$$(a)(ii): \quad \begin{aligned} E(r_2/SE) &= a_0 \\ E(p_2/SE) - E(r_2/SE) &= b_0 + b_1Y_1 + b_2Y_2 + \dots + b_nY_n \end{aligned}$$

damit ergibt sich:

$$E(C/SE) = a_0 + (b_0 + b_1Y_1 + b_2Y_2 + \dots + b_nY_n) \cdot SE,$$

damit also die Annahme, daß die Proportion r_2 der CDU-Wähler unter den Nicht-Selbständigen wahlberechtigten Bevölkerung statistisch konstant ist, währenddessen der Anteil der CDU-Wähler unter den Selbständigen systematisch variiert linear abhängig von den Interaktionen zwischen Y_1, Y_2, \dots, Y_n und SE .

Es bedarf wohl kaum weiteren Erläuterungen, daß dies eine äußerst unrealistische Annahme darstellt.

$$(b)(ii): \quad \begin{aligned} E(r_2/SE) &= a_0 + a_1X_1 + a_2X_2 + \dots + a_nX_n \\ E(p_2/SE) - E(r_2/SE) &= b_0 + b_1Y_1 + b_2Y_2 + \dots + b_nY_n \end{aligned}$$

damit ergibt sich:

$$\begin{aligned}
 E(C/SE) &= a_0 + a_1X_1 + a_2X_2 + \dots + a_nX_n \\
 &\quad + (b_0 + b_1Y_1 + b_2Y_2 + \dots + b_nY_n) \cdot SE \\
 &= a_0 + a_1X_1 + a_2X_2 + \dots + a_nX_n \\
 &\quad + b_0SE + b_1X_1SE + \dots + b_nY_nSE
 \end{aligned}$$

Dies ist also das Modell, welches annimmt, daß r_2 und p_2 beide linear abhängig von einer Menge von Vorhersagevariablen sind, aber in unterschiedlicher Weise.

Eine solche Annahme ist gleichbedeutend mit einer multiplen Regression des CDU-Stimmenanteils auf die Menge von Vorhersagevariablen $\{X_1, X_2, \dots, X_n\}$ und Interaktionen zwischen einer weiteren Menge von Vorhersagevariablen $\{Y_1, Y_2, \dots, Y_n\}$ mit dem Selbständigenanteil SE. Die Menge $\{Y_1, Y_2, \dots, Y_n\}$ kann die Menge $\{X_1, X_2, \dots, X_n\}$ ganz oder teilweise enthalten.

Solche "Doppelerklärungen" des Stimmenanteils durch $\{X_1, X_2, \dots, X_n\}$ und $\{Y_1, Y_2, \dots, Y_n\}$ sind sehr wahrscheinlich. Wenn der Anteil der CDU-Wähler unter den Selbständigen nämlich unter dem Einfluß einer anderen Variablen, z.B. dem Anteil der Universitätsabsolventen variiert, dann wäre es schon verwunderlich, wenn der Anteil der Universitätsabsolventen keinen Einfluß auf den Anteil der CDU-Wähler unter den Nicht-Selbständigen hat.

Ein weiteres Problem taucht bei den letzten beiden betrachteten Modellen auf, nämlich das der Multikolarität. In den beiden Regressionsansätzen tauchen zahlreiche Terme auf, die die Interaktionen von bestimmten Vorhersagevariablen mit dem Selbständigenanteil beinhalten. Es ist zu erwarten, daß diese Terme hoch miteinander korreliert sind.

Miller(1972) beschreibt die Anwendung von *Ridge-Regression* zur Lösung dieses Problems.

Im Vorfeld der Regressionsanalyse wurden im vorliegenden Fall schon einige Vorhersagevariablen aufgrund ihrer Korrelation mit dem Wahlergebnis aus der Menge aller aus der Volkszählung erhobenen Variablen selektiert.

Diese ausgewählten Variablen wurden durch Clusteranalyse in Gruppen eingeteilt. Dies geschah unter dem Vorsatz, im Vorfeld der Regressionsanalyse dem Problem der Multikolarität vorzubeugen.

Zusätzlich stehen zur Analyse die Ergebnisse der Kommunalwahl in Dortmund von 1984, soweit damals die Stimmbezirksgrenzen identisch mit denen

der Wahl 1989 waren, zur Verfügung, um bei der Analyse, wenn nötig, auch eine Variable "politische Prägung" eines Stimmbezirkes berücksichtigen zu können.

So berücksichtigt die nachfolgende Analyse neben dem Anteil der Selbständigen in den 620 Dortmunder Stimmbezirken zusätzlich von 3 durch Korrelation selektierte Variablen und die Stimmenanteile von SPD, CDU, GRÜNEN, F.D.P., sowie der Nichtwähler bei der Kommunalwahl 1984.

Aus der Gruppe dieser 36 Variablen wurden die Vorhersagevariablen für das endgültige Regressionsmodell gewählt.

Dies geschah schrittweise jeweils ausgehend von den oben angesprochenen Modellen durch Analyse der Residuen, d.h. der Abweichungen zwischen tatsächlichen und prognostizierten Stimmenanteilen, der Modelle der jeweils vorangegangenen Stufe.

5.3.3 Der Aussagewert der Residuen

Unter den *Residuen* der Analyse versteht man die Abweichungen zwischen den tatsächlichen und prognostizierten Werten, in unserem Fall also des CDU-Stimmenaufkommens in jedem Stimmbezirk, also im Stimmbezirk i den Wert $c_i - \hat{c}_i$, auf jeder Stufe des Modellbildungsprozesses.

Diese Residuen können interpretiert werden, als der Anteil am CDU-Stimmenanteil, den das jeweilige Modell nicht erklären kann. Somit enthält die Menge der Residuen der Stimmbezirke alle Aussagen über Fehlspezifikationen des zugrundegelegten Modells.

Wenn das Modell korrekt wäre, würde man die Residuen als Beobachtungsfehler auffassen. Es ist daher nützlich zwischen verschiedenen Ursachen zu unterscheiden, die in ihrer Summe den Störterm ϵ_i der Regression ausmachen:

1. Meßfehler (z.B. durch zu große Meßskalen),
2. Fehler bei der Datenerhebung,
3. Stichprobenverzerrungen
4. der Einfluß der im Modell nicht spezifizierten Variablen,
5. falsche Form des spezifizierten Modells (z.B. aufgrund tatsächlichen quadratischen Beziehungen im Gegensatz zur linearen Spezifikationen),
6. durch (1) bis (5) nicht erklärte Fehler, also echte Zufallskomponenten.

Da die vorliegenden Daten im Rahmen einer Volkszählung erhoben worden sind, ist es nicht sehr wahrscheinlich, daß die drei erstgenannten Fehlerquellen hohe Fehler produzieren¹⁵. Die an vierter Stelle genannte Ursache ist schon weit eher wahrscheinlich und natürlich auch operativ besser zu berücksichtigen. Diese Fehlerquelle kann reduziert werden, indem man den Selbständigenanteil durch andere Variable ersetzt oder durch andere Variablen ergänzt, von denen man aus anderen Quellen weiß, daß diese einen Einfluß auf das Wahlergebnis haben.

Dies könnten z.B. Variablen der Bildung, der Einstellung zum familiären Leben, aber auch Variablen der Popularität des lokalen Kandidaten oder der in einem Stimmbezirk manifestierten lokalen politischen Tradition sein.

Der fünfte Fehler kann auftreten, wenn der Stimmenanteil z.B. nicht linear von Vorhersagevariablen abhängig ist, sondern quadratisch. Diesem Problem kann durch Transformation der Stimmenanteile oder auch der Vorhersagevariablen begegnet werden.

Ebenso fällt unter diesen Punkt die Nichtberücksichtigung von Interaktionen zwischen verschiedenen Vorhersagevariablen.

¹⁵Der Autor verläßt sich bei dieser Behauptung natürlich auf die Sorgsamkeit bei der Datenerhebung durch das Statistische Amt der Stadt Dortmund bei der Datenerhebung und des Statistischen Landesamtes in Düsseldorf bei der Datenaufbereitung.

5.4 Modellbildung mittels Residuen

5.4.1 Sukzessive Definition der Einflußvariablen

1. Schritt:

Aus der Identifikation der Residuen mit dem Fehlerterm des zugrundeliegenden theoretischen Regressionsmodells folgt, daß man die statistischen Eigenschaften des Fehlerterms bei den Residuen nachprüfen kann. Diese sind: Normalverteilung (für Tests und Konfidenzintervalle), sowie ein Mittelwert von Null und eine konstante Varianz.

Erfüllt die Verteilung der Residuen diese Bedingungen nicht, ist die Modellspezifikation falsch.¹⁶

Es gibt eine Vielzahl von graphischen und anderen statistischen Techniken, um die genannten Eigenschaften der Residuen nachzuprüfen.¹⁷ In dieser Studie sollten auf der Stufe des Übergangs von einer einfachen zu einer multiplen Regression des CDU-Stimmenanteils auf den Selbständigenanteil weitere Vorhersagevariable aus der Menge der 36 bereits selektierten Variablen aus der Volkszählung 1987 und der Kommunalwahl 1984 ausgewählt werden. Zu diesem Zweck wurden die Residuen gegen diese Variablen und gegen die durch das Modell vorhergesagten CDU-Stimmenanteile geplottet und zusätzlich die Korrelationen zwischen den Residuen und diesen Variablen berechnet.

Es ergaben sich bei den folgenden Variablen Werte von $|r| \geq 0.29$:

¹⁶Sind umgekehrt diese Voraussetzungen erfüllt, stellt dies allein allerdings noch keine Modellverifikation dar. Dies sei hier ausdrücklich betont!

¹⁷Siehe z.B. *Draper, N.R. und Smith, H. (1981)*, Seite 141-183

Variable	Korrelation zu den Residuen der einfachen Regression
WAHLSC84	0,33
CDU84	0,67
FDP84	0,29
NICHTW84	-0,48
UNI	0,29
ARBEITER	-0,48
ARBEITSL	-0,42
LED_MAEN	-0,34
SEKTOR2	-0,29

Tabelle 3: Korrelationen verschiedener Prädikatoren zu den Residuen der einfachen Regression

Die Residuen der einfachen Regression des CDU-Stimmenanteils auf den Selbständigenanteil sind also deutlich erkennbar nicht unabhängig von der politischen Struktur in einem Wahlkreis, wie sie vor der Wahl bestand. Oder anders ausgedrückt: der bisherige einfache Modellansatz berücksichtigt diese Struktur keineswegs hinreichend.

Andererseits ergeben sich lineare Abhängigkeiten der Residuen zu den Variablen UNI, dem Anteil der Universitätsabsolventen an der wahlberechtigten Bevölkerung, ARBEITER, dem Arbeiteranteil, ARBEITSL, dem Arbeitslosenanteil, LED_MAEN, dem Anteil der ledigen Männer und SEKTOR2, dem Anteil der Wahlberechtigten, die im produzierenden Sektor beschäftigt sind. Nun seien die Korrelationen zwischen diesen 5 Variablen betrachtet:

	ARBEITER	ARBEITSL	LED_MAEN	SEKTOR2
UNI	-0,64	-0,38	0,11	-0,30
ARBEITER		0,51	0,22	0,77
ARBEITSL			0,50	0,21
LED_MAEN				0,11

Tabelle 4: Korrelationen wichtiger Variablen der Berufs- und Familienstruktur untereinander

Man erkennt, daß der Arbeiteranteil, der Anteil der im produzierenden Sektor Beschäftigten und auch der Anteil der Wahlberechtigten mit Universitätsabschluß hoch untereinander korreliert sind. Diese Variablen beschreiben wahrscheinlich einen gemeinsamen Faktor, was die Erklärung des Wahlergebnisses betrifft.

Der Anteil der ledigen Männer wie auch der Anteil der Arbeitslosen scheint in dieser Hinsicht eine selbständige Variable zu sein. Diese beiden Variablen sind zwar auf der Aggregatebene Stimmbezirk untereinander korreliert, und der Arbeitslosenanteil ist mit dem Arbeiteranteil korreliert. Es ist aber anzunehmen, daß dies einfach daran liegt, daß mehr Männer im Erwerbsleben stehen und daran, daß Arbeiter stärker von der Arbeitslosigkeit betroffen sind. Also Gründe, die nichts mit der Wahlpräferenz gemeinsam haben.

Der Arbeiteranteil wurde aus der oben genannten Gruppe zur weiteren Analyse ausgewählt. Zum einen weist er die höchste Korrelation mit den Residuen auf, zum anderen wird er auch in vielen anderen Analysen zur Erklärung benutzt (z.B. bei *Crewe/Payne(1976)*, *Heere(1976)*).

Aus diesen Gründen sollen die Variablen aus der Vorwahl, WAHLSC84, SPD84, CDU84, GRUENE84, FDP84, sowie NICHTW84 und die Variablen ARBEITER, ARBEITSL und LED_MAEN in der weiteren Analyse zur Verbesserung des Modellansatzes der einfachen Regression des CDU-Stimmenanteils auf den Arbeiteranteil benutzt werden.

Besonders hohe Werte der Residuen ϵ_i , die durch jeden der o.a. genannten 6 Gründe verursacht sein können, werden als "Ausreißer" der Regression bezeichnet. In der vorliegenden einfachen Regression sind:

- 45 Residuen betragsmäßig größer als 8, (7,3 % der Beobachtungen)
- 30 Residuen betragsmäßig größer als 9, (4,8 %)
- 21 Residuen betragsmäßig größer als 10, (3,4 %)
- 13 Residuen betragsmäßig größer als 11, (2,1 %)
- 9 Residuen betragsmäßig größer als 12, (1,5 %)
- 3 Residuen betragsmäßig größer als 15. (0,5 %)

Diese Ausreißer sollten einer genaueren Analyse unterzogen werden. Sind klar substantielle Gründe für die hohen Residuen zu erkennen, so kann dies

eine wertvolle Hilfe sein, das bisherige Modell zu verbessern. Lassen sich jedoch keine Gründe für sehr hohe Ausreißer erkennen, so kann es von Vorteil sein und ist auch nicht ungewöhnlich, diese Stimmbezirke aus der Analyse auszuschließen¹⁸, da diese einen unproportional hohen Einfluß auf die Parameterschätzungen haben.

Bei einer absoluten Abweichung von 9 Prozentpunkten (PP) nimmt die Anzahl der Stimmbezirke mit höheren Residuen deutlich ab. Die nachfolgende Tabellen listen diejenigen 10 Stimmbezirke auf, bei denen der CDU-Stimmenanteil um über Prozentpunkten überschätzt wurde, bzw. diejenigen 20 Stimmbezirke, bei denen der CDU-Stimmenanteil um mehr als 9% unterschätzt wurde. Dies sind also Stimmbezirke, in denen die Regression auf den Selbständigenanteil den CDU-Stimmenanteil besonders schlecht erklärt.

Um diese Tabellen besser interpretieren zu können, sind die am höchsten mit den Residuen korrelierenden Variablen, nämlich der Anteil der CDU-Stimmen an den im Wählerverzeichnis am Wahltag eingetragenen Personen¹⁹, der Anteil der Nichtwähler, sowie der Arbeiteranteil in der Tabelle hinzugefügt.

¹⁸Siehe *Anscombe, F.J. (1960)*.

¹⁹Also der Anteil der CDU-Stimmen an der Wahlberechtigten Bevölkerung abzüglich derjenigen Wähler, die einen Wahlschein beantragt haben und somit aus der Wählerverzeichnis gestrichen wurden.

Stimmbezirk	Wahllokal (Statist. Bezirk)	Stimmenanteil überschätzt um	Anteil CDU ^a KW '89	Nichtwähler KW '89	Arbeiter- anteil
455	Dietrich-K.-Haus (Nordmarkt)	14,6 PP	8,9 %	58,5 %	18,1 %
1306	Jakobstuben (Lindenhorst)	11,1 PP	6,3 %	62,6 %	23,1 %
2302	Zum Postillon (Scharnhorst)	10,6 PP	4,6 %	46,9 %	34,6 %
8302	Städt. Betriebsh. (Deusen)	10,6 PP	10,3 %	57,0 %	24,2 %
5413	Haus Hunke (Syburg)	10,6 PP	21,7 %	31,5 %	9,3 %
5205	Keglerstuben (Hörde)	10,3 PP	6,6 %	56,3 %	26,8 %
3	Fritz-Henzler-Haus (City)	9,6 PP	12,4 %	60,8 %	13,1 %
9107	Haus Block (Oestrich)	9,5 PP	7,4 %	41,9 %	16,9 %
7206	Hellweg Schänke (Lütgendortmund)	9,4 PP	8,5 %	59,5 %	29,6 %
5103	Zum Phönix (Hörde)	9,1 PP	6,7 %	58,2 %	19,2 %

Tabelle 5: Stimmbezirke, in denen der CDU-Stimmenanteil bei der einfachen Regression um über 9 Prozentpunkte überschätzt wurde

^aZum Vergleich: Im gesamten Stadtgebiet wurde die CDU von 14,3 % der Wahlberechtigten gewählt. 41,7 % der Wahlberechtigten gaben ihre Stimme nicht ab.

Stimmbezirk	Wahllokal (Statist. Bezirk)	Stimmenanteil unterschätzt um	Anteil CDU KW '89	Nichtwähler KW '89	Arbeiter- anteil
6313	Augustinum (Lücklemberg)	21,5 PP	40,6 %	30,5 %	3,1 %
2204	Haus Buchbinder (Kurl-Husen)	18,8 PP	32,5 %	33,4 %	10,8 %
2112	Brockhaus (Lanstrop)	16,2 PP	34,0 %	23,8 %	12,8 %
2113	Alte Post (Lanstrop)	14,7 PP	29,8 %	36,9 %	18,2 %
9117	Altenheim Burgring (Mengede)	14,0 PP	23,0 %	35,3 %	6,9 %
115	Zum Volmerst. Platz (Westfalenhalle)	13,5 PP	23,5 %	33,6 %	8,6 %
6311	Kirchhörder GSCH (Kirchhörde)	12,2 PP	33,7 %	26,9 %	4,4 %
902	Landgrafen GSCH (Ruhrallee)	12,2 PP	26,8 %	33,1 %	6,4 %
6301	Olpketal GSCH (Lücklemberg)	11,7 PP	35,9 %	25,9 %	4,1 %
6191	Fritz-Heuner-Heim (Barop)	11,6 PP	23,0 %	19,3 %	2,0 %
6310	Olpketal GSCH (Kirchhörde) ^a	11,6 PP	30,9 %	31,9 %	4,0 %
703	Stadtgymnasium (Kaiserbrunnen)	10,8 PP	24,8 %	37,8 %	6,1 %
819	Landgrafen GSCH (Ruhrallee)	10,7 PP	28,7 %	36,6 %	5,3 %
2205	Zur Mühle (Kurl-Husen)	10,7 PP	26,4 %	33,9 %	11,8 %
4312	Busenberg GSCH (Berghofen)	10,6 PP	25,6 %	24,8 %	5,0 %
7315	Alt Marten (Marten)	9,8 PP	22,7 %	34,5 %	14,0 %
5204	KGV zum G.Entenpoth (Hörde)	9,4 PP	20,2 %	34,8 %	15,5 %
6114	Ostenberg GSCH (Barop)	9,3 PP	25,9 %	27,1 %	6,2 %
6308	Minna-Sattler-Zentr. (Lücklemberg)	9,3 PP	21,0 %	26,0 %	7,8 %
915	Zum Landgrafen (Ruhrallee)	9,0 PP	22,6 %	36,4 %	6,7 %

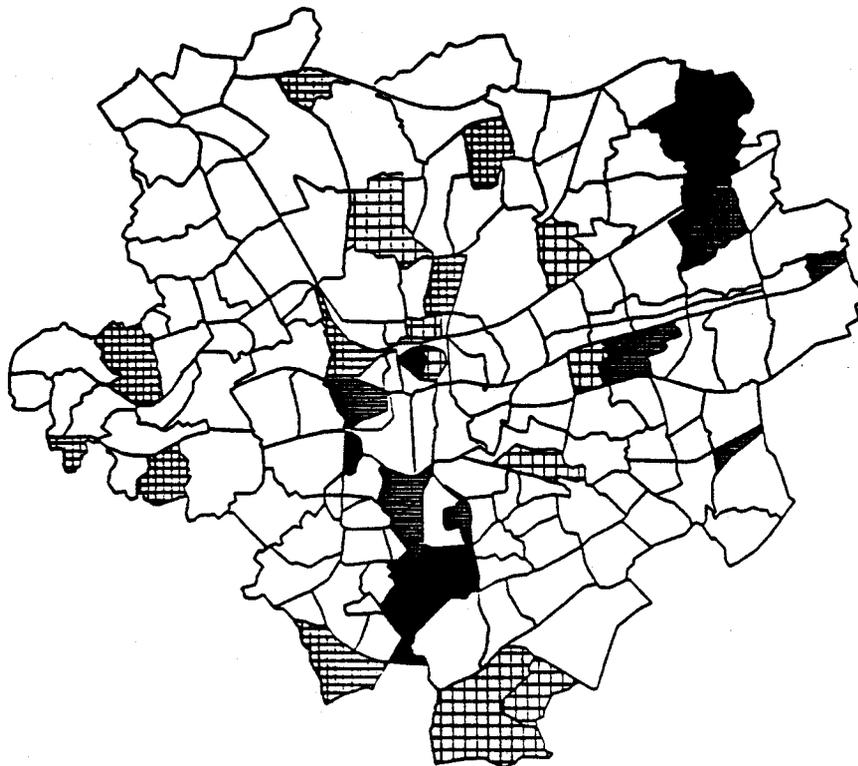
Tabelle 6: Stimmbezirke, in denen der CDU-Stimmenanteil bei der einfachen Regression um über 9 Prozentpunkte unterschätzt wurde

^aDieser Stimmbezirk hat sein Wahllokal zwar im gleichen Schulgebäude wie 6301, die "Bevölkerungsschwerpunkte" der zugehörigen Stimmbezirke liegen aber in Lücklemberg (6301) bzw. in Kirchhörde (6310).

Diese Auflistungen erweisen sich als besonders aufschlußreich. Gerade diejenigen Stimmbezirke, in denen das bisherige einfache Regressionsmodell versagt, beleuchten dessen Defizite.

Um das Modell durch die Analyse dieser Modellinsuffizienzen zu verbessern, ist es nützlich, regionale Schwerpunkte von Verschätzungen aufzuzeigen. Eine Hilfe dazu bietet die Abbildung 4, die diejenigen Statistischen Unterbezirke mit den höchsten und diejenigen mit den niedrigsten Residuen zeigt.

**EXTREME RESIDUEN DER EINF. REGRESSION
IN DEN STATISTISCHEN UNTERBEZIRKEN
MITTELWERTE DER RESIDUEN AUF STIMMBEZIRKSEBENE**



CDU-VERSCHAETZT:  SEHR HOCH UEBER  HOCH UEBERSCH.  HOCH UNTERSCH.
 SEHR HOCH UNTER  MITTEL O. K.A. 1)

1) K.A. BEDEUTET, DASS IN DIESEM GEBIET KEIN
BEVOELKERUNGSSCHWERPUNKT EINES STIMMBEZIRKES LIEGT

QUELLE: STADT DORTMUND, AMT FUER STATISTIK UND WAHLEN

Abbildung 4: Darstellung der extremen Residuen bei der einfachen Regression

Ein erster Bereich, in dem das Modell offensichtlich versagt, bilden im Stadtbezirk 6 (Hombruch) die Unterbezirke Brünninghausen, Kirchhörde und Lücklemberg.

So tauchen in Tabelle 6 die Stimmbezirke 6313 und 6311, 6301 und 6310, sowie 6308 aus diesem Schwerpunkt auf. Weiterhin weisen höhere Überschätzungen auf: 6314 (8,8 PP), 6309 (7,9 PP), 6315 (6,9 PP), 6302 (5,7 PP) sowie 6306 (5,4 PP) im gleichen Bereich. Dem Unterbezirk Bittermark sind zwei Stimmbezirke zugeordnet, in denen der CDU-Stimmenanteil durch die einfache Regression um 6,6 PP unterschätzt wird (6304) bzw. um 8,7 PP überschätzt wird (6305). Niedrigere Residuen weisen nur auf die Stimmbezirke 6303 (Lücklemberg), 4,2 PP, 6307 (Brünninghausen), 2,7 PP und 6312 (Kirchhörde) mit 3,8 PP.

Betrachtet man die Verteilung der in der Analyse verbliebenen Variablen, so fällt auf, daß in dem Bereich Brünninghausen/Kirchhörde/Lücklemberg sehr niedrige Arbeiteranteile zu finden sind. In den Stimmbezirken 6301, 6309, 6310, 6311, 6313 und 6314 sind weniger als 5 % der wahlberechtigten Bevölkerung Arbeiter, in den Stimmbezirken 6302, 6305 und 6308 liegt der Arbeiteranteil zwischen 7 % und 8 %. Im Mittel aller Stimmbezirke beträgt der Arbeiteranteil 15,7 % der wahlberechtigten Bevölkerung, liegt also deutlich höher als in den angesprochenen Gebieten. Der höchste Arbeiteranteil in diesem Gebiet ist im Stimmbezirk 6304 mit 15,1 % der Wahlberechtigten zu finden. Diese Tatsache unterstreicht die Notwendigkeit, den Arbeiteranteil als erklärende Variable in das multiple Regressionsmodell mit einzubeziehen. Dies geht auch konform mit dem Ansatz von *Crewe und Payne (1976)*, die ein ähnliches Modell entwickeln und als Hauptvorhersagevariable den Arbeiteranteil wählen.

Die Stimmbezirke 6303, 6308, 6313 und 6314 enthalten zudem Altersheime. Die Stimmbezirke 6191, sowie 9117, in denen der CDU-Stimmenanteil ebenfalls deutlich unterschätzt wurde, sind reine Altersheime. Hier tritt die Aussagekraft der Berufsstruktur (ausgedrückt durch den Selbständigenanteil) deutlich in den Hintergrund. Diese Tatsache wurde in der bisherigen Analyse noch gar nicht berücksichtigt.

Unter den 618 Stimmbezirken, die in die Analyse einfließen, befinden sich 25, die Altersheime enthalten. Darunter sind 8 Stimmbezirke, in denen alle Wahlberechtigten im Altersheim leben. Tabellen 7 und 8 listen alle 25 Stimmbezirke mit Altersheimen zusammen mit den entsprechenden Residuen, CDU- und Nichtwähleranteil bei der KW 1989, sowie den Anteilen von Selbständigen und Arbeitern in diesen Bezirken auf.

Stimmbez.	Wahllokal (Statist. Bezirk)	Reines Altenh.	Resid.	Anteil CDU KW '89	Nichtw. KW '89	Arbeiter- anteil	Selbst.- anteil
4	Käthe-Kollw.-G. (City)	nein	3,5 PP	16,6 %	44,3 %	7,8 %	2,3 %
8	Verwaltungssch. (City)	nein	4,4 PP	18,4 %	43,7 %	12,4 %	2,8 %
9	Mütterber.-st. (City)	nein	8,5 PP	23,4 %	48,0 %	6,7 %	3,3 %
190	Theod.-Flied.-H. (Westfalahalle)	ja	5,3 PP	14,3 %	50,2 %	1,6 %	0,0 %
406	Hauptsch.-Lütz. (Hafen)	nein	- 3,3 PP	8,1 %	57,4 %	22,1 %	2,4 %
505	Burgholz-Sch. (Nordmarkt)	nein	- 5,5 PP	7,4 %	56,4 %	16,4 %	2,2 %
590	Altenh. Burgh. (Nordmarkt)	ja	.a	12,1 %	42,9 %		
754	Seniorenwohnheim (Kaiserbrunnen)	nein	2,3 PP	15,6 %	49,0 %	10,7 %	2,4 %
801	Berswordt-GSCH. (Kaiserbrunnen)	nein	6,4 PP	19,1 %	37,5 %	8,9 %	2,1 %
1190	Seniorenwohnst. (Eving)	ja	- 2,9 PP	7,7 %	22,1 %	0,9 %	0,9 %
2107	Michaelsklause (Derne)	nein	- 8,0 PP	5,5 %	61,0 %	20,4 %	3,7 %
4107	Haus Hans (Aplerbeck)	nein	5,3 PP	17,8 %	39,5 %	7,1 %	2,0 %
4290	Altenh. Sölde (Sölde)	ja	.b	21,5 %	21,5 %		
4302	Zum armen Dorfsw. (Aplerbeck)	nein	0,5 PP	15,8 %	41,6 %	9,5 %	3,6 %
6191	Fritz-Heuner-H. (Barop)	ja	11,6 PP	23,0 %	19,3 %	2,0 %	1,3 %

Tabelle 7: Stimmbezirke, die Altersheime enthalten (1. Teil)

^aHier sind keine Personendaten verfügbar, da das gesamte Altenheim weniger als eine Blockseite darstellt.

^bDas Altenheim Sölde besteht in Bezug auf die Volkszählung aus weniger als einer Blockseite. Daher lassen sich hier keine Selbständigen bzw. Arbeiteranteile zuordnen und mithin fällt dieser Stimmbezirk aus der Analyse.

Stimmbez.	Wahllokal (Statist. Bezirk)	Reines Altenh.	Resid.	Anteil CDU KW '89	Nichtw. KW '89	Arbeiter- anteil	Selbst- anteil
6192	Städt. Altenh. (Eichlinghofen)	ja	3,7 PP	13,6 %	44,0 %	0,5 %	0,5 %
6303	Forsth. Lückl. (Lücklemberg)	nein	4,2 PP	25,7 %	34,1 %	5,1 %	7,1 %
6308	Minna-S.-Zentrum (Brünninghaus.)	nein	9,3 PP	21,0 %	26,0 %	7,8 %	1,5 %
6313	Augustinum (Lücklemberg)	nein	21,5 PP	40,6 %	30,5 %	3,1 %	5,7 %
6314	Altenh. Kirchh. (Kirchhörde)	nein	8,8 PP	29,8 %	30,3 %	4,8 %	6,8 %
6391	Herman.-K.-Haus (Lücklemberg)	ja	4,0 PP	12,9 %	29,2 %	0,4 %	0,0 %
6392	Erna-David-Zentr. (Lücklemberg)	ja	1,9 PP	10,8 %	40,6 %	2,7 %	0,0 %
7209	Alte Ziegelei (Lütgendortmu.)	nein	- 2,5 PP	12,2 %	55,0 %	12,3 %	3,2 %
8304	Huckarder Krone (Huckarde)	nein	5,5 PP	17,2 %	49,1 %	12,1 %	1,5 %
9117	Altenheim Burgr. (Mengede)	nein	14,0 PP	23,0 %	35,3 %	6,9 %	0,0 %

Tabelle 8: Stimmbezirke, die Altersheime enthalten (2. Teil)

Gerade in denjenigen Bezirken, die nur Wahlberechtigte aus Altersheimen beinhalten, liegt der Anteil der Selbständigen nahezu bei Null, was nicht weiter überrascht. Man könnte vermuten, daß die Vorhersagen der einfachen Regression deshalb besonders schlecht werden würden. Das ist aber nicht der Fall. Nur 12 dieser 25 weisen Verschätzungen des CDU-Stimmenanteils um mehr als 5 PP auf, nur 7 Verschätzungen um über 8 PP. Diese Tatsache zeigt, daß das vorliegende Modell den Stimmenanteil der CDU innerhalb der Nicht-Selbstständigen Bevölkerung schon recht gut durch den Absolutwert der Regression, nämlich den Wert a_0 geschätzt wird²⁰.

Die bei weitem größte Verschätzung des CDU-Stimmenanteils um 21,5 PP ergab sich im Bezirk 6313, wo die CDU ihren höchsten Anteil an den Wahlberechtigten mit 40,6 % erreichte. In diesem Stimmbezirk erreichte auch die F.D.P. mit 9,4 % ein stadtweit sehr gutes Ergebnis. Abgesehen davon gibt es keine weitere Variable der Sozialstruktur aus der Volkszählung, die hier einen Extremwert aufweist. Lediglich der CDU-Stimmenanteil bei der Vorwahl (43,7 %) fällt ähnlich aus dem Rahmen. Dies unterstreicht eindringlich die Notwendigkeit, eine Variable, die den politischen Kontext in einem Stimmbezirk beschreibt, mit in die Analyse einfließen zu lassen.

Vier von 11 Stimmbezirken aus dem Bereich Lanstrop/Kurl/Husen weisen Unterschätzungen des CDU-Stimmenanteils um mehr als 9 PP auf. Es gibt keinen weiteren Stimmbezirk aus diesem Bereich, der geschätzte Werte aufweist, die um mehr als 5 PP von den tatsächlichen CDU-Stimmenanteilen abweichen.

Die Betrachtung der anderen Variablen in diesen Stimmbezirken, ebenso wie im Stimmbezirk 115 (Wahllokal: Zum Volmarsteiner Platz) und 819/902 (Wahllokal: Landgrafengrundschule), sowie 703 (Wahllokal: Stadtgymnasium) bietet keinen Hinweis auf die Gründe für die Unterschätzung. Für diese Stimmbezirke erscheint es daher besonders interessant, zu beobachten, wie sich hier die Residuen des multivariaten Modells darbieten.

Soweit die Betrachtung derjenigen Stimmbezirke, in denen der CDU-Stimmenanteil stark unterschätzt wurde.

Betrachtet man die aufgelisteten Stimmbezirke, in denen der CDU-Stimmenanteil stark überschätzt wurde, so fällt auf, daß sich diese Bezirke ebenfalls nicht in einfacher Weise durch andere Variable charakterisieren lassen.

²⁰Diese Qualitätsaussage trifft mindestens auf den Bereich der Senioren in Altersheimen zu.

2. Schritt:

Nach der Analyse der Residuen des einfachen Regressionsmodells kann nun ein multipler Ansatz gemäß den Annahmen (b)(i) oder auch (b)(ii) aus Kapitel 5.3.2 spezifiziert werden.

Dazu erscheint es aufgrund der Residualanalyse notwendig, zusätzlich die Variablen UNI, ARBEITER und LED_MAEN zu verwenden und die politische Ausgangslage in einem Stimmbezirk in geeigneter Weise durch Variablen, die die Stimmenanteile bei der Vorwahl (KW 1984) beschreiben, zu spezifizieren.

Die obige Residualanalyse hat aufgezeigt, daß CDU-Stimmenanteile dort *unterschätzt wurden, wo:*

1. der Anteil der Universitätsabsolventen sehr hoch liegt oder
2. der Arbeiteranteil sehr niedrig ist oder
3. es deutlich weniger ledige Männer gibt und vor allem dort, wo
4. die CDU schon 1984 eine sehr überdurchschnittliche Stellung hatte.

Überschätzungen (zumeist geringeren Ausmaßes) ergaben sich dementgegen dort, wo:

1. der Anteil der Universitätsabsolventen sehr niedrig ist oder
2. der Arbeiteranteil sehr hoch ist oder
3. der Anteil lediger Männer deutlich höher liegt und vor allem dort, wo
4. die CDU auch schon bei der Vorwahl 1984 unterdurchschnittlich schlecht abgeschnitten hat und auch der Anteil der Nichtwähler sehr hoch lag.

Zudem gibt es Assoziationen zwischen bestimmten geographischen Bereichen und höheren Residuen. So treten, wie bereits dargelegt, in den Bereichen Brünninghausen/Kirchhörde/Lücklemburg und Kurl/Husen Unterschätzungen größeren Ausmaßes auf.

Da nun Indikatoren zur Verbesserung des Regressionsansatzes aufgezeigt worden sind, kann das Regressionsmodell erweitert und verbessert werden. Die

theoretischen Grundlagen dazu wurden in Kapitel 5.3.2 eruiert. Bei der Modellentwicklung wurde nun wie folgt verfahren:

Da auf der bisherigen Stufe der Modellbildung keine Gründe offensichtlich darauf hindeuten, daß der CDU-Stimmenanteil innerhalb der Gruppe der Selbständigen anders durch bestimmte Vorhersagevariablen zu beschreiben ist, als innerhalb der Gruppe der Nicht-Selbständigen, wird hier nach dem Modellansatz (b)(i) verfahren und (b)(ii) außer acht gelassen.

Die Erklärung des "erwarteten CDU-Stimmenanteils" wurde um weitere Vorhersagevariablen schrittweise erweitert. Die Güte der dadurch eingetretenen Verbesserung wurde durch den Anstieg der durch das Modell erklärten Varianz, also durch den Anstieg des Wertes R^2 (Siehe 5.1.3) gemessen. Außerdem wurde durch einen t-Test die statistische Signifikanz jedes neu eingeführten Parameters getestet²¹.

In dieser Weise wurde fortgefahren mit dem Ziel, im endgültigen Modell die folgenden Kriterien hinreichend zu erfüllen:

- (a) "sparsame" Modellspezifikation (d.h. möglichst wenige Vorhersagevariablen),
- (b) ein möglichst hoher Wert R^2 an durch das Modell erklärter Varianz,
- (c) signifikant von 0 verschiedene Regressionskoeffizienten zu den einzelnen Vorhersagevariablen,
- (d) Erfüllung der durch das Modell vorausgesetzten statistischen Kriterien, nachgeprüft durch Tests auf der Basis der Residuen.

Zudem sollen die Vorhersagevariablen einen substantiellen Wert haben, also leicht interpretierbar sein und der Effekt von globalen Variablen soll im Modell vollständig berücksichtigt sein, damit die Residuen dieses "vollständigen" Modells zur Beurteilung des Wahlerfolges der einzelnen Parteien dienen können.

Betragsmäßig größere Residuen verdienen in jedem Fall eine genauere Nachforschung der "betroffenen" Parteien.

Diese Strategie kann wie folgt interpretiert werden: Jedem Wähler wird eine bestimmte *objektive Wahrscheinlichkeit* zugeordnet, mit der dieser CDU wählt. Diese Wahrscheinlichkeit verändert sich abhängig davon, ob der Wähler selbständig oder nicht selbständig ist, und nimmt zu oder ab abhängig von bestimmten Attributen des Stimmbezirks, in dem dieser Wähler seine

²¹Siehe *Draper/Smith(1981)*, Seite 104-107

Stimme abgibt. Die Variablen zur Bestimmung dieser Attribute wurden bereits selektiert. Nun wurden diese Variablen, um die Interpretation erheblich zu vereinfachen, auf nominale Skalen reduziert. Dies geschah, obwohl sich die Variablen auch intervallskaliert, d.h. in ihrem genauen Wert in einem Stimmbezirk, ausdrücken ließen.

Diese Vereinfachung geschah, da man nicht einfach annehmen kann, daß der Effekt der Vorhersagevariablen notwendigerweise linear ist, der Wert der entsprechenden Regressionskoeffizienten also der gleiche ist für alle vorkommenden Werte von Vorhersagevariablen. Dieser Schritt hin zu einer nominalen Skalierung vereinfacht auch die Interpretation, erlaubt er doch die Klassifizierung der Stimmbezirke nach Zugehörigkeit zu einer bestimmten Kategorie (hier einer Kombination von Vorhersagevariablen).

Ebenso wird ohne weitere Begründung davon ausgegangen, daß die Effekte der Vorhersagevariablen additiv sind. Interaktionseffekte zwischen verschiedenen Vorhersagevariablen wurden im Rahmen dieser Untersuchung auch nicht berücksichtigt²².

Somit sind im geschilderten Ansatz einige Vereinfachungen enthalten, insbesondere in der Definition der Vorhersagevariablen und in der Beschränkung auf eine streng lineare Form des Modells. Etwas Genauigkeit mußte geopfert werden für die vereinfachte Interpretation, die das Modell dadurch erlaubt.

Daher flossen auch die hohen Residuen in den Bereichen Brünninghausen/ Kirchhörde/Lücklemburg und Kurl/Husen nicht in die Analyse ein. Man könnte regionale Erklärungsfaktoren in die Modellspezifikation einfließen lassen. Dies geschah aber nicht. Zum einen sollte das gewünschte Modell nämlich möglichst schlicht formuliert und damit leicht erklärbar sein, d.h. vor allem möglichst wenige Erklärungsfaktoren enthalten. Die Berücksichtigung von regionalen Besonderheiten hätte aber im größten Fall die Einbeziehung von mindestens 12 weiteren Erklärungsvariablen (für jeden Stadtbezirk eine) bedeutet. Zum anderen soll das Modell auch nur solche Variablen enthalten, die den Wahlerfolg einer Partei stadtweit beeinflussen. Die Residuen des "kompletten" Modells auf regionaler Ebene (bzw. für jeden Stimmbezirk) sind dann nämlich als individueller Wahlerfolg einer Partei bzw. der Nichtwähler zu interpretieren.

²²Mit einer Ausnahme: Die Beschreibung der politischen Ausgangssituation berücksichtigt auch Interaktionen zwischen den Variablen der Vorwahl.

Nominalskalierung der Vorhersagevariablen und Klassifizierung der Stimmbezirke

Bei der Residualanalyse fanden sich die höchsten Modellabweichungen in solchen Stimmbezirken mit besonders hohen oder niedrigen Werten der wichtigen Vorhersagevariablen.

Da nur diese besonders hohen oder niedrigen Werte in die Analyse einfließen sollen, wurde wie folgt verfahren:

Für die Berücksichtigung der Vorhersagevariablen UNI, ARBEITER und LED_MAEN wurden 6 nominalskalierte Variablen PRAE1 bis PRAE6 eingeführt, die alle nur den Wert 1 oder 0 annehmen. Und zwar wurden die folgenden Werte zugeordnet:

PRAE1 = 1 , wenn der Anteil der Universitätsabsolventen in einem Stimmbezirk zu den 25 % höchsten Anteilen aller Stimmbezirke zählt, also über dem 75 %-Quantil der Verteilung von UNI, hier über 5,32 % liegt. Der Wert von UNI, also überdurchschnittlich *hoch* ist,

PRAE1 = 0 , sonst.

PRAE2 = 1 , wenn der Wert von ARBEITER überdurchschnittlich *hoch* ist, d.h. hier über 19,75 % liegt,

PRAE2 = 0 , sonst.

PRAE3 = 1 , wenn der Wert von LED_MAEN überdurchschnittlich *hoch* ist, d.h. hier über 15,17 % liegt,

PRAE3 = 0 , sonst.

PRAE4 = 1 , wenn der Wert von UNI unterdurchschnittlich *niedrig* ist, d.h. unter dem 25 % Quantil der Verteilung von UNI, hier also unter 1,38 % liegt,

PRAE4 = 0 , sonst.

PRAE5 = 1 , wenn der Wert von ARBEITER unterdurchschnittlich *niedrig* ist, d.h. hier unter 11,50 % liegt,

PRAE5 = 0 , sonst.

PRAE6 = 1 , wenn der Wert von LED_MAEN unterdurchschnittlich *hoch* ist, d.h. hier über 19,75 % liegt,

PRAE6 = 0 , sonst.

Läßt man die so definierten Variablen in einen Regressionsansatz gemäß (b)(ii) einfließen, paßt also ein Modell der Form:

$$CDU = a_0 + b_0 \cdot SE + b_1 \cdot PRAE1 + b_2 \cdot PRAE2 + b_3 \cdot PRAE3 + b_4 \cdot PRAE4 + b_5 \cdot PRAE5 + b_6 \cdot PRAE6 \quad (5)$$

an, so ergibt sich ein Bestimmtheitsmaß von $R^2 = 0,57$. Dieses Modell erklärt also 57 % der gesamten Varianz, im Gegensatz zum einfachen Regressionsmodell, daß nur 35 % der Gesamtvarianz erklärte, also eine deutliche Verbesserung.

Der Parameter b_6 ist aber in diesem Modell nicht signifikant von 0 verschieden, Stimmbezirke mit sehr niedrigen Anteilen an ledigen Männern weichen in ihren CDU-Anteilen also nicht signifikant von den anderen ab. Deshalb wurde ein Modell der Form:

$$CDU = a_0 + b_0 \cdot SE + b_1 \cdot PRAE1 + b_2 \cdot PRAE2 + b_3 \cdot PRAE3 + b_4 \cdot PRAE4 + b_5 \cdot PRAE5 \quad (6)$$

angepaßt. Dies erklärt 56,8 % der Gesamtvarianz und die Parameterschätzungen sind die folgenden:

$$\begin{aligned} a_0 &= 11,58 & b_0 &= 1,05 & b_1 &= 1,37 & b_2 &= -1,27 \\ b_3 &= -3,17 & b_4 &= -2,08 & b_5 &= 3,26 \end{aligned}$$

Um den Anteil der erklärten Varianz noch weiter zu erhöhen, ist aber zudem nötig, die Vorwahl in geeigneter Weise mit in die Analyse einfließen zu lassen. Wie schon dargelegt, hat die Analyse ergeben, daß dort, wo die CDU 1984 schon schwach war, ihre Stimmenanteile durch die Regression überschätzt wurden. Dort wo sie 1984 schon stark war, wurden ihre Stimmenanteile unterschätzt.

Dies kann verschiedene Gründe haben. Zum einen ist es möglich, daß in diesen Stimmbezirken eine "CDU-unfreundliche Situation" herrscht, durch nicht-populäre Kandidaten, usw., die einen größeren Teil der Wähler dazu bringt, nicht CDU zu wählen, sondern SPD oder nicht zur Wahl zu gehen. Zum anderen aber können die potentiellen CDU Wähler auch den Wert ihrer Stimme in diesen Stimmbezirken als so niedrig beurteilen, daß sie gar nicht

erst zur Wahl antreten²³. Ebenso können sehr hohe Anteile der CDU an den Stimmen der Vorwahl, die ja in Dortmund erst eine "echte Konkurrenz" zwischen SPD und CDU darstellen, mehr potentielle Wähler zur Stimmabgabe für die CDU bewegen, da sich in solchen Stimmbezirken die Stimmabgabe dann "lohnt".

Bei der Beschreibung der politischen Struktur durch Diasporagebiete und Hochburgen, wie bei den bisher hinzugenommenen Variablen, tauchen zwei Schwierigkeiten auf.

Zum einen sind für 32 Stimmbezirke keine Vorwahlergebnisse verfügbar, da sich die Stimmbezirksgrenzen seit der letzten Wahl geändert haben. Daher muß sich die Analyse dieser Stimmbezirke darauf beschränken, die Residuen einer multiplen Regression unter Berücksichtigung des Selbständigenanteils und der genannten 5 zusätzlichen Vorhersagevariablen aufzulisten. Siehe hierzu Tabelle 9.

Nummer	Residuum	Nummer	Residuum
1105	-1,94	7115	0,27
1106	1,83	7116	-7,42
1107	0,35	7303	-1,60
1108	-3,16	7304	-2,90
1110	-4,60	7305	-4,80
1115	-2,42	7306	-10,09
1190	-6,05	7307	-2,04
3214	-1,28	7308	-5,13
3404	2,22	7317	-2,21
3408	5,23	8309	6,54
4307	2,46	8310	-2,21
4319	-2,90	8312	0,43
4404	-0,75	8313	-0,06
6105	-4,12	9116	-1,67
7110	0,16	9117	10,22
7111	3,49	9303	4,51

Tabelle 9: Stimmbezirke ohne verfügbare Vorwahlergebnisse und Residuen des für diese Bezirke möglichen Modells für die CDU

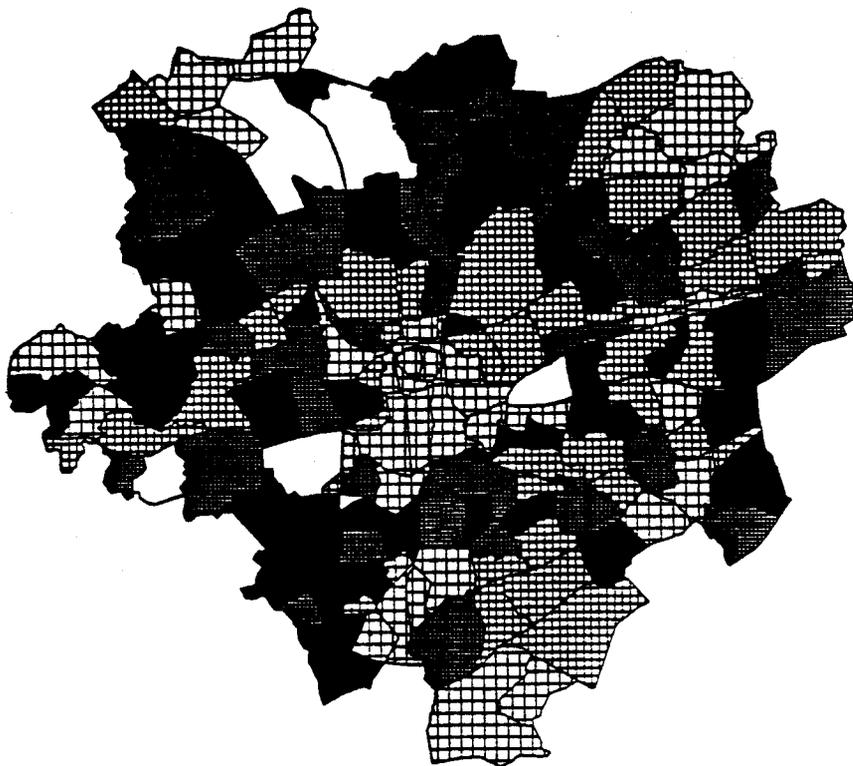
²³z.B. Selbständige in absoluten SPD-Hochburgen, die die SPD-Ratsmehrheit im Rat als unveränderbar ansehen.

Zum anderen ist die Beschreibung nicht so einfach, wie für die bisher hinzugenommenen Variablen UNI, ARBEITER und LED_MAEN.

Es erscheint sinnvoll die Anteile von SPD, CDU und Nichtwählern gemeinsam zu berücksichtigen. Die Abbildungen 5 bis 7 zeigen die entsprechenden Hochburgen und Diasporagebiete.

STELLUNG DER SPD BEI DER KW 1984 IN DEN STATISTISCHEN UNTERBEZIRKEN

SPD - WAHLERANTEIL AN DER WAHLBER. BEVOELKERUNG ABZGL. BEANTR. WAHLSCHEINE



STELLUNG SPD:  DIASPORA HOCHBURG:  UNTERDURCHSCHN.:  UEBERDURCHSCHN.: 
K.A. MOEGLICH 1): 

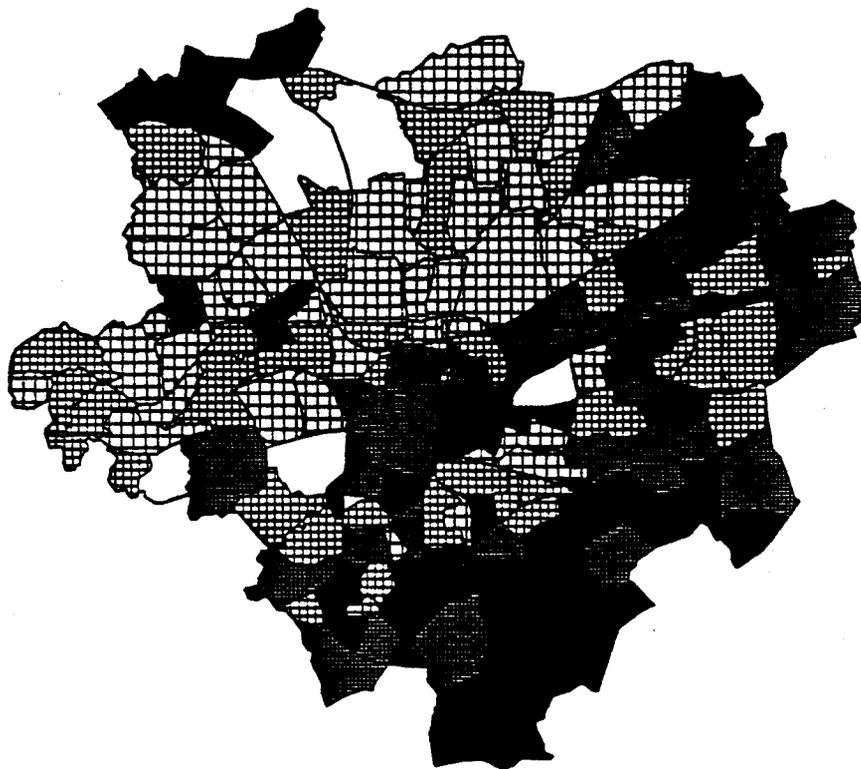
1) STIMMBEZIRKSGRENZEN VON 1984 STIMMEN HIER NICHT MIT DEN 1989ERN UEBEREIN
ODER IN DIESEM GEBIET LIEGT KEIN BEVOELKERUNGSSCHWERPUNKT EINES STIMMBEZIRKES

QUELLE: STADT DORTMUND, AMT FUER STATISTIK UND WAHLEN

Abbildung 5: Hochburgen und Diaspora der SPD bei der KW 1984

STELLUNG DER CDU BEI DER KW 1984 IN DEN STATISTISCHEN UNTERBEZIRKEN

CDU - WAHLERANTEL AN DER WAHLBER. BEVOELKERUNG ABZGL. BEANTR. WAHLSCHENE



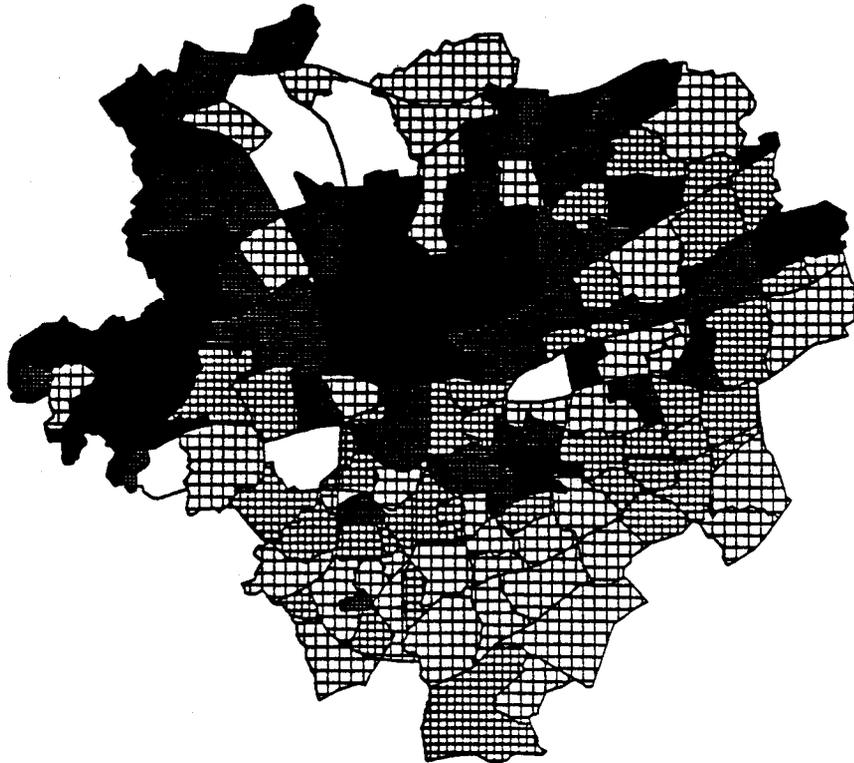
STELLUNG CDU:  UNTERDURCHSCHN.  UEBERDURCHSCHN.
 DIASPORA HOCHBURG  K.A. MOEGLICH 1)

1) STIMBEZIRKSGRENZEN VON 1984 STIMMEN HIER NICHT MIT DEN 1989ERN UEBEREIN
 ODER IN DIESEM GEBIET LIEGT KEIN BEVOELKERUNGSSCHWERPUNKT EINES STIMBEZIRKES

QUELLE: STADT DORTMUND, AMT FUER STATISTIK UND WAHLEN

Abbildung 6: Hochburgen und Diaspora der CDU bei der KW 1984

NICHTWAEHLER BEI DER KW 1984 IN DEN STATISTISCHEN UNTERBEZIRKEN ANTEIL AN DER WAHLBER. BEVOELKERUNG ABZGL. BEANTR. WAHLSCHEINE



ANTEIL:  SEHR NIEDRIG  NIEDRIG  HOCH
 SEHR HOCH  K.A. MOEGLICH 1)

1) STIMMBEZIRKSGRENZEN VON 1984 STIMMEN HIER NICHT MIT DEN 1989ERN UEBEREIN
 ODER IN DIESEM GEBIET LIEGT KEIN BEVOELKERUNGSSCHWERPUNKT EINES STIMMBEZIRKES

QUELLE: STADT DORTMUND, AMT FUER STATISTIK UND WAHLEN

Abbildung 7: Hochburgen und Diaspora der Nichtwähler bei der KW 1984

Nun ist es möglich, daß ein Stimmbezirk gleichzeitig Hochburg von SPD und CDU oder von SPD und Nichtwählern oder von CDU und Nichtwählern ist, was aus den o.a. Erwägungen als Interaktion berücksichtigt werden muß. Dies trifft gleichermaßen auf die Diasporagebiete zu. Deshalb reicht es nicht, wie bei der Behandlung der drei bisherigen Vorhersagevariablen, jeweils zwei Nominalvariablen pro Vorhersagevariable einzuführen, sondern es müssen zur Berücksichtigung der Stimmenanteile von SPD, CDU und Nichtwählern bei der Vorwahl insgesamt zunächst 12 neue nominal skalierte Variable eingeführt werden.

Diese wurden folgendermaßen definiert:

HOCH1-HOCH6 geben die Hochburgensituation an. Dabei nimmt HOCH1 den Wert 1 an, wenn der SPD-Stimmenanteil 1984 über dem 75 %-Quantil der Verteilung der Stimmenanteile der SPD bei der KW 1984 innerhalb der 586 betrachteten Stimmbezirke liegt, hier also über 38,06 % und wenn die Anteile für die CDU bzw. der Nichtwähler 1984 unter den 75 %-Quantilen der entsprechenden Verteilungen lagen. Also hier unter 31,35 % bei der CDU und unter 47,2 % bei den Nichtwählern. HOCH2-HOCH6 wurden entsprechend für die anderen Hochburgen definiert, also

HOCH1 = 1 , wenn dieser Stimmbezirk 1984 nur SPD-Hochburg war,
= 0 , wenn nicht,

HOCH2 = 1 , wenn dieser Stimmbezirk 1984 nur CDU-Hochburg war,
= 0 , wenn nicht,

HOCH3 = 1 , wenn dieser Stimmbezirk 1984 nur Nichtwähler-Hochburg war,
= 0 , wenn nicht,

HOCH4 = 1 , wenn dieser Stimmbezirk 1984 Hochburg von SPD und CDU und nicht der Nichtwähler war,
= 0 , wenn nicht,

HOCH5 = 1 , wenn dieser Stimmbezirk 1984 Hochburg von SPD und Nichtwählern war,
= 0 , wenn nicht,

HOCH6 = 1 , wenn dieser Stimmbezirk 1984 Hochburg von CDU und Nichtwählern war,
= 0 , wenn nicht.

Entsprechend wurden die Variablen DIAS1-DIAS6 für die Diasporasituation definiert.

DIAS1 = 1 , wenn dieser Stimmbezirk 1984 nur SPD-Diasporagebiet war, (SPD \leq 27,6 %, CDU \geq 12,3 %, Nichtwähler \geq 35,1 %)
= 0 , wenn nicht,

DIAS2 = 1 , wenn dieser Stimmbezirk 1984 nur CDU-Diaspora war,
= 0 , wenn nicht,

DIAS3 = 1 , wenn dieser Stimmbezirk 1984 nur Nichtwähler-Diasporagebiet war,
= 0 , wenn nicht,

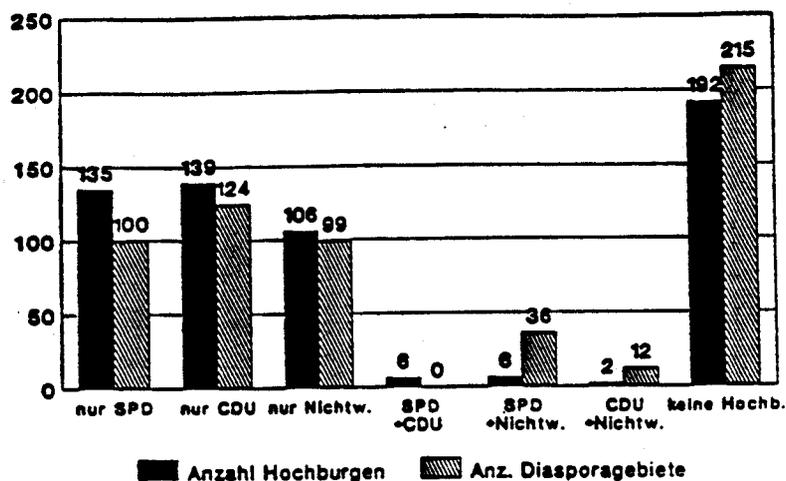
DIAS4 = 1 , wenn dieser Stimmbezirk 1984 Disporagebiet von SPD und CDU und nicht der Nichtwähler war,
= 0 , wenn nicht,

DIAS5 = 1 , wenn dieser Stimmbezirk 1984 Diaspora von SPD und Nichtwählern war,
= 0 , wenn nicht,

DIAS6 = 1 , wenn dieser Stimmbezirk 1984 Diaspora von CDU und Nichtwählern war,

Abbildung 8 zeigt dabei für die 586 Stimmbezirke, für die Daten verfügbar sind, die entsprechenden Häufigkeiten.

Hochburgen und Diaspora in Dortmund bei der Kommunalwahl 1984



Basis: 588 Direktwahlbezirke

Abbildung 8: Häufigkeitsstruktur von Hochburgen bzw. Diaspora der Parteien bei der KW 1984

DIAS4 braucht also in der Analyse nicht berücksichtigt zu werden, da es keinen Stimmbezirk gibt, wo gleichzeitig SPD und CDU in Diaspora waren.

Nun wurden die CDU-Stimmenanteile einem Modell der Form:

$$\begin{aligned}
 CDU = & a_0 + b_0 \cdot SE + b_1 \cdot PRAE1 + b_2 \cdot PRAE2 + b_3 \cdot PRAE3 \\
 & + b_4 \cdot PRAE4 + b_5 \cdot PRAE5 \\
 & + b_6 \cdot HOCH1 + b_7 \cdot HOCH2 + b_8 \cdot HOCH3 + b_9 \cdot HOCH4 \\
 & + b_{10} \cdot HOCH5 + b_{11} \cdot HOCH6 \\
 & + b_{12} \cdot DIAS1 + b_{13} \cdot DIAS2 + b_{14} \cdot DIAS3 \\
 & + b_{15} \cdot DIAS5 + b_{16} \cdot DIAS6
 \end{aligned} \quad (7)$$

angepaßt. Dies beschreibt 76,9 % der Gesamtvarianz. Allerdings sind jetzt die Parameter für b_1 , b_2 , b_4 und b_{11} nicht mehr signifikant von 0 verschieden, wurden also aus der Regression ausgeschlossen. Daß b_{11} , der Parameter für gemeinsame Hochburgen von CDU und Nichtwählern, keinen signifikanten Einfluß hat, ist dadurch zu erklären, daß er nur für zwei Stimmbezirke den Wert 1 annimmt. Die Tatsache, daß b_1 , b_2 und b_4 aus der Analyse

fallen, mag überraschen, ist aber dadurch zu erklären, daß die Vorwahlsituation zusammenfällt mit den entsprechenden Werten für sehr wenige und sehr viele Hochschulabsolventen bzw. sehr viele Arbeiter.

Das endgültige Regressionsmodell für die CDU-Stimmenanteile

So bleibt letztlich ein Modell übrig, in dem neben dem Selbständigenanteil und den Hochburgen und Diasporagebieten bei der Vorwahl nur berücksichtigt wird, ob im Stimmbezirk sehr viele ledige wahlberechtigte Männer leben oder nicht und ob dort sehr wenige wahlberechtigte Arbeiter leben oder nicht.

Dieses Modell hat die Form:

$$\begin{aligned}
 CDU = a_0 + b_0 \cdot SE + b_1 \cdot PRAE3 + b_2 \cdot PRAE5 \\
 + b_3 \cdot HOCH1 + b_4 \cdot HOCH2 + b_5 \cdot HOCH3 \\
 + b_6 \cdot HOCH4 + b_7 \cdot HOCH5 \\
 + b_8 \cdot DIAS1 + b_9 \cdot DIAS2 + b_{10} \cdot DIAS3 \\
 + b_{11} \cdot DIAS5 + b_{12} \cdot DIAS6
 \end{aligned} \tag{8}$$

und erklärt 76,6 % der Gesamtvarianz.

Betrachtet man die Residuen dieser multiplen Regression und den Einfluß bestimmter Stimmbezirke auf die Ergebnisse der Regression, so fallen die Stimmbezirke 2112, 2204, 5413, 6220, 6301 und 6313 derart aus dem Rahmen, daß sie als Ausreißer aus der Analyse ausgeschlossen wurden. Dadurch erhöht sich der Anteil der erklärten Varianz auf 77,9 % der Gesamtvarianz. Die Parameterschätzungen und die Interpretation dieses "vollständigen" Modells finden sich in Kapitel 4!

Modelle für die Stimmenanteile der anderen Parteien

Tabelle 2 zeigt die Möglichkeiten, Stimmenanteile zu schätzen. Für die CDU wurde die bisherige Analyse exemplarisch durchgeführt, da sich der Selbständigenanteil als besonders hoch mit dem CDU-Stimmenanteil erwies.

Erwartungswerte für die Stimmenanteile der SPD und der Nichtwähler sollten ebenfalls durch multiple Regressionsmodelle ausgehend von allen Vorher-

sagevariablen PRAE1-PRAE6, HOCH1-HOCH5 und DIAS1-DIAS4, sowie DIAS5 und DIAS6 geschätzt werden.

Dabei sollten wieder diejenigen Vorhersagevariablen aus der Regression ausgeschlossen werden, deren Parameter nicht signifikant von 0 verschieden waren und Stimmbezirke mit extremen Residuen sollten ebenfalls als Ausreißer aus der Regression ausgeschlossen werden.

Für den SPD-Stimmenanteil ging dies recht gut und das letzte Modell erklärt 72 % der Varianz innerhalb der SPD-Stimmenanteile bei der KW 1989. Die Ergebnisse finden sich in Kapitel 4.

Bei der Erklärung des Anteils der Nichtwähler zeigt sich jedoch, daß der Faktor, der die Vorhersagevariable Selbständige berücksichtigt, nicht signifikant von 0 verschieden ist. Also ist das entwickelte Modell untauglich zur Erklärung des Nichtwähleranteils, da Selbständige offensichtlich sehr unterschiedlich zur Wahl gehen.

Regressionen der Stimmenanteile der F.D.P. bzw. der GRÜNEN führen zu keinem befriedigenden Ergebnis. Hier werden durch Modelle der bisher verwendeten Struktur nur Anteile an erklärter Varianz unter 55 % erreicht.

Um den Stimmenanteilen dieser kleineren Parteien einen Erwartungswert zuzuordnen, bedarf es also klar einer anderen Modellstruktur, die an dieser Stelle nicht entwickelt werden soll.

5.4.2 Vorteile der Modellbildung durch Residualanalyse und Unterschiede zu anderen Modellbildungstechniken

Die Methode *„ein multiples Regressionsmodell, d.h. die Auswahl bzw. Definition der Vorhersagevariablen, durch Analyse der Residuen zu spezifizieren*, ist nicht der übliche Weg, den man innerhalb der politikwissenschaftlichen oder soziologischen Forschung normalerweise beschreitet.

In diesem Bereich ist es üblich, die Modelle auf der Basis von multiplen Korrelationskoeffizienten zu konstruieren. Die von P. Klemmer und H.F. Eckey durchgeführte Regression in der *Wahlanalyse der Stadt Hamm (1976)* ist beispielhaft für dieses Vorgehen. Hier werden aus allen möglichen Variablen aus der Volkszählung diejenigen ermittelt, die am höchsten mit dem Wahlergebnis korreliert sind. Im genannten Fall geschieht dies für 12 Wahlbezirke als Aggregationsstufe und es werden 51 Variablen selektiert. Diese 51 Variablen sind mit hoher Wahrscheinlichkeit untereinander hoch korreliert. Gerade auf dieser groben Aggregatstufe sind die Indikatoren der Bildung und der Berufsstruktur von gleichem Aussagewert für die Höhe der Stimmenan-

teile einzelner Parteien. Solche "Autokorrelationen" innerhalb der Vorhersagevariablen verletzen auch die Voraussetzungen zur Anwendung der Regressionsanalyse empfindlich. Mithin kann ein hoher Wert des Bestimmtheitsmaßes R^2 erreicht werden, das zugehörige Regressionsmodell, bzw. dessen Schätzwerte sind aber kaum mehr interpretierbar, da die Einflüsse der Vorhersagevariablen in der Modellspezifikation untereinander "vermengt" sind²⁴. Diese Vermengung ist ziemlich willkürlich und daher ist die Interpretation der gewonnenen Schätzwerte äußerst schwierig.

So sind im Modell von *Eckey und Klemmer* beispielsweise die Vorhersagevariablen, wie Wohnbevölkerung mit überwiegendem Lebensunterhalt des Ernährers im Handel, in den Wirtschaftsabteilungen Verkehr und Nachrichtenübermittlung, in der Wirtschaftsabteilung Dienstleistungen, und in Gebietskörperschaften und Sozialversicherungen enthalten. Würde man die Korrelationen dieser Variablen untereinander betrachten, so könnte man sicherlich leicht nachweisen, daß alle diese Variablen einen gemeinsamen Faktor (z.B. Beschäftigte im tertiären Sektor) beschreiben. Dieser Faktor wird jedoch von den beiden Autoren noch zusätzlich ins Modell einbezogen, was die Vermengung völlig willkürlich macht. So sind die Anteile der Erwerbstätigen im Handel, die Anzahl der Erwerbstätigen in Verkehr und Nachrichtenübermittlung, die Anzahl der Beschäftigten im Versicherungsgewerbe und die Anzahl der Beschäftigten im Dienstleistungsgewerbe, jeweils pro Wahlbezirk noch zusätzlich im Modell berücksichtigt. Die ermittelten Schätzwerte sind daher sehr schlecht als Erwartungswerte für Stimmenanteile zu interpretieren, denn eine komplett andere Modellspezifikation (einige Vorhersagevariablen weggelassen und neue an anderer Stelle hinzugefügt) könnte ein gleich hohes Bestimmtheitsmaß, aber deutlich andere "Erwartungswerte" liefern.

Trotz der angesprochenen Probleme ist eine solche Vorgehensweise, um eine sorgfältige Analyse der Autokorrelationen innerhalb der Vorhersagevariablen erweitert, sicherlich ein denkbarer Weg zur Wahlanalyse. Die Frage ist nur, welche Qualität eine solche Analyse innerhalb der empirischen Wahlforschung haben soll, falls hinter den selektierten Vorhersagevariablen keine substantielle Erklärung steht. In der Analyse von *Eckey und Klemmer* ist die Anzahl der Privathaushalte mit Telefonanschluß als Vorhersagevariable berücksichtigt, obwohl sicher niemand im Ernst annehmen wird, daß es das Wahlverhalten einer Familie beeinflusst, ob ein Telefon in der Wohnung vorhanden ist, oder nicht. Vielmehr ist die entsprechende Maßzahl nur

²⁴Siehe *Hartung/Elpelt (1984)*

einer von vielen möglichen Indikatoren für den "sozialen Wohlstand" der Wähler, die alle in das Modell einfließen. Besser wäre es, nur einen, und zwar den "besten" Indikator für den genannten Einfluß des Wohlstandes auf die Wahlentscheidung zu definieren und im Modell zu berücksichtigen. Zum anderen ist das Problem ökologischer Fehlansagen in diesen Modellansätzen überhaupt nicht berücksichtigt.

Die vorliegende Studie hat eine ganze Reihe von Vorzügen gegenüber den genannten üblichen regressionsanalytischen Ansätzen aus dem Bereich der empirischen Wahlforschung.

Sie beschäftigt sich explizit mit dem Problem des ökologischen Fehlschlusses und präsentiert eine mögliche Lösung. Andere Studien erkennen das Problem zwar, ignorieren es dann aber in der nachfolgenden Analyse.

Die Studie entwickelt ein Erklärungsmodell für individuelles Wahlverhalten eines Wählers abhängig davon, ob dieser selbständig ist oder nicht, und abhängig von bestimmten Attributen seines Stimmbezirkes, also sehr kleinräumig.

Die Interpretation des entwickelten Modells ist klar und eindeutig. Diese Eigenschaften ergeben sich, weil schon der Ausgangspunkt der Analyse, ausgehend von ökologischen Problemen, die späteren Interpretationsmöglichkeiten waren.

So beantwortet die vorliegende Analyse die Frage, wie die Wahrscheinlichkeit eines Wählers, seine Stimme für eine bestimmte Partei abzugeben, abhängig von bestimmten Attributen seines Stimmbezirkes variiert und nicht die normalerweise übliche Frage, welche Menge von Vorhersagevariablen den höchsten Anteil an der Gesamtvarianz erklären.

Auf dem methodischen Weg zum endgültigen Modell werden dabei aufgrund der Residualanalyse eine Reihe von wichtigen Sonderproblemen (Altersheimen, regionale Verschätzungen) offengelegt, die für eine weitere Analyse sehr wichtig sind, und bei einer anderen Analyse wahrscheinlich ignoriert worden wären.

6 Schlußkommentar

Es wurde in dieser Studie ein multiples Regressionsmodell im ökologischen Kontext der Gegebenheiten bei der Wahlanalyse entwickelt. Dabei sollte aufgezeigt werden, daß eine Modellbildung über Residualanalyse wesentlich überzeugendere Ergebnisse liefert als der übliche Weg über multiple Korrelationsmatrizen. Somit soll auch die Methodik der multiplen Regressionsanalyse für die Wahlanalyse rehabilitiert werden gegenüber anderen multivariaten Modellansätzen, und auch vor allem der nicht-empirischen Theoriebildung.

Die in dieser Studie entwickelten Regressionsmodelle für die Stimmenanteile von SPD und CDU bei der Kommunalwahl 1989 in Dortmund erklären mit 72 % bzw 78 % bedeutsame Teile der Gesamtvarianz.

Dabei flossen die hohen Anteile der Nichtwähler, wie sie gerade bei Kommunalwahlen schon sehr üblich sind, mit in die Analyse ein. Die vorliegenden Ergebnisse sind leicht zu interpretieren und nachvollziehbar.

Andererseits ging natürlich durch den Übergang auf Indikatorvariablen etwas Genauigkeit verloren. Es bleibt auch zu prüfen, ob durch Veränderung in der Variablendefinition (z.B. Verwendung anderer als der 25 % - und 75 % -Quantile) die Ergebnisse deutlich verändern würden.

Weiterhin wäre ein Nachprüfen der präsentierten Ergebnisse durch Umfragen o.ä. sehr nützlich. Gerade auch deshalb, weil sich seit 1987 gewisse Veränderungen der Bevölkerungsstruktur ergeben haben.

Es soll an dieser Stelle noch einmal betont werden, daß die beiden schließlich entwickelten Modelle nicht statisch sind. Besteht die begründete Annahme, daß wichtige Faktoren in diese Modelle noch nicht eingeflossen sind, läßt sich die Modellstruktur problemlos erweitern, beispielsweise um generative Faktoren.

Andererseits kann das Modell keinesfalls einfach auf andere Wahlen oder Städte übertragen werden ohne vorheriges sorgfältiges Nachprüfen, ob die Voraussetzungen dort überhaupt vergleichbar sind.

Wenn in wenigen Wochen die Ergebnisse der ersten gesamtdeutschen Wahl vorliegen, wäre es sehr interessant nachzuprüfen, inwieweit die vorliegende Modellstruktur auch geeignet ist, die Stimmenanteile der Parteien bei der zu erwartenden hohen Ausnutzung des sogenannten Wählerpotentials zu beschreiben.

Referenzen

- [1] Anscombe, F.J.(1960), "The Rejection of Outliers", *Technometrics*, 2, Seite 123-147
- [2] Bürklin, W.(1988). *Wählerverhalten und Wertewandel*, Leske und Budrich, Opladen
- [3] Crewe, I. (1976). Another Game with Nature: An Ecological Regression Model of the British Two Party Vote Ratio in 1970, *British Journal of Political Science* 6, Seite 43-81
- [4] Diederich, N.(1965). *Empirische Wahlforschung - Konzeptionen und Methoden im internationalen Vergleich*. Westdeutscher Verlag, Köln und Opladen
- [5] Dillon, W.-R. und Goldstein, M.(1984). *Multivariate Analysis. Methods and Applications*, Wiley, New York
- [6] Dinse, J.(1989). *Kleinräumige Strukturanalysen - Nutzung von VZ-Ergebnissen für Wahlanalysen*. Berichtspapier zur Vortrag auf der Verbandstagung, Ausschuß Wahlforschung, Verband Deutscher Städtestatistiker, Frankfurt a.M., 1.12.1989
- [7] Draper, N.R. und Smith, H.(1981). *Applied Regression Analysis*, 2nd Edition. Wiley, New York
- [8] Falter, J. u.a.(1983). Arbeitslosigkeit und Nationalsozialismus. Eine empirische Analyse des Beitrags der Massenarbeitslosigkeit zu den Wahlerfolgen der NSDAP 1932 und 1933, *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 35, Seite 525-551
- [9] Frey, B.S. und Weck, H.(1981). Hat Arbeitslosigkeit den Aufstieg des Nationalsozialismus bewirkt? *Jahrbuch für Nationalökonomie und Statistik*, 196, Stuttgart und New York, Seite 1-31
- [10] Goodman, L.A.(1953). "Ecological Regression and the Behavior of Individuals", *American Sociological Review*, XXVIII, Seite 663-65
- [11] Goodman und Kruskal, W.H.(1954), "Measures of Association for Cross Classifications", *Journal of the American Statistical Association*, XLIX, Seite 732-64

- [12] Goodman L.A.(1959). "Some Alternatives to Ecological Correlation", *American Journal of Sociology*, LXIV, Seite 610-15
- [13] Hahn, H. und Kemper, F.J.(1985). "Sozialökonomische Struktur und Wahlverhalten am Beispiel der Bundestagswahlen von 1980 und 1983 in Essen", *Arbeiten zur Rheinischen Landeskunde*, Heft 53, Dümmlers, Bonn
- [14] Hartung, J. und Elpelt, B.(1984). *Multivariate Statistik*, Oldenbourg, München und Wien
- [15] Heere, K.(1977). *Anwendung Statistischer Modelle für Wahlanalytische Untersuchungen auf der Basis von Aggregatdaten*, Dissertation am Fachbereich für Rechts- und Wirtschaftswissenschaften, TH Darmstadt
- [16] Holler, M.J.(Hrsg.)(1984). *Wahlanalyse - Hypothesen, Methoden und Ergebnisse*, tuduv, München
- [17] Kaltefleiter, W. und Nißen, P.(1980). *Empirische Wahlforschung - Eine Einführung in Theorie und Technik*, Schönigh, Paderborn
- [18] Marinell, G.(1986). *Multivariate Verfahren*, 2. Auflage, Oldenbourg, München und Wien
- [19] Mielke, G.(1987). *Sozialer Wandel und politische Dominanz in Baden Württemberg*. Duncker & Humblot, Berlin
- [20] Miller, W.L.(1972). "Measures of Electoral Change Using Aggregate Data", *Journal of the Royal Statistical Society, Series A*, 135, Seite 122-42
- [21] Miller, W.L. u.a.(1974). "Voting Research and the Population Census 1918-1971: Data for Constituency Analysis", *Journal of the Royal Statistical Society, Series A*, 137, Seite 384-411
- [22] Robinson, W.S.(1950). "Ecological Correlations and the Behavior of Individuals", *American Sociological Review*, XV, Seite 351-57
- [23] SAS Institute Inc.(Hrsg.)(1985). *SAS (R) User's Guide: Basics. Version 5 Edition*, Cary, NC
- [24] SAS Institute Inc.(Hrsg.)(1985). *SAS/GRAPH (R) User's Guide. Version 5 Edition*, Cary, NC

- [25] SAS Institute Inc.(Hrsg.)(1985). *SAS (R) User's Guide: Statistics. Version 5 Edition*, Cary, NC
- [26] Schuchard-Ficher, C. u.a.(1985). *Multivariate Analysemethoden*, 3. Auflage, Springer, Berlin und Heidelberg
- [27] Srivastava, M.S. / Carter, E.M.(1983). *An Introduction to Applied Multivariate Statistics*, North Holland, Amsterdam
- [28] Stadt Hamm, OStd., Hauptamt (Hrsg.)(1976). "Wahlanalyse der Stadt Hamm", *Schriftenreihe der Stadt Hamm*, 27, Reimann & Co., Hamm
- [29] Stadt Dortmund, OStd. ,Amt für Statistik und Wahlen (Hrsg.)(1990). "Die Kommunalwahl am 1. Oktober 1989", *Dortmunder Statistik*, Sonderheft 123
- [30] Stadt Düsseldorf, OStd., Amt für Statistik und Wahlen (Hrsg.)(1990). "Lebensbesingungen und Wahlentscheidung", *Beiträge zur Statistik und Stadtforschung*, Heft 36
- [31] Stadt Essen, Amt für Entwicklungsplanung, Statistik, Stadtforschung, Wahlen und Stadtarchiv (Hrsg.) (1989). *Die Essener Ergebnisse der Kommunalwahl vom 1. Oktober 1989*, Eigenverlag

A Detaillierte Einzelergebnisdarstellung

SAS			SAS		
NUMMER	SPD_RES	CDU_RES	NUMMER	SPD_RES	CDU_RES
1	0.4732	-0.3214	312	2.948	0.5105
2	-5.5738	-0.7375	313	2.198	-1.4205
3	-3.3586	-1.3322	314	3.700	-1.8700
4	0.7677	-3.1784	315	7.808	-2.6495
5	0.0678	-1.9691	316	4.543	-3.4457
6	-0.6988	2.4192	401	-2.686	-1.7286
7	-1.8427	3.1470	402	-10.008	4.4645
8	2.4505	1.0731	403	-4.422	2.2721
9	-5.9466	3.2336	404	-7.735	3.4709
10	-3.9410	-0.0535	405	1.087	-0.9549
11	-4.9107	3.1370	406	-1.339	1.2332
12	-2.0507	-3.3638	407	-4.139	-0.7082
13	0.2793	-7.3985	408	-1.502	-1.3451
14	-0.2777	-0.4806	451	1.420	-1.0906
15	-1.5800	-0.0440	452	-1.455	-1.0610
16	-3.4409	0.9944	453	-1.384	-2.7098
101	0.5618	0.8328	454	-4.447	-2.4985
102	1.5951	-0.0576	455	-1.351	-3.9104
103	-4.9780	-1.2326	456	-4.305	-0.1360
104	-0.6411	-5.7233	457	-3.247	-1.8001
105	2.4507	-1.0315	458	1.841	0.3584
106	-1.4847	-0.8476	459	-0.244	-0.9954
107	-2.1964	1.4490	460	0.705	-1.2977
108	5.1735	-1.3363	461	-4.823	-0.2446
109	1.2655	-1.3809	462	-2.326	3.0651
110	0.4469	-2.6690	501	-3.856	-0.9997
111	-0.9471	-1.4131	502	-0.111	-4.1178
112	-3.0880	-1.0872	503	-5.720	-5.3369
113	-1.1249	-5.6166	504	-5.617	-1.3996
114	-1.7909	0.5659	505	0.605	0.1658
115	0.5241	3.7637	506	-0.597	-2.1654
190	-0.8486	0.4684	507	0.618	-1.0428
201	-6.4678	-1.2300	508	0.200	0.6179
202	-2.4872	3.3461	509	-0.108	0.6638
203	-1.7909	-2.2187	510	-2.632	0.3755
204	2.4870	1.9031	511	0.859	-2.7347
205	-0.8357	0.8722	512	4.810	-0.9590
206	-0.0021	2.6829	601	-7.981	2.5343
207	-1.6633	3.2211	602	-2.005	0.1450
208	-0.7012	-2.3538	603	-0.734	-1.1959
209	-4.5564	0.9085	604	-1.205	-1.2538
210	-0.8266	-3.9757	605	-6.097	1.3431
211	-0.6464	-3.3663	606	-0.116	2.8884
212	-2.5707	-3.0247	607	0.208	-3.2298
301	0.3422	1.9976	608	2.176	-2.6363
302	-3.1301	2.9229	609	0.879	-2.4650
303	-5.7262	2.7882	610	1.946	3.7871
304	-5.5209	1.2699	611	-3.345	3.9015
305	-0.4548	0.5575	612	-1.878	-2.2809
306	-3.8613	6.2866	613	-1.659	-1.0098
307	2.1379	0.1431	701	-0.251	-0.4585
308	1.3464	1.7711	702	2.107	-2.0045
309	-0.3960	-0.2217	703	-4.789	4.8486
310	3.2294	0.3236	704	0.272	2.8472
311	-2.6077	2.4067	705	1.148	-0.2668

SPD_RES sind die Abweichungen in Prozentpunkten vom Erwartungswert für die SPD und
 CDU_RES sind die Abweichungen in Prozentpunkten vom Erwartungswert für die CDU und

Tabelle 10: Liste der Residuen der endgültigen Modelle für SPD- und CDU-Stimmenanteile auf Stimmbezirksebene

SAS			SAS		
NUMMER	SPD_RES	CDU_RES	NUMMER	SPD_RES	CDU_RES
706	-0.2173	4.1035	906	-2.1895	2.7597
707	2.4570	4.9425	907	-4.0704	-0.9544
708	-1.0855	3.6985	908	1.9128	2.5558
709	-3.5855	-1.5801	909	-1.8302	-0.6705
710	1.1515	-1.1096	910	-1.7568	-3.9071
711	-0.8025	-2.8900	911	-6.0164	0.6602
712	-2.7806	-4.9893	912	2.2203	-1.7901
713	3.0873	3.5461	913	0.4805	0.5951
714	-3.8942	-0.3120	914	-0.5832	-0.2562
715	-1.8772	3.0675	915	-0.8615	-2.0951
716	4.1774	0.2002	916	2.7526	0.2015
751	0.8573	2.4774	1101	7.1979	-1.9179
752	1.8104	2.6322	1102	-2.3164	0.0412
753	3.5062	0.0603	1103	3.4046	-3.8282
754	-3.0313	-0.4106	1104	-2.5489	-1.4118
755	-2.2175	-4.4086	1109	3.1195	-0.8723
756	-3.1223	-1.8629	1111	4.6306	-1.5758
757	1.7042	0.4880	1112	4.5465	-2.7496
758	0.7793	-0.5044	1113	8.1457	-1.1188
759	-1.8628	0.2176	1114	8.5771	-0.9935
760	-2.9644	0.3530	1201	3.3349	-0.6228
761	-5.8576	-2.2658	1202	0.0055	-2.2822
762	-5.5375	0.9453	1203	0.9147	-1.8761
763	-8.0157	2.6630	1204	3.6841	-1.0366
764	-1.0886	-0.5795	1205	3.4645	-2.1367
765	-3.7361	-0.1905	1206	5.2627	0.3400
766	3.9781	-0.5299	1207	-1.5290	-3.8859
767	-5.7460	3.7405	1208	-0.0969	-1.4702
768	-2.2530	1.5133	1209	2.6500	-1.9046
769	.	3.9047	1210	10.3779	0.4787
770	.	2.8004	1211	-1.0979	-0.8897
801	-3.5634	-1.2711	1301	-1.3656	0.9730
802	-1.1891	4.6030	1302	3.7443	-0.0587
803	-1.7592	-1.6695	1303	2.5832	-2.1967
804	5.5490	-2.9522	1304	-3.9798	1.3271
805	-6.4113	1.5864	1305	0.7500	-1.9073
806	-6.9569	1.3891	1306	-4.4066	-5.1328
807	-0.9947	0.1439	1307	6.4025	1.5481
808	-2.8402	6.3615	1308	0.1798	-2.7629
809	-3.7241	2.2576	1309	-1.7636	-2.8260
810	-0.3290	-1.3649	1310	7.2387	-0.1113
811	-1.0852	1.6843	1311	0.5797	-0.3358
812	-2.6628	1.8570	1312	5.8657	-4.4060
813	-2.8996	-0.2481	1313	6.3002	-0.3132
814	2.7981	-1.4846	2101	4.4209	-0.7452
815	0.7926	-5.7309	2102	8.9854	-1.7013
816	-6.3145	3.5016	2103	4.3968	0.8360
817	-0.8311	0.4565	2104	3.1505	0.2767
818	-5.2396	0.7743	2105	2.3187	-1.4254
819	-5.9745	2.9663	2106	-2.1913	0.5589
901	3.9528	-1.3241	2107	-0.5443	-3.5268
902	1.0147	1.8227	2108	2.2848	-1.8378
903	-2.2598	2.1535	2109	5.0159	-2.6597
904	0.9181	-5.1318	2110	6.9858	-4.8795
905	-0.1403	-0.2357	2111	3.2755	-1.9891

SAS			SAS		
NUMMER	SPD_RES	CDU_RES	NUMMER	SPD_RES	CDU_RES
2112	1.762	.	3209	0.2549	2.1930
2113	-4.240	5.9086	3210	1.2972	5.9256
2114	1.842	2.8449	3211	7.5863	-3.0103
2115	-5.132	1.9206	3212	-1.9650	3.4563
2116	-1.061	3.6990	3213	6.8912	-3.7536
2117	-3.662	3.4373	3215	-1.4778	-0.3560
2201	2.410	-1.3059	3301	-4.6307	2.1218
2202	7.405	2.5385	3302	0.7366	1.0003
2203	6.387	2.1581	3303	-0.8890	3.2288
2204	-1.966	.	3304	-4.4038	2.2983
2205	3.439	2.4182	3305	-2.1998	-2.9921
2206	-7.153	3.0944	3306	0.8525	1.7193
2207	-1.722	1.3462	3307	-3.7669	3.9333
2208	-4.967	0.9629	3308	4.8279	0.0921
2209	-10.543	0.7820	3309	0.2554	-0.6809
2210	-1.261	0.5572	3310	2.1367	3.9992
2211	-4.302	2.4102	3311	1.5021	0.2866
2212	-3.042	-3.2308	3312	0.3232	1.3505
2301	-4.362	-0.7431	3313	8.4866	-3.4302
2302	-4.034	-3.2487	3401	-5.0291	2.9013
2303	-1.234	1.4964	3402	0.9104	-2.8664
2304	1.819	-3.8907	3403	1.9522	-1.0667
2305	3.311	-0.8867	3405	-0.3070	2.9469
2306	5.295	-2.8160	3406	-4.7154	7.5582
2307	-0.300	2.2431	3407	6.2081	6.4720
2308	6.706	-0.0331	3409	-2.0659	3.5455
2309	2.189	-1.7444	3410	-0.0207	-0.1093
2310	-3.823	2.8565	3411	-1.7112	-0.2622
2311	-4.367	-0.1311	3412	1.6930	1.5978
2312	-4.962	-4.0234	3413	2.9107	5.3021
2313	-1.320	0.2546	3414	-2.1333	2.1811
2314	2.185	1.7349	3415	-1.4200	2.4276
2315	-0.069	3.6783	4101	-0.6184	2.0184
2316	3.774	-2.5623	4102	2.0230	1.5569
3101	-4.482	-0.7962	4103	-0.1305	-0.5270
3102	6.019	-0.6247	4104	-0.6720	-1.8722
3103	5.315	-1.5648	4105	-0.7664	-0.3451
3104	1.602	0.4522	4106	-0.2342	-6.4167
3105	0.679	3.1506	4107	-3.7591	-0.5996
3106	-0.519	5.0596	4108	-3.8213	2.8352
3107	0.934	1.8088	4109	4.1847	0.0951
3108	-2.443	0.5479	4110	-1.9938	0.1355
3109	3.238	-3.6247	4111	3.0519	-1.8999
3110	-1.291	4.6783	4112	-5.8980	1.5471
3111	-2.650	-2.3181	4113	-4.2889	-0.0166
3112	-3.723	2.7889	4114	4.8641	0.6271
3113	0.257	-1.0036	4201	0.8957	-1.4361
3201	4.254	-1.3259	4202	2.0288	1.5237
3202	0.384	0.9049	4203	7.6167	0.1679
3203	-2.904	4.1695	4204	6.7317	1.8938
3204	0.078	-3.6317	4205	0.0845	-0.2036
3205	0.217	1.3349	4206	-4.6326	2.6790
3206	-1.422	1.0350	4207	5.1278	3.2860
3207	1.222	-3.9669	4208	1.6498	-1.5866
3208	3.403	-4.8638	4209	0.4362	-6.7388

SAS			SAS		
NUMMER	SPD_RES	CDU_RES	NUMMER	SPD_RES	CDU_RES
4210	1.197	-0.3114	5215	2.6329	-2.3055
4211	-0.037	2.2018	5216	0.6108	-1.5701
4301	2.703	-1.3152	5217	1.9748	1.9990
4302	0.077	-3.2229	5301	1.3385	-1.8715
4303	2.482	-4.6913	5302	-1.1649	10.4502
4304	1.447	-1.9336	5303	2.8435	-2.0988
4305	-1.773	1.9589	5304	3.1352	-6.1740
4306	3.155	-1.8371	5305	2.3916	1.1466
4308	-1.167	-2.3760	5306	6.4898	0.2205
4309	-1.681	-2.9010	5307	-1.1698	2.2010
4310	-2.764	-4.5771	5308	-3.2576	-3.0027
4311	2.957	-0.2232	5309	1.4606	-1.3215
4312	-1.343	4.7009	5310	-1.0823	0.9444
4313	-1.666	6.9624	5311	-3.1596	1.3226
4314	1.104	1.6258	5312	2.5103	-1.4178
4315	4.281	-3.5954	5313	-4.7167	-2.8673
4316	7.301	-0.0978	5314	-2.4524	-0.6407
4317	2.866	1.4203	5315	-3.5633	1.8478
4318	2.220	-6.2800	5401	2.5403	-2.9744
4401	5.965	-0.0727	5402	4.2677	1.5033
4402	9.612	-2.1612	5403	2.3508	-3.5929
4403	-0.222	-1.9669	5404	7.6696	-1.1565
4405	4.847	0.4942	5405	4.1168	-1.2369
4406	0.008	-1.8469	5406	6.0723	-2.4736
4407	0.182	-1.2704	5407	3.6882	-3.4501
4408	-4.600	-4.0040	5408	2.4160	-1.6013
4409	2.468	-0.0926	5409	-3.2586	0.8937
4410	-8.466	-2.7212	5410	0.1351	0.5204
5101	-0.116	1.8336	5411	1.7147	-1.8294
5102	-2.011	-2.1614	5412	-1.0064	-0.1370
5103	-1.077	-5.9050	5414	-0.1670	1.9914
5104	8.093	0.4953	6101	3.4512	-1.0665
5105	1.665	-2.3418	6102	1.4704	-0.0508
5106	-0.035	1.2500	6103	2.1251	2.6348
5107	-4.256	-0.0974	6104	1.9880	-0.7465
5108	-12.484	0.6326	6106	0.6624	1.7223
5109	-1.168	0.5888	6107	4.1255	0.8223
5110	-0.510	-0.9759	6108	1.0946	2.6630
5111	1.803	-3.3362	6109	-3.8306	-2.3587
5112	0.415	-2.6702	6110	6.0297	-0.2762
5113	2.749	0.0622	6111	2.2889	-0.4548
5201	0.146	1.6446	6112	1.6486	-1.2678
5202	-7.678	1.6117	6113	-0.8348	2.7673
5203	-3.826	-0.7718	6114	-5.9802	0.4649
5204	-9.467	6.7384	6115	-0.6736	1.0990
5205	-0.914	-1.5082	6116	2.9279	1.5959
5206	3.659	-1.4498	6117	2.2026	-0.2097
5207	0.746	-2.9405	6191	9.2936	3.4687
5208	0.718	-1.1039	6192	-3.2275	4.3231
5209	-3.116	0.0036	6201	1.9388	-0.1068
5210	1.477	1.5359	6202	-0.2391	-1.4495
5211	3.343	-1.8463	6203	5.7014	-1.6184
5212	-8.054	1.4066	6204	4.8907	0.4685
5213	-2.526	-0.1455	6205	-0.1956	2.8640
5214	1.684	-2.2605	6206	-5.8514	5.1329

SAS			SAS		
NUMMER	SPD_RES	CDU_RES	NUMMER	SPD_RES	CDU_RES
6207	4.875	1.7678	7212	6.3898	-1.3951
6208	0.720	0.4809	7213	5.0740	-0.1132
6209	3.730	4.1267	7214	4.2896	0.0412
6210	0.541	1.6315	7215	0.6249	-1.2116
6211	-1.071	5.8619	7216	-1.6739	1.3518
6212	-0.948	4.8096	7217	-2.1622	-0.4838
6213	6.907	-1.3263	7301	-6.6988	0.1177
6214	-1.844	4.3930	7302	1.7179	-1.7871
6215	-2.122	1.3037	7309	-0.3665	-1.3151
6216	3.943	1.3408	7310	7.6816	-1.8592
6217	-6.175	2.2227	7311	-4.5640	-1.3949
6218	-0.091	2.1596	7312	-2.9907	2.6614
6219	-6.619	-2.2569	7313	2.7298	2.1446
6220	5.476	.	7314	1.4758	0.4742
6221	0.350	0.9229	7315	-0.0946	3.4664
6301	-1.589	.	7316	4.3786	-2.7178
6302	0.411	3.6017	8101	-4.2615	1.4447
6303	-1.511	4.0796	8102	6.7106	-3.3301
6304	4.106	0.2986	8103	8.6610	0.6030
6305	1.741	4.9600	8104	2.3006	-2.1572
6306	-0.679	2.1685	8105	2.1261	2.6357
6307	10.028	2.1349	8106	1.0380	3.7888
6308	3.948	6.1134	8107	2.3947	2.7004
6309	-4.155	3.8706	8108	-8.6263	2.5337
6310	-3.801	4.8741	8109	-3.0268	0.9212
6311	-2.181	7.2141	8201	-2.9368	-0.7325
6312	2.986	2.7370	8202	.	-0.2849
6314	-1.950	3.4040	8203	.	-2.3843
6315	5.361	4.0376	8204	9.7786	-0.8162
6316	-5.588	1.1630	8205	2.7374	0.0987
6391	-3.529	-2.8421	8206	-0.3531	-0.1350
6392	6.880	-1.4998	8207	8.0082	-0.4250
7101	6.423	-4.8214	8208	-1.2653	2.0244
7102	-0.899	1.2460	8209	3.6468	-0.0515
7103	-4.556	1.6344	8210	6.6979	-1.6584
7104	-2.097	7.7889	8211	0.7923	0.9026
7105	-1.490	-0.8624	8212	6.2746	1.3846
7106	-7.292	2.4779	8213	2.3852	-4.2547
7107	-4.774	0.5031	8301	5.2441	2.9012
7108	3.304	-0.0518	8302	3.8426	-3.1761
7109	-1.077	0.8983	8303	-0.9056	2.2540
7112	1.535	1.0276	8304	-3.4405	-1.1085
7113	6.062	0.2939	8305	-4.8618	5.0239
7114	-10.001	0.9700	8306	1.2980	0.9699
7201	-0.311	-3.1331	8307	6.4997	-4.2459
7202	1.062	-0.2220	8308	-1.1490	-1.4374
7203	-0.235	-2.1800	8311	1.2611	-2.6003
7204	1.291	3.0490	8314	2.1262	0.4082
7205	0.911	1.8859	9101	-3.1749	-2.6059
7206	2.805	-3.0741	9102	-3.2119	-6.2466
7207	2.642	-0.0688	9103	-5.5148	-4.9672
7208	5.687	1.8498	9104	2.1499	-4.6021
7209	-2.440	-1.3619	9105	1.1342	-3.6288
7210	-0.001	-1.3406	9106	-3.1338	-1.6014
7211	6.426	-0.1654	9107	-7.8450	-7.4490

A DETAILLIERTE ERGEBNISDARSTELLUNG

69

NUMMER	SAS	
	SPD_RES	CDU_RES
9108	-4.160	-3.3461
9109	-8.901	-2.1349
9110	-1.710	-1.7028
9111	-7.904	1.9436
9112	-6.256	-2.4674
9113	5.085	-1.0823
9114	1.106	-3.9116
9115	-4.886	-3.3925
9201	-6.352	1.9197
9202	-5.215	0.8966
9203	-0.808	1.0421
9204	1.076	-3.1105
9205	4.834	-1.4054
9206	-0.532	-0.8560
9207	-2.841	-2.0137
9208	-1.735	-0.0092
9209	-5.617	1.4494
9210	3.380	-1.3291
9211	0.885	-1.2230
9212	-2.122	-3.0598
9213	-6.604	-0.7164
9301	3.232	-0.0030
9302	-1.469	-2.8888
9304	-2.347	0.6829
9305	3.273	1.1054
9306	-1.147	2.1448
9307	0.827	0.9926
9308	-6.818	0.4796
9309	3.571	1.3672
9310	7.618	-2.0431
9311	-12.753	0.6998
9312	-3.400	5.4489
9313	-3.680	-0.2530
9314	-5.983	1.7691