

Jochen Groß

Die Prognose
Wahlergebnisse

Ansätze und
Leistungsfähigkeit

Jochen Groß

Die Prognose von Wahlergebnissen

VS RESEARCH

Forschung und Entwicklung in der Analytischen Soziologie

Herausgegeben von

Prof. Dr. Monika Jungbauer-Gans, Christian-Albrechts-Universität Kiel

Die Reihe nimmt die Forderung der Analytischen Soziologie auf, dass sich die soziologische Theoriediskussion stärker auf erklärende soziale Mechanismen konzentrieren sollte. Die Analytische Soziologie sucht nach präzisen, handlungstheoretisch fundierten Erklärungen für soziale Phänomene. Dabei soll eine Balance zwischen einer abstrahierenden und einer realitätsgerechten Theoriebildung gehalten werden. Im Vordergrund der Reihe steht nicht die Theorieentwicklung und -diskussion, sondern die empirische Umsetzung, die sich den skizzierten theoretischen Grundsätzen verpflichtet fühlt. Der handlungstheoretischen Fundierung widerspricht nicht, dass auch Makrophänomene und insbesondere die Wechselwirkungen zwischen Strukturen und Individuen untersucht werden. Die Reihe bietet in Folge dessen ein Forum für NachwuchswissenschaftlerInnen, welche die theoretischen Überlegungen der Analytischen Soziologie konsequent in empirischen Untersuchungen umsetzen.

Jochen Groß

Die Prognose von Wahlergebnissen

Ansätze und empirische
Leistungsfähigkeit

Mit Geleitworten von Prof. Dr. Monika Jungbauer-Gans
und Prof. Dr. Thomas Hinz

VS RESEARCH

Bibliografische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der
Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über
<<http://dnb.d-nb.de>> abrufbar.

Dissertation Universität Konstanz, 2009

1. Auflage 2010

Alle Rechte vorbehalten

© VS Verlag für Sozialwissenschaften | Springer Fachmedien Wiesbaden GmbH 2010

Lektorat: Verena Metzger / Dr. Tatjana Rollnik-Manke

VS Verlag für Sozialwissenschaften ist eine Marke von Springer Fachmedien.

Springer Fachmedien ist Teil der Fachverlagsgruppe Springer Science+Business Media.

www.vs-verlag.de



Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung außerhalb der engen Grenzen des Urheberrechtsgesetzes ist ohne Zustimmung des Verlags unzulässig und strafbar. Das gilt insbesondere für Vervielfältigungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.

Die Wiedergabe von Gebrauchsnamen, Handelsnamen, Warenbezeichnungen usw. in diesem Werk berechtigt auch ohne besondere Kennzeichnung nicht zu der Annahme, dass solche Namen im Sinne der Warenzeichen- und Markenschutz-Gesetzgebung als frei zu betrachten wären und daher von jedermann benutzt werden dürften.

Umschlaggestaltung: KünkelLopka Medienentwicklung, Heidelberg

Gedruckt auf säurefreiem und chlorfrei gebleichtem Papier

Printed in Germany

ISBN 978-3-531-17273-6

Geleitwort

Die analytische Soziologie hat sich das Ziel gesetzt, Erklärungen für soziale Phänomene zu finden, indem soziale Mechanismen herausgearbeitet werden, denen zu Folge die Handlungen von Akteuren regelmäßig zu den beobachteten Phänomenen führen. In der vorliegenden Dissertation von Jochen Groß wird ein Prognosemodell für Wahlergebnisse entwickelt und umfragebasierten Prognosen sowie Wahlbörsen gegenübergestellt.

Prognosen können als eine rigorose Prüfung von theoretischen Erklärungsmodellen angesehen werden. Sowohl die verschiedenen umfragebasierten Verfahren, die in der Regel die berühmte Sonntagsfrage verwenden und häufig nicht ihre methodischen Vorgehensweisen z. B. in Bezug auf die Prinzipien der Gewichtung offen legen, als auch die von Ökonomen initiierten Wahlbörsen, die durch „reale“ Transaktionen das künftige Wahlverhalten simulieren, zielen nur auf die Vorhersage des blanken Wahlergebnisses ab. Sie versuchen aber nicht, das zugrunde liegende Verhalten der Wählerinnen und Wähler zu erklären. Die Entwicklung eines Prognosemodells hingegen hat das Ziel, den inneren Zusammenhängen des Entscheidungsprozesses ein Stück weit näher zu kommen, indem z. B. die sozialstrukturelle Zusammensetzung der Wählerschaft vor dem Hintergrund der jeweils typischen Bedürfnisse, Überzeugungen und Opportunitäten der durch diese Merkmale gekennzeichneten Wählergruppen unter dem Stichwort „politisierte Sozialstruktur“ berücksichtigt wird. Weitere langfristige strukturelle Faktoren sind das Alter der Bundesrepublik zur Abbildung institutioneller Sklerose und die langfristige Parteibindung der Wählerinnen und Wähler. Darüber hinaus werden im vorliegenden Modell mittel- und kurzfristige Einflüsse berücksichtigt, wie die ökonomische Situation (operationalisiert anhand der Arbeitslosenrate), politisch relevante Großereignisse und die Wählersympathie für die jeweiligen Spitzenkandidaten.

Ein wesentliches Resultat der Arbeit ist die Aufforderung zu einer vertieften Ausarbeitung der theoretischen Bezüge zur Erklärung der sozialen Mechanismen, die im Aggregat zu einem Wahlergebnis führen. Wichtig erscheint es dabei, den Konnex zwischen der Nachfrage nach politischer Willensbildung auf Seiten der Wählerinnen und Wähler und dem Angebot an Umsetzungsstrategien auf Seiten der politischen Parteien stärker zu berücksichtigen.

Monika Jungbauer-Gans

Geleitwort

Die sozialwissenschaftliche Wahlforschung hat bislang um eine systematische Auseinandersetzung mit Wahlprognosen einen großen Bogen gemacht. Dies liegt vor allem daran, dass die Erklärung des tatsächlichen Wahlverhaltens theoretischen und empirischen Vorrang besitzt, denn ohne die Entwicklung von Mikromodellen erscheint die wissenschaftlich solide Vorhersage zukünftiger Wahlausgänge nicht sinnvoll. Jochen Groß tritt einen Schritt zurück und untersucht die unterschiedlichen Methoden und Datengrundlagen von Wahlprognosen, *ohne* ein Mikromodell politischer Wahlen bereits vorauszusetzen. Umfragebasierte Wahlprognosen sind *per se* ein vielschichtiger Forschungsgegenstand und eng mit der Geschichte der Umfrageforschung verbunden. In der Nachkriegszeit wurden Wahlprognosen fester Bestandteil der politischen Kultur der Bundesrepublik Deutschland. Die politischen Parteien, die Medien und die Öffentlichkeit sind heute mehr denn je an der Erfassung von politischen Stimmungen und Wahlabsichten interessiert, gleichwohl werden Umfrageergebnisse und ihre Aussagekraft nach Belieben in Zweifel gezogen. Ja, wir wissen, Politiker wollen die Wahlen gewinnen und nicht die Umfragen. Dabei wird den Wahlprognosen selbst eine immer wichtigere Rolle zugeschrieben, da sie ihrerseits Wahlausgänge beeinflussen können. In jüngster Zeit haben sich alternativ zu Umfragen auch Wahlbörsen etabliert, bei denen ein Markt für politische Parteien simuliert wird. Die Preisbildung in diesem virtuellen Markt wird in Anlehnung an Friedrich Hayek ebenfalls als Mittel zur Vorhersage von Wahlergebnissen herangezogen.

Ein Maßstab, um die verschiedenen Prognosemodelle zu beurteilen, ist naheliegend. Man kann die aus Umfragen und Wahlbörsen gewonnenen Prognosen leicht mit den tatsächlichen Wahlausgängen abgleichen und ihre Treffgenauigkeit beschreiben. Die vorliegende Studie leistet diese Aufgabe für einen langen Zeitraum, sehr systematisch, mit großer Akkuratheit sowie hohem methodischem Problembewusstsein. Die Analysen von Jochen Groß erlauben nebenbei auch einen Einblick in die „Geheimkammern“ der Demoskopie, wenn klar wird, welche

herausragende Bedeutung den letztlich unbekanntem Gewichtungsfaktoren der Umfragedaten zukommt.

Mit den angestellten Auswertungen samt den durch sie aufgeworfenen Rätseln wird das Desiderat einer größeren Transparenz in der empirischen Wahlforschung offenkundig. Selbst klare Regelverletzungen in der Surveyforschung wie Quotenstichproben führen nicht unbedingt zu schlechteren Prognosen. Die Analysen unterstreichen alles in allem, dass für eine produktive Weiterentwicklung der Prognosefähigkeit der Sozialwissenschaft im Allgemeinen neben den komplexen, theoretischen Mikromodellen auch „einfache“ Fragen der Aggregation und der Datenqualität auf der Agenda stehen. Darüber auf der Grundlage der vorliegenden Arbeit nachzudenken, ist überaus anregend.

Thomas Hinz

Vorwort

Ausgangspunkt der vorliegenden Arbeit war der überraschende Ausgang der Bundestagswahl 2005. Entgegen der zuvor publizierten Ergebnisse von Meinungsumfragen schnitt insbesondere die CDU/CSU deutlich schlechter ab als erwartet. Gemeinhin wurde diese Bundestagswahl als eines von vielen Beispielen für das Versagen der Demoskopien und ihrer Methoden gewertet.

Die Berichterstattung über die mangelhafte Leistungsfähigkeit der politischen Meinungsforschung in Folge der Bundestagswahl von 2005 weckte das Interesse an der Thematik und führte zur Suche nach systematischen Analysen zur Prognosegüte der gescholtenen Sonntagsfragen. Es zeigte sich, dass bisher kaum entsprechende Untersuchungen vorliegen. Zudem überraschte, dass seitens der Politikwissenschaft bisher kaum Versuche zur Vorhersage von Wahlergebnissen unternommen wurden. Angesichts der Prognosefreudigkeit der Ökonomen und ihrer zunehmenden Betätigung auch auf angestammten Gebieten der Soziologie und der Politikwissenschaft wurde im Laufe der Bearbeitungszeit der vorliegenden Dissertation immer mehr die Notwendigkeit einer genuin politikwissenschaftlichen Prognoseaktivität deutlich. Beide Defizite werden in der vorliegenden Dissertationsschrift herausgearbeitet und bearbeitet. So wird eine systematische Evaluation vorliegender Sonntagsfragen durchgeführt sowie die Überwindung des politikwissenschaftlichen Prognoseskeptizismus angestrebt, indem ein erstes Modell zur Prognose von Zweitstimmenanteilen einzelner Parteien bei Bundestagswahlen entwickelt wird.

Das Verfassen dieser Arbeit wäre nicht möglich gewesen ohne die Unterstützung zahlreicher Personen, denen mein herzlicher Dank gilt. Zunächst bedanke ich mich bei meinen Betreuern, Prof. Norman Braun, Ph.D. und Prof. Dr. Thomas Hinz, welche diese Arbeit maßgeblich inspiriert, gestützt und engagiert begleitet haben. Norman Braun danke ich darüber hinaus für die Möglichkeit der Arbeit an seinem Lehrstuhl und die hilfreichen substanziellen Vorschläge, welche maßgeblich in die Arbeit eingeflossen sind. Thomas Hinz ist indes neben der

Übernahme der Betreuung dieser Arbeit für die langjährige Unterstützung und wissenschaftliche Motivation seit meinem Studium herzlich zu danken.

Besonderer Dank geht an Christian Ganser, dessen kritische wie genaue Lektüre des Manuskripts diese Arbeit wesentlich verbessert hat. Darüber hinaus schätze ich sehr den kontinuierlichen Austausch mit ihm in unserer Büro-WG, wodurch sich die Freuden des universitären Arbeitens erst in vollem Umfang entfalten konnten und manches Ärgernis leichter verarbeiten ließ. Weiterhin bedanken möchte ich mich bei meinen Kollegen Johannes Bauer, Roger Berger, Werner Fröhlich, Marc Keuschnigg, Eva Negele, Thomas Wimmer und Tobias Wolbring für die produktive und angenehme Arbeitsatmosphäre, die durch Kollegialität und gegenseitige Hilfsbereitschaft gekennzeichnet ist. Meinen ehemaligen Kollegen am Lehrstuhl Christiane Gross, Peter Kriwy und Sonja Pointner gilt für ihre Unterstützung insbesondere zu Beginn der Promotion ebenfalls mein Dank. Ein spezieller Dank geht an die studentischen Hilfskräfte Benjamin Gedon, Anja Hänsel und Patrick Riordan, die mich bei der Literaturrecherche sowie der Datenerhebung unterstützt haben. Hiervon habe ich ebenso profitiert wie von den Diskussionen im Rahmen der Seminare an der Venice International University zur „Rational Choice Sociology“. Allen Teilnehmern möchte ich für ihre hilfreichen Anmerkungen und die konstruktive Kritik danken.

Katrin Auspurg danke ich für den guten Draht nach Konstanz zur Bewältigung aller formalen Hürden und die gute Zusammenarbeit auch im Rahmen von gemeinsamen Lehrveranstaltungen. Eine zentrale Bürde hat Mariana Grgic auf sich genommen: Sie hat die Arbeit Korrektur gelesen – auch hierfür mein herzlichster Dank.

Prof. Dr. Werner Georg sowie Prof. Dr. Peter Selb danke ich für die Abnahme der mündlichen Promotionsprüfung an der Universität Konstanz. Prof. Dr. Monika Jungbauer-Gans danke ich für die Möglichkeit der Veröffentlichung der vorliegenden Dissertationsschrift in ihrer Reihe „Forschung und Entwicklung in der Analytischen Soziologie“ und ihre Unterstützung im Rahmen der Manuskripterstellung und -überarbeitung.

Die Unterstützung in nicht-fachlicher Hinsicht ist zum erfolgreichen Abschluss einer solchen Arbeit unabdingbar. Hierfür möchte ich zunächst meinen Eltern für ihre fortwährende Unterstützung danken. Schließlich möchte ich mich bei meiner großen Liebe Eva bedanken, ohne deren stetige Motivation und bedingungslosen Rückhalt in allen Lebensphasen diese Arbeit nicht entstanden wäre.

Jochen Groß

Inhaltsverzeichnis

1	Einleitung und Überblick	15
1.1	Ziele und Aufbau der Arbeit	15
1.2	Zum Begriff der Prognose	18
2	Ansätze zur Prognose von Wahlergebnissen	27
2.1	Umfragebasierte Vorhersagen	28
2.1.1	Meinungsforschung und empirische Wahlforschung in Deutschland	29
2.1.2	Kritik an der Meinungsforschung	37
2.1.2.1	Beeinflussung des Wählers	39
2.1.2.2	Beeinflussung politischer Entscheidungen	44
2.1.2.3	Beeinflussung der staatlichen Konstitution	46
2.1.2.4	Verwendung der Ergebnisse von Meinungsumfragen ..	47
2.1.3	Wahlabsichtsfragen	48
2.1.4	Wahlerwartungsfragen	57
2.1.5	Wahltagsbefragungen	58
2.1.6	Zum prognostischen Wert von umfragebasierten Vorhersagen .	60
2.1.7	Vorliegende Evidenz zur Güte von umfragebasierten Vorhersagen	65
2.2	Wahlbörsen	71
2.2.1	Funktionsweise von Wahlbörsen	72
2.2.2	Zum prognostischen Wert von Wahlbörsen	77
2.2.3	Vorliegende Evidenz zur Güte von Wahlbörsen	80

2.3	Prognosemodelle	86
2.3.1	Modellvorschläge für die USA	87
2.3.2	Das Prognosemodell für Bundestagswahlen von Gschwend und Norpoth	92
2.3.3	Zum prognostischen Wert von Prognosemodellen	98
2.3.4	Vorliegende Evidenz zur Güte von Prognosemodellen	100
3	Die Prognosegüte von Wahlabsichtsfragen	103
3.1	Einflussfaktoren auf die Prognosegüte von Wahlabsichtsfragen	104
3.2	Forschungsstand	117
3.3	Methodik	121
3.3.1	Daten	121
3.3.2	Maßzahlen zur Bestimmung des Prognosefehlers	126
3.3.2.1	Der mittlere absolute Fehler	127
3.3.2.2	Der mittlere absolute prozentuale Fehler.....	129
3.3.2.3	Der Vergleichsquotient v	130
3.3.2.4	Das A-Maß	132
3.3.3	Auswertungsverfahren	135
3.3.3.1	Zeitreihenanalyse	135
3.3.3.2	Mehrebenenanalyse	143
3.4	Empirische Resultate	146
3.4.1	Die politische Gewichtung der Wahlabsichtsfrage	147
3.4.2	Die Prognosegüte von Wahlabsichtsfragen	173
3.4.3	Einflussfaktoren auf die Prognosegüte von Wahlabsichtsfragen	196
4	Ein Prognosemodell für Bundestagswahlen	213
4.1	Die Modellvariablen	217
4.1.1	Normalwahl.....	218
4.1.1.1	Die langfristige Parteibindung im sozialpsycho- logischen Ansatz	218
4.1.1.2	Operationalisierung	222

4.1.2	Politierte Sozialstruktur	225
4.1.2.1	Entstehung und Verfestigung der politisierten Sozialstruktur	226
4.1.2.2	Dynamik der politisierten Sozialstruktur	228
4.1.2.3	Empirische Befunde	230
4.1.2.4	Operationalisierung	232
4.1.3	Wirtschaftslage und Regierungstätigkeit	236
4.1.3.1	Retrospektives Wählen	236
4.1.3.2	Operationalisierung	239
4.1.4	Institutionelle Sklerose	241
4.1.4.1	Die These vom Altern von Demokratien	241
4.1.4.2	Operationalisierung	243
4.1.5	Wahlbeeinflussende Ereignisse	245
4.1.5.1	Von Rallye- und Priming-Phänomenen	245
4.1.5.2	Operationalisierung	246
4.1.6	Simultanes Wirken vielfältiger Einflüsse	249
4.1.6.1	Mostellers Gesetz	249
4.1.6.2	Operationalisierung	250
4.1.7	Die Relevanz des Rangplatzes	254
4.1.7.1	Zipfs Gesetz und Kandidatenorientierung	254
4.1.7.2	Operationalisierung	255
4.2	Empirische Befunde zur Prognosegüte des Modells	257
4.2.1	Retrospektive Bewertung des Modells	260
4.2.2	Vergleich mit wahlabsichtstfragenbasierten Vorhersagen	266
5	Zusammenfassung und Fazit	275
5.1	Zusammenfassung zentraler Ergebnisse	275
5.2	Schlussfolgerungen und Ausblick	280
	Literaturverzeichnis	287

1 Einleitung und Überblick

1.1 Ziele und Aufbau der Arbeit

Wahlen erlauben die direkte Beteiligung der Bürger am politischen Prozess von Demokratien, wodurch ihnen eine herausragende Stellung zukommt. Entsprechend ihrer Bedeutung konzentriert sich auch das öffentliche Interesse auf Wahlen und die Vorwahlzeit. Die Wahlforschung stellt ein bedeutendes wissenschaftliches Feld innerhalb der Sozialwissenschaften dar und beschäftigt sich insbesondere mit der Beschreibung und Erklärung des Wählerverhaltens. In der Bundesrepublik Deutschland stehen dabei die Wahlen zum Bundestag im Mittelpunkt des Interesses. Gewählt wird dabei die erste Kammer der Legislative, welche wiederum den Bundeskanzler wählt.

Entgegen der Auffassung, dass Prognosen als ein zentrales Ziel wissenschaftlichen Arbeitens zu zählen sind, finden sich bisher nur wenige wissenschaftliche Versuche, den Ausgang bundesdeutscher Wahlen zu prognostizieren. Dies überrascht, da einerseits die empirische Wahlforschung gerne als eine der am weitesten entwickelten sozialwissenschaftlichen Disziplinen aufgefasst wird und andererseits, da entsprechende Versuche in anderen Staaten, etwa den USA, mittlerweile als Bestandteil der Wahlforschung fest verankert sind.

Nichtsdestotrotz erfreuen sich Spekulationen über den möglichen Wahlausgang auch in Deutschland bereits seit langem großer Beliebtheit. Hierbei sind insbesondere via Umfragen generierte Voraussagen relevant. Diese beruhen in der Regel auf der Erfassung der Wahlabsicht von ausgewählten Probanden (so genannte Sonntagsfrage). Trotz der zunehmenden Zahl an Veröffentlichungen solcher Sonntagsfragen und damit auch einer offenkundigen steigenden Nachfrage nach Ergebnissen politischer Umfragen durch die Medien stehen Wahlvoraussagen auf Grundlage der Sonntagsfrage im Kreuzfeuer der Kritik – und mit ihr die Meinungsforscher. Kaum eine Wahl vergeht, ohne dass die Demoskopien

aufgrund ihrer anscheinend unzutreffenden Voraussagen zum eigentlichen Wahlverlierer erklärt werden.

Festzuhalten ist also einerseits, dass Prognosen ein Ziel wissenschaftlicher Betätigung sind, die deutsche Wahlforschung diesbezüglich jedoch bisher kaum Vorschläge für Prognosemodelle unterbreitet hat. Andererseits erfreuen sich dennoch Voraussagen von Wahlergebnissen auf Grundlage von Wahlabsichtsfragen zunehmender Beliebtheit bei gleichzeitig anhaltender Kritik dieser. Erstaunlicherweise beruht diese Kritik nicht auf systematischen Evaluationen des Ansatzes der Wahlabsichtsfrage oder deren Vorhersagegüte im Zeitablauf.

In dieser Arbeit werden deshalb zwei Zielsetzungen verfolgt, welche die genannten Lücken schließen sollen. *Erstens* soll der herausgehobenen Bedeutung von umfragebasierten Voraussagen von Wahlergebnissen in der Wahlkampfzeit entsprochen werden und eine diesbezüglich systematische Evaluation des Ansatzes, seiner Güte und der dabei wirksamen Einflussfaktoren im Zeitablauf erfolgen. In diesem Rahmen wird deutlich werden, dass Sonntagsfragen – zumindest in der in Deutschland praktizierten Art und Weise – keine wissenschaftlichen Prognosen darstellen. Zum anderen zeigt sich jedoch auch, dass diese insbesondere pragmatischen Überlegungen folgende Art der Wahlvoraussage im empirischen Mittel eine durchaus beachtliche Vorhersagegüte erreicht.

Zweitens wird ein Modell zur Vorhersage von Stimmenanteilen relevanter Parteien für Bundestagswahlen entwickelt, welches die aus wissenschaftstheoretischer Perspektive relevanten Bedingungen an eine wissenschaftliche Prognose möglichst erfüllt. Ziel dabei ist es, auf Grundlage fundamentaler Theorien zum Wählerverhalten und statistischer Gesetzmäßigkeiten einen Prognoseansatz zu entwickeln, welcher nicht nur aus wissenschaftstheoretischer Perspektive als Prognose bezeichnet werden kann, sondern sich auch hinsichtlich seiner Prognosegüte mit den dominierenden Sonntagsfragen als zumindest konkurrenzfähig erweist. Einschränkend muss erwähnt werden, dass im Rahmen dieser Arbeit die Konkurrenzfähigkeit des präsentierten Ansatzes nur post hoc anhand so genannter *out-of-sample*-Prognosen geprüft werden kann – eine Feuerprobe des Ansatzes mit einer „wahren“ Prognose einer zukünftigen Wahl steht noch aus.

Zunächst wird knapp der Prognosebegriff erörtert (Abschnitt 1.2). Hierbei werden ausgehend von wissenschaftstheoretischen und methodischen Überlegungen Anforderungen an eine wissenschaftliche Prognose erarbeitet. Diese

werden im Weiteren zur Bewertung der verschiedenen Ansätze zur Vorhersage von Bundestagswahlen herangezogen.

Abschnitt 2 nimmt verschiedene Wahlprognoseansätze in den Blick. Im Einzelnen werden umfragebasierte Voraussagen (Abschnitt 2.1), in jüngster Zeit an Bedeutung gewinnende Wahlbörsen (Abschnitt 2.2) und schließlich Prognosemodelle auf Basis theoretischer Überlegungen zum Wählerverhalten (Abschnitt 2.3) diskutiert. Hierbei werden jeweils zentrale Merkmale und gegebenenfalls Ausdifferenzierungen der Ansätze dargestellt. Vor dem Hintergrund der entwickelten Kriterien an wissenschaftliche Prognosen werden die Ansätze hinsichtlich ihrer Erfüllung dieser Kriterien bewertet. Abgerundet wird die Darstellung jeweils durch die Präsentation vorhandener empirischer Befunde zu deren Prognosegüte. Auch wenn die Praxis in Deutschland im Fokus des Interesses steht, wird zu allen Ansätzen auch der internationale Forschungsstand in den Blick genommen, um einerseits die bestehenden Forschungslücken aufzuzeigen und um andererseits grundlegende bestehende Erkenntnisse nicht zu vernachlässigen.

Eine wesentliche Erkenntnis stellt dabei die Tatsache dar, dass eine systematische Analyse der prognostischen Qualität von Sonntagsfragen sowie eine befriedigende Identifikation von relevanten Einflussfaktoren für Deutschland bisher nicht vorgelegt wurden. Deshalb werden in Abschnitt 3 zunächst anhand theoretischer und methodischer Überlegungen potenzielle Einflussfaktoren auf die Prognosegüte von Sonntagsfragen identifiziert (Abschnitt 3.1). Hieran schließt eine Zusammenfassung der vorliegenden empirischen Befunde zur Wirkung dieser Einflussfaktoren an (Abschnitt 3.2). Vor der eigentlichen Analyse ist noch darauf einzugehen, wie die Prognosegüte zu messen ist und welche Daten der folgenden Analyse zugrunde liegen (Abschnitt 3.3). Schließlich werden die empirischen Befunde präsentiert. Hierbei werden einleitend differenzierte deskriptive Ergebnisse präsentiert, welche später zur Beurteilung der Brauchbarkeit des entwickelten Modells dienen. Weiterhin werden multivariate Resultate diskutiert, die Einflussfaktoren auf die Prognosegüte von Sonntagsfragen auch im Zeitablauf offen legen (Abschnitt 3.4).

Daran anschließend wird auf Grundlage von theoretischen Ansätzen zur Erklärung des Wählerverhaltens und einem Modell zur Wirkung des Zufalls ein erster Vorschlag für ein Prognosemodell zur Vorhersage von Parteistimmenanteilen bei Bundestagswahlen entwickelt und – zumindest post hoc – dessen empirische Brauchbarkeit geprüft (Abschnitt 4). Hierzu werden zunächst wesentliche

theoretische Einsichten der empirischen Wahlforschung aufgegriffen (Abschnitt 4.1). Dabei werden jedoch nur zentrale Aspekte erörtert, welche auch als empirisch bestätigt gelten können – eine umfassende und ausdifferenzierte Darstellung der theoretischen Ansätze zur Erklärung des Wählerverhaltens kann und soll im Rahmen dieser Arbeit nicht geleistet werden (vgl. hierzu bspw. die Aufsätze in Falter/Schoen 2005). Als weitere theoretische Komponente des Modells wird der erwartete Stimmenanteil von Parteien bei einer gegebenen Anzahl von antretenden Parteien nach Mosteller (1965: 63ff.) herangezogen. Das hieraus entwickelte Modell zur Prognose von Stimmenanteilen für einzelne Parteien bei Bundestagswahlen wird auf seine empirische Brauchbarkeit anhand von *out-of-sample*-Prognosen für vergangene Bundestagswahlen getestet (Abschnitt 4.2). Zentrale Vergleichsgröße zur Bewertung der Leistungsfähigkeit des Modells stellen die Ergebnisse zur Prognosegüte von Wahlabsichtsfragen dar.

Abschließend erfolgt eine Zusammenfassung und Diskussion der zentralen Befunde (Abschnitt 5.1). Zudem werden Entwicklungsmöglichkeiten des vorgestellten Modells aufgezeigt und Anregungen für die weitere Beschäftigung mit der Prognose sozialer Phänomene geliefert (Abschnitt 5.2).

1.2 Zum Begriff der Prognose

Die Beschreibung, Erklärung und Prognose sozialer Phänomene stellen zentrale Ziele wissenschaftlichen Arbeitens dar. Beschreibungen gehen dabei zumeist Erklärungen voraus. Weiterhin wird in den Sozialwissenschaften häufiger das Verstehen sozialer Phänomene als eigenständiges Ziel wissenschaftlicher Betätigung angesehen (Toulmin 1968). Umstritten ist jedoch, ob Verstehen als eigene Kategorie wissenschaftlicher Zielsetzungen betrachtet werden soll (Bunge 1996: 150ff.; Opp 2005: 66ff.). Da im Weiteren wissenschaftliche Prognosen im Mittelpunkt stehen, wird diese wissenschaftstheoretische Auseinandersetzung nicht aufgegriffen.

Im Folgenden werden Anforderungen an wissenschaftliche Prognosen benannt. Eine derartige Festlegung von Kriterien ist notwendig, um wissenschaftliche von nicht-wissenschaftlichen Prognosen zu unterscheiden, um in Folge vorhandene Prognoseansätze bewerten zu können und auch einen eigenen wissenschaftlichen Prognoseansatz entwickeln zu können. Lewis-Beck (2005: 145)

sieht nicht-wissenschaftliche Prognosen dadurch gekennzeichnet, dass diese „guesses based on hunches, intuition, stargazing, casual conversation, nonsystematic interviews, insider talk or coincidence“ darstellen. Zwar können nicht-wissenschaftliche Prognosen durchaus über einen gewissen Zeitraum empirisch zutreffen, jedoch ist über einen längeren Zeitraum zu erwarten, dass nur wissenschaftliche Prognosen, welche auf empirisch bestätigte und korrekte kausale Mechanismen verweisen, auch substanzielle Vorhersagen liefern können. Zudem stellen nur wissenschaftliche Prognosen die intersubjektive Nachprüfbarkeit sicher und können somit neben ihrem praxeologischen Wert wesentliche Hinweise zur Verbesserung theoretischer Ansätze liefern und damit zum angestrebten kumulativen Wissensfortschritt beitragen.

Für wissenschaftliche Prognosen gilt, dass sie logisch mit Erklärungen verknüpft sind, weshalb zunächst zumindest knapp auf die grundlegende Struktur wissenschaftlicher Erklärungen eingegangen werden muss. In den Sozialwissenschaften besteht keine Einigkeit hinsichtlich der Ausgestaltung der Struktur einer wissenschaftlichen Erklärung. Vielmehr wird hierüber ein anhaltender wissenschaftstheoretischer Disput geführt (vgl. etwa Kühn 1970; Salmon 1990). Auch dieser Disput wird im Folgenden nicht aufgegriffen, sondern es wird lediglich das hier zugrunde gelegte Verständnis von wissenschaftlichen Erklärungen begründet und nachgezeichnet.

Eine *Erklärung* gibt die Antwort auf die Frage „Warum ist etwas der Fall?“. Damit wird impliziert, dass Erklärungen kausale Aussagen liefern. Diese erlauben in der Folge Vorhersagen. Ein brauchbares Verständnis von Erklärung muss ein zu erklärendes soziales Phänomen auf eine Ursache zurückführen. Von einer Ursache ist Bunge (1983: 136f.) zufolge dann zu sprechen, wenn alle notwendigen und hinreichenden Bedingungen für das Eintreten eines Ereignisses angeführt werden. Üblicherweise wird als Ausgangspunkt für eine Erklärung, welche dieser Auffassung entspricht, auf das von Hempel und Oppenheim (1948) vorgeschlagene deduktiv-nomologische Schema rekuriert.

Hempel und Oppenheim (1948) fordern für eine Erklärung die logisch korrekte Ableitung eines in einem singulären Satz ausgedrückten zu erklärenden Ereignisses (Explanandum) aus einer nomologischen Hypothese und den Antezedensbedingungen (zusammen das Explanans). Weiterhin formulieren Hempel und Oppenheim (ebd.: 137f.) logische und empirische Bedingungen für die Adäquanz von wissenschaftlichen Erklärungen. Akzeptabel sind demnach Erklärungen

nur, wenn folgende Anforderungen erfüllt werden (vgl. dazu auch Lenk 1972: 17; Stegmüller 1983: 124ff.):

- Der Schluss vom Explanans auf das Explanandum muss logisch korrekt erfolgen.
- Das Explanans muss zumindest eine allgemeine Gesetzesaussage enthalten, die benötigt wird, um das Explanandum herzuleiten, und die Antezedensbedingungen müssen zumindest einen singulären Satz umfassen, der die konkreten Sachverhalte beschreibt.
- Das Explanans muss empirischen Gehalt aufweisen, also grundsätzlich durch Beobachtung überprüfbar sein.
- Die Sätze des Explanans müssen sich zumindest gut bewährt haben.

Wesentlich für die deduktiv-nomologische Erklärung ist die deterministische Beziehung in der nomologischen Hypothese, die eine logisch korrekte Deduktion des Explanandums aus dem Explanans erlaubt. In den Sozialwissenschaften sind jedoch derartige deterministische Gesetze kaum bekannt (vgl. Opp 2005: 52), weshalb in der Disziplin häufig auf induktiv-statistische Erklärungen rekurriert wird. Deren Grundstruktur folgt dem deduktiv-nomologischen Schema, fordert allerdings nur mehr probabilistische Gesetzesaussagen im Explanans. Daraus folgt, dass „the explanans confers upon the explanandum a more or less high degree of inductive support or of logical (inductive) probability“ (Hempel 1965: 385).

Aus wissenschaftstheoretischer Sicht ist diese Abschwächung der Anforderungen problematisch, da eine Deduktion des Explanandums aus einer probabilistischen Gesetzesaussage und den Antezedensbedingungen nicht mehr möglich ist (vgl. dazu bspw. Lenk 1972: 21ff.). Vielmehr besteht nur eine „Bestätigungsrelation“ (Opp 2005: 56) zwischen Explanans und Explanandum. Dennoch bleibt die Logik des Erklärungsschemas auch bei induktiv-statistischen Erklärungen bestehen, bedingt jedoch einige zusätzliche Unsicherheiten (detailliert bei Hempel 1965: 376ff.; Stegmüller 1983: 777ff.).

Das deduktiv-nomologische Schema liefert lediglich die logische Struktur für gültige Erklärungen. Auch bei Erfüllung der genannten Adäquatheitsbedingungen resultieren hieraus keineswegs zwingend befriedigende Erklärungen. Dies liegt daran, dass auch bei Einhaltung der deduktiven Struktur auf falsche kausale Aussagen verwiesen werden kann (vgl. bspw. Braun 2008: 379f.; Woodward 2003: 154f.). Deshalb kann zwar das deduktiv-nomologische Erklärungsschema als

Idealbild für die logische Struktur einer Erklärung dienen, muss jedoch ergänzt werden um die Anforderung kausal korrekter Aussagen (Bunge 1997: 443).

Eine Möglichkeit diese Lücke zu schließen ist der Rückgriff auf mechanismus-basierte Erklärungen (Bunge 1997, 2004; Hedström 2005; Hedström/Swedberg 1998; Mayntz 2004; Schmid 2006).¹ Unter einem kausalen Mechanismus wird ein Prozess in einem System verstanden, welcher Wandel in dem gesamten System oder auch Teilsystemen bedingt oder verhindert (Bunge 1997: 414). Wesentlich hierbei ist weiterhin, dass der identifizierte Mechanismus regelmäßig zu dem zu erklärenden sozialen Phänomen führt (Braun 2008: 380). Bunge betont überdies, dass ein Mechanismus nicht kausal, das heißt deterministisch im engeren Sinne sein muss, sondern vielmehr häufig aus zufälligen Prozessen entsteht. Ebenso verweist er darauf, dass Mechanismen nicht universal sein können, sondern immer gegenstandsabhängig und systemspezifisch sind (Bunge 2004: 195).

Damit stellt sich auch die Frage nach dem Verhältnis von Mechanismus und Gesetz. Gesetze sind Hempel und Oppenheim (1948) zufolge notwendig für gültige Erklärungen. In der wissenschaftstheoretischen Diskussion wird jedoch sowohl die Existenz solcher Gesetze bezweifelt als auch deren Notwendigkeit zur Erklärung sozialer Phänomene. Elster (1998, 2007: 44ff.) betrachtet Gesetze als Gegenteil von Mechanismen, negiert in der Folge die Möglichkeit von Gesetzen in den Sozialwissenschaften und sieht diese ersetzt durch Mechanismen. Bunge (2004) hingegen betont die Interdependenz von Mechanismen und Gesetzen. Ihm zufolge ist in den Wissenschaften zunächst von der Existenz einer Realität (z. B. natürliche Ereignisse, soziale Prozesse) auszugehen (Prämisse des Materialismus). Weiterhin ist für die Wissenschaft entscheidend von der Existenz von Zusammenhängen in der Realität auszugehen, welche die Menschen

1 Gerade vor dem Hintergrund dieser Schwäche des deduktiv-nomologischen Erklärungsschemas sind weitere Überlegungen zur Kausalität zwingend. Einen Überblick über diese Diskussion liefert Salmon (1990). Woodward (2003) hat eine Theorie der Kausalität vorgelegt, welche unter anderem auf der Arbeit von Pearl (2000) aufbaut. Hier wird grundlegend der Begriff der Kausalität diskutiert und eine Verbindung von statistischer Inferenz und Wissenschaftstheorie gesucht, um etwa die Frage zu beantworten, welche empirische Evidenz notwendig ist, um die Inferenz eines Ursache-Wirkungs-Zusammenhangs zu akzeptieren. Eine aktuelle Erörterung des Verhältnisses von Kausalität und Wahrscheinlichkeit in anderen Wissenschaftsdisziplinen liefern die Aufsätze in Russo und Williamson (2007), während Morgan und Winship (2007) speziell die Aufdeckung kausaler Beziehungen in den Sozialwissenschaften vor dem Hintergrund verschiedener Methoden diskutieren.

entdecken können. Dies bedingt die Prämisse, dass alle Ereignisse Gesetzen folgen (vgl. Bunge 2004: 197).

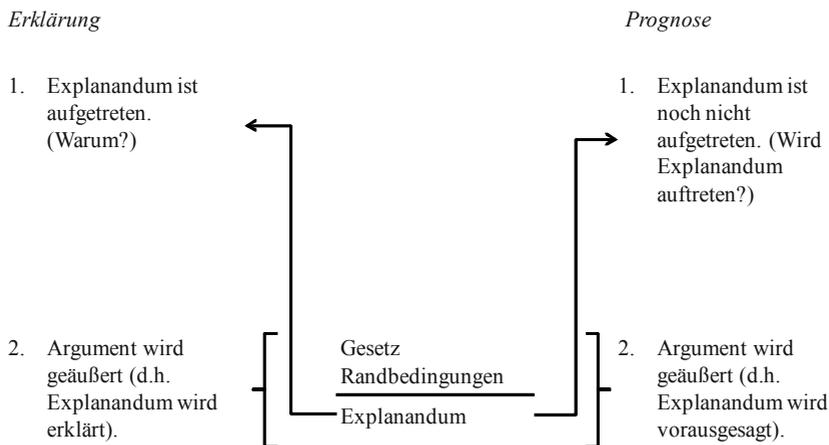
Entscheidend ist also, dass für eine vollständige Erklärung im Idealfall der logischen Struktur des deduktiv-nomologischen Schemas gefolgt wird, jedoch zusätzlich auf einen kausalen Mechanismus im hier definierten Sinn verwiesen werden muss. Denn nur die Spezifizierung eines korrekten kausalen Zusammenhangs erlaubt die Beantwortung von „Warum“-Fragen (Braun 2008). Prognosen wiederum sind ebenfalls nur auf Basis bekannter kausaler Zusammenhänge möglich.

Hempel (1965: 234) zufolge bestehen *wissenschaftliche Prognosen* darin, eine Aussage über ein zukünftiges Ereignis abzuleiten aus „(1) statements describing certain known (past or present) conditions (...) , and (2) suitable general laws (...)“. Auch Bunge (1959: 307) definiert wissenschaftliche Prognosen in dieser Art und Weise: „Scientific prediction may be defined as the deduction of propositions concerning as yet unknown or unexperienced facts, on the basis of general laws and of items of specific information“. Wesentlich für Prognosen ist demnach zunächst, dass diese dieselbe Struktur wie Erklärungen aufweisen – Hempel (1965: 367) spricht in diesem Zusammenhang von der „*thesis of the structural identity*“ (Hervorhebung im Original).

Die Identität zwischen Erklärung und Prognose bezieht sich auf deren logische Struktur, nicht auf deren „Natur“ (Bunge 1959: 307). Der Unterschied zwischen Erklärungen und Prognose liegt zunächst darin, dass im Rahmen einer Prognose das Explanans bekannt und das Explanandum gesucht wird, während eine Erklärung durch die Suche nach dem Explanans bei einem gegebenen Explanandum gekennzeichnet ist (vgl. Abbildung 1.1). Hempel (1977: 40) betont in diesem Zusammenhang, dass an Prognosen zusätzliche Bedingungen geknüpft sind, weshalb deduktiv-nomologische Erklärungen stets nur „potentielle Vorhersagen“ sein können. Schließlich sind Prognosen mit größeren Unsicherheiten verbunden als Erklärungen – und dies gilt insbesondere für Prognosen in den Sozialwissenschaften. So hat Merton (1936) darauf hingewiesen, dass bei sozialen Prozessen nicht ausgeschlossen werden kann, dass die Äußerung der Prognose selbst Einfluss auf das vorherzusagende Ereignis nimmt. Denkbar ist hier sowohl die Form der sich selbst erfüllenden als auch der sich selbst widerlegenden Prognose. In Zusammenhang mit diesem Problem ist auch die häufige Unsicherheit der zeitlichen Abfolge von Ursache und Wirkung zu sehen (vgl. auch Braun 2008: 378).

Bei Prognosen muss sichergestellt sein, dass das Auftreten des Explanandums zeitlich der Äußerung der Vorhersage, also dem Vorliegen der Antezedensbedingung nachgelagert ist. Auch wenn diese Forderung trivial erscheint, ist dennoch bei der Bewertung der Ansätze, mit welchen Wahlergebnisse vorhergesagt werden, zu prüfen, ob diese Bedingung erfüllt wird, zumal die Zeitpunkte, zu denen entsprechende Voraussagen geäußert werden, stark variieren.

Abbildung 1.1: Strukturelle Identität von Erklärung und Prognose



Quelle: Darstellung nach Opp (2005: 77).

Weiterhin ist wesentlich, dass alle Bedingungen spezifiziert werden, die für das Auftreten des Explanandums relevant sind, und auch die Wirkungen der Beziehungen zwischen diesen Bedingungen korrekt benannt werden. Diese Anforderung bedingt, dass Prognosen auf korrekte kausale Mechanismen verweisen, welche empirisch als bestätigt gelten können. Dies ist gerade im Kontext der Vorhersage von Wahlergebnissen zentral, denn es existieren einige prominente Beispiele, die zwar auf „Gesetzmäßigkeiten“ verweisen, welche jedoch keine kausalen Beziehungen spezifizieren. Zu denken ist hier etwa an die Ableitung des zukünftigen US-Präsidenten aus dem Gewinner des Superbowls oder dem Postulat bestimmter für ein Elektorat repräsentativer Stimmbezirke (vgl. Lewis-Beck/Rice 1992: 3ff.).

Prognosen müssen den postulierten zeitlichen Abstand bis zum Eintreten des Explanandums korrekt angeben. In Bezug auf die Vorhersage von Wahlresultaten stellt diese Anforderung keine zentrale Herausforderung dar, da die Wahltermine in der Regel frühzeitig festgelegt werden und somit exakt angegeben werden können. Es gibt jedoch auch Ausnahmen von der Regel. So wurde beispielsweise am 21. Juli 2005 der 15. Deutsche Bundestag von Bundespräsident Horst Köhler vorzeitig aufgelöst und Neuwahlen für den 18. September 2005 angesetzt – die 15. Legislaturperiode hätte eigentlich ein Jahr länger gedauert. In diesem Fall wäre es möglich, dass vor dem 21. Juli 2005 erstellte Prognosen für den Ausgang der nächsten Bundestagswahl den Zeitpunkt des Eintritts des Explanandums zeitlich falsch spezifiziert hätten.

Auch in Bezug auf die Anfangsbedingungen sind Anforderungen zu nennen, die für den Erfolg von Prognosen entscheidend sind. Zunächst ist in diesem Kontext die richtige Erhebung der Anfangsbedingungen zu nennen. Hinsichtlich der Vorhersage von Wahlresultaten ist auch dies durchaus kritisch zu prüfen. Wie bereits erläutert stützen sich die meisten Voraussagen auf Umfrageergebnisse – unterstellt werden muss demnach, dass die jeweiligen Variablen, etwa die Wahlabsicht, korrekt erfasst wurde. Gerade vor dem Hintergrund möglicherweise systematischer Non-Response oder verzerrend wirkendem Antwortverhalten aufgrund sozialer Erwünschtheit ist diese Anforderung keineswegs zu vernachlässigen.

Zu vermuten ist schließlich, etwa in Anschluss an Mertons (1936) Beitrag zu sich selbst erfüllenden und widerlegenden Prophezeiungen, dass sich nach der Äußerung einer Prognose, jedoch vor Eintritt des Explanandums, relevante Antezedensbedingungen verändern. Eine erfolgreiche Prognose muss dies erfassen und entsprechend berücksichtigen. Nicht zuletzt diese Anforderung ist umstritten – und zwar nicht in Bezug auf die logische Notwendigkeit ihrer Erfüllung im Rahmen der These der strukturellen Identität von Erklärung und Prognose als vielmehr als genereller Angriffspunkt gegen die Identitätsthese.

So wendet Wild (1970) ein, dass die Symmetrie von Erklärung und Prognose nicht aufrecht erhalten werden könne aufgrund bestehender Unsicherheitsmomente im Rahmen von Prognosen, welche sich unmittelbar aus dieser Bedingung ergeben. Von Knorring (1972) wendet hiergegen ein, dass entsprechende „Unsicherheitsmomente“ (Wild 1970: 559) auch bei Erklärungen bestehen, da es sich bei Randbedingungen um singuläre Sätze handelt, welche nach Popper

(1989: 20f.) intersubjektiv nachprüfbar sein müssen. Dies bedingt wiederum, dass eine Verifizierung der Randbedingungen nicht möglich ist, sondern nur deren Falsifizierung. Demnach unterliegen also auch Erklärungen zwangsläufig einer gewissen Vorläufigkeit und unterscheiden sich in dieser Hinsicht nicht von Prognosen. Von Knorring (1972: 148) betont in diesem Zusammenhang:

„Es ist jedoch denjenigen unsicheren Voraussagen den Vorzug zu geben, deren Unsicherheit lediglich in einer potentiellen Falsifizierung der zugrunde liegenden singulären und universellen Sätze begründet liegt, deren Ursachen also aufgrund bislang bestätigter oder nicht falsifizierter Hypothesen und ihrer Anwendungsbedingungen angegeben werden können und damit Entscheidungshilfe bei der Verhinderung unerwünschter zukünftiger Ereignisse leisten.“

Dennoch erscheint die von Wild (1970) problematisierte Bedingung erfolgreicher Prognosen bedenkenswert, denn gerade die Sozialwissenschaften verfügen über keine derart exakten Theorien, die notwendig wären, um neben einer korrekten Vorhersage des Explanandums zusätzlich die Veränderung *aller* relevanten Anfangsbedingungen *korrekt* zu erfassen. Entsprechend ist davon auszugehen, dass die Erstellung dieser so genannten unbedingten Prognosen zumindest in den Sozialwissenschaften nicht möglich ist.

Vielmehr werden in der Regel Prognosen an bestimmte Bedingungen hinsichtlich der Veränderung von Anfangsbedingungen zwischen Prognosezeitpunkt und Eintritt des Explanandums geknüpft. Üblicherweise wird dabei vereinfachend unterstellt, dass die relevanten Anfangsbedingungen konstant bleiben. Entsprechende Vorhersagen werden konditionale Prognosen genannt (von Knorring 1972: 147).²

Zusammenfassend ist für die Anforderungen an wissenschaftliche Prognosen damit Folgendes festzuhalten. Prognosen sind Aussagen über zukünftige Ereignisse, welche idealerweise der logischen Struktur des deduktiv-nomologischen Schemas folgen und auf einen kausalen Mechanismus verweisen. Weiterhin muss gewährleistet sein, dass die Prognose vor dem Eintritt des Explanandums geäußert wird, der Eintrittszeitpunkt des Explanandums korrekt angegeben wird und schließlich die Antezedensbedingungen korrekt erhoben werden.

2 Mit diesen Einwänden gegen die Identitätsthese wurde die wissenschaftstheoretische Diskussion hierüber nur gestreift. Für eingehendere Diskussionen vgl. Hempel (1977: 43ff.), Kühn (1970: 63ff.), Lenk (1972: 30ff.), Popper (1980: 26) oder Stegmüller (1983: Kap. I).

2 Ansätze zur Prognose von Wahlergebnissen

Entsprechend der herausragenden Bedeutung von Wahlen im politischen Prozess von Demokratien und dem öffentlichen Interesse an ihnen haben sich verschiedene Methoden zur Voraussage von Wahlergebnissen etabliert. Diese werden im Folgenden entlang einer Klassifikation von Lewis-Beck (2005) dargestellt, vor dem Hintergrund der Anforderungen an eine wissenschaftliche Prognose (vgl. Abschnitt 1.2) bewertet und anhand des Forschungsstandes bezüglich ihrer empirischen Prognosegüte diskutiert. In diesem Kontext ist zu betonen, dass die methodologische Bewertung der Ansätze nicht zwingend in einem engen gleichsinnigen Zusammenhang mit der empirischen Prognosegüte stehen muss. Vielmehr ist durchaus denkbar, dass auch Ansätze, welche im Folgenden nicht als wissenschaftliche Prognosen qualifiziert werden, bisher eine hohe Prognosequalität aufweisen. Auch Lewis-Beck und Rice (1992: 1ff.) verweisen auf diese Möglichkeit, indem sie Beispiele von populären Daumenregeln nennen, welche auf scheinbare Gesetzmäßigkeiten verweisen, jedoch keine korrekten kausalen Mechanismen benennen. Die Prognoseansätze müssen daher sowohl methodologischen als auch empirischen Ansprüchen gerecht werden.

Im Hinblick auf die diskutierten Ansätze werden zunächst „auf Umfragen basierte Vorhersagen“ (Fessel 1969: 176) aufgegriffen. Aufgrund ihrer Dominanz im Rahmen der medialen Berichterstattung wird auf entsprechende Ansätze besonders ausführlich eingegangen (Abschnitt 2.1). In jüngerer Zeit gewinnen Wahlbörsen auch in Deutschland als Prognoseinstrument zunehmend an Bedeutung. Herausfordernd für die umfragebasierten Vorhersagen sind Wahlbörsen insbesondere aufgrund der zumeist angeführten Überlegenheit im Hinblick auf die empirische Prognosegüte. In Abschnitt 2.2 werden sowohl die Funktionsweise als auch die empirische Evidenz zu Wahlbörsen erörtert. Schließlich werden theoretische Prognosemodelle, wie sie insbesondere in den USA verbreitet sind, in Betracht gezogen. Neben einer Darstellung wesentlicher Konstruktionsmerkmale

dieser Modelle wird detailliert auf das bisher für Deutschland vorgelegte Modell eingegangen (Abschnitt 2.3).

Schließlich muss darauf hingewiesen werden, dass als weiterer – hier nicht berücksichtigter – Ansatz Experteneinschätzungen angeführt werden können. Zwar werden in der Vorwahlberichterstattung immer wieder Experten, etwa Politikwissenschaftler oder Politiker selbst, zu ihren Einschätzungen des Wahlausgangs interviewt und medial hierüber berichtet, jedoch werden in Deutschland Expertenmeinungen nicht systematisch erhoben oder dokumentiert, so dass eine empirische Analyse ihrer Vorhersagen nicht möglich erscheint. Gleichwohl ist unmittelbar einsichtig, dass Voraussagen durch einzelne Experten keine wissenschaftlichen Prognosen darstellen, da sie auf der subjektiven Einschätzung Einzelner beruhen und sich nicht auf gesetzesartige Aussagen zurückführen lassen, sondern vielmehr von deren Intuition geleitet werden. In den USA allerdings werden systematische und standardisierte Expertenbefragungen durchgeführt, welche beispielsweise für das Vorhersageverfahren *PollyVote* genutzt werden (Jones et al. 2007). Auf *PollyVote* wird im Rahmen der Darstellung theoretischer Prognosemodelle eingegangen.

2.1 Umfragebasierte Vorhersagen

Zur Vorhersage von Wahlergebnissen kann man verschiedene umfragebasierte Ansätze unterscheiden, die in diesem Abschnitt thematisiert werden. Zu differenzieren sind im Rahmen von Vorwahlumfragen die erhobene Wahlabsicht (Sonntagsfrage) und die Wahlerwartungen sowie Wahltagsbefragungen, in denen unmittelbar nach der Wahl das tatsächliche Wahlverhalten erfasst wird (so genannte *exit polls*).

In diesem Abschnitt wird zunächst knapp die Entwicklung der Meinungsforschung³ und der empirischen Wahlforschung in Deutschland sowie die Kritik hieran nachgezeichnet, da dieser Kontext wesentlich für das Verständnis der Relevanz und Praxis von umfragebasierten Vorhersagen von Wahlergebnissen in

3 Die Begriffe Meinungsforschung und Demoskopie werden im Rahmen dieser Arbeit synonym verwendet. Die Adäquanz des letzteren Begriffs ist jedoch umstritten: Während das Institut für Demoskopie Allensbach die von Dodd (1946: 473) eingeführte Bezeichnung Demoskopie bevorzugt (Petersen 2008), lehnt Eberlein (2001: 23f.) diese strikt ab.

Deutschland ist (Abschnitte 2.1.1 sowie 2.1.2). Die weiteren Schritte umfassen die Erörterung der Praxis der Wahlabsichtsfragen (Abschnitt 2.1.3), der Wahlerwartungsfragen (Abschnitt 2.1.4) sowie von Wahltagsbefragungen, welche auch als *exit polls* bezeichnet werden (Abschnitt 2.1.5). Hierzu ist anzumerken, dass der Schwerpunkt der Darstellung auf Wahlabsichtsfragen liegt, da diese Form der Prognose die in Deutschland mit großem Abstand dominierende Form der umfragebasierten Vorhersagen ist. In Abschnitt 2.1.6 werden die umfragebasierten Vorhersagen vor wissenschaftstheoretischem Hintergrund bewertet und abschließend der Forschungsstand zu deren empirischer Vorhersagequalität in Blick genommen (Abschnitt 2.1.7).

2.1.1 Meinungsforschung und empirische Wahlforschung in Deutschland

Die Entwicklung der kommerziellen Umfrageforschung in Deutschland nach dem Zweiten Weltkrieg ist eng verknüpft mit wahlsoziologischen Fragestellungen und muss zudem im Zusammenhang mit der Entwicklung der empirischen Sozialforschung im Allgemeinen sowie der empirischen Wahlforschung im Speziellen gesehen werden. Diese bis heute besondere Verzahnung von zumindest Teilen der kommerziellen mit der wissenschaftlichen Wahlforschung und auch der Politik ist wesentlich für die heutige Praxis der Sonntagsfrage.

Ein Beispiel für diese engen Verbindungen stellt die Forschungsgruppe Wahlen e. V. dar, welche 1974 als *non-profit*-Organisation von Mitarbeitern der Fakultät für Sozialwissenschaften der Universität Mannheim aus dem Wahlforschungsteam von Rudolf Wildenmann für die (Vor-) Wahlberichterstattung des ZDF gegründet wurde und diese bis heute gestaltet (Forschungsgruppe Wahlen 2008b). Weiterhin zu denken ist an das lange von Elisabeth Noelle-Neumann geführte Institut für Demoskopie Allensbach. Von 1964 bis zu ihrer Emeritierung 1983 hatte Noelle-Neumann einen Lehrstuhl für Publizistik an der Universität Mainz inne und war zudem seit den 1950er Jahren in Bezug auf die politische Meinungsforschung Beraterin erst von Konrad Adenauer, später von Helmut Kohl. Wildenmann (1992: 61) spricht im Zusammenhang der Darstellung der Gründungsphase der Meinungsforschungsinstitute entsprechend von einem „wissenschaftlich-politisch-kommerzielle[n] Komplex“.

Als entscheidend für die Durchsetzung (moderner) demoskopischer Methoden wird die erfolgreiche Vorhersage des Präsidentschaftswahlgewinners in den USA des Jahres 1936 durch das von George H. Gallup 1934 gegründete *American Institute of Public Opinion* gesehen (Gallus 2002; Kaase 1999a; Keller 2001: 31ff.; Noelle-Neumann 1988, 2000; Noelle-Neumann/Petersen 2005: 209; Scheuch 1956: 52ff.; 1990). Bei dieser Voraussage stützte sich Gallups Institut auf Wahlabsichtsfragen auf Grundlage einer Quotenstichprobe, womit sich die standardisierte Befragung von nach einem systematischen Verfahren ausgewählten Probanden durchsetzte.

Gleichwohl handelt es sich hierbei nicht um die Geburtsstunde der modernen Meinungsforschung, da bereits 1912 in den USA erste Umfragen auf Grundlage von Zufallsstichproben mit standardisierten Fragebögen durchgeführt wurden. Auch Elmo Ropers Meinungsforschungsinstitut veröffentlichte bereits 1935 Resultate von Umfragen (Eberlein 2001: 50ff.). Die Geschichte der Umfrageforschung weist zeitlich noch deutlich weiter zurück. So ist als erstes Dokument der Vorgeschichte der Umfrageforschung ein Fragebogen von Karl dem Großen bekannt, welcher im Jahre 811 offenbar die Stimmung seiner Untertanen erfassen wollte (vgl. Petersen et al. 2004). Noch weiter in die Vergangenheit reichen Eberlein (2001: 37ff.) zufolge die Wurzeln der Meinungsforschung. Ihm zufolge sind bereits in der Bibel Hinweise auf den Einsatz von frühen Formen der Erfassung der Volksmeinung zu finden.

Auch das Zeitalter von Wahlvorhersagen begann bereits lange vor der Präsidentschaftswahl von 1936. Vorhersagen der Gewinner erfreuten sich in den USA bereits zuvor großer Beliebtheit. Diese beruhten häufig auf so genannten *straw polls*, Probeabstimmungen,⁴ die vor allem von großen Zeitschriften initiiert wurden, die entsprechende „Abstimmungsaufrufe“ an ihre Leser versandten und die zurückgeschickten „Stimmzettel“ auszählten und als Vorhersage veröffentlichten (vgl. Smith 1990). Die Zeitschrift *Literary Digest* setzte auf solche *straw polls* und versuchte, mit groß angelegten Probeabstimmungen neue Leser zu gewinnen, indem sie ihre „Stimmzettel“ nicht nur an eigene Leser, sondern auch etwa an alle Telefon- oder Autobesitzer verschickte. Durch die immense Zahl der verschickten und ausgewerteten *straw polls* durch *Literary Digest* – vor der

4 Im Prinzip sind *straw polls* ein Pendant zu den noch heute beliebten TED-Umfragen beispielsweise im Rahmen von Fernsehsendungen ohne kontrollierte Stichprobenziehung.

Wahl 1936 wurden zehn Millionen Personen angeschrieben und es wurden über zwei Millionen „Stimmzettel“ ausgezählt – und die korrekte Vorhersage des Siegers Herbert Hoover bei der Präsidentschaftswahl 1928 galt sie als seriöse Institution mit besonders aussagekräftigen Vorhersagen. 1936 prophezeite die Zeitschrift dem Kandidaten der republikanischen Partei, Alfred M. Landon, einen klaren Wahlsieg (vgl. Keller 2001: 33ff.).

Herausgefordert wurde *Literary Digest* durch Gallups *American Institute of Public Opinion*, welches in der *Washington Post* eine Vorhersage des Wahlergebnisses auf Basis einer Befragung von mehreren tausend Personen publizierte, die den demokratischen Kandidaten Franklin D. Roosevelt als zukünftigen Präsidenten sah. Roosevelt konnte am Wahltag etwa zehn Millionen Stimmen mehr (*popular votes*)⁵ als sein Konkurrent Landon für sich gewinnen und *Literary Digest* lag mit seiner Vorhersage des Stimmenanteils von Roosevelt um fast 20 Prozentpunkte daneben.

Nicht nur dieser Erfolg des *American Institute of Public Opinion* begründet den rasanten Aufstieg und Bedeutungszuwachs von kommerziellen Meinungsumfragen in den USA und in allen westlichen Demokratien (vgl. Keller 2001: 36ff.; Worcester 1983). Gallup verstand es bereits im Vorfeld, durch geschickt über die *Washington Post* verbreitete Artikel die *straw polls* aufgrund der ihnen inhärenten selektiven Stichprobe zu diskreditieren und pries das von ihm verwendete Quotenverfahren als wissenschaftliche Methode zur Erfassung des politischen Meinungsbilds an, das nicht nur validere, sondern zudem schnellere und kostengünstigere Ergebnisse versprach. In der Folge wurden die Ergebnisse der Demoskopien nicht nur von Medien, sondern auch von Politikern und Parteien rege nachgefragt (vgl. Gallus/Lühe 1998; Kruke 2007: 61ff.; Raupp 2007: 121ff.).

Bereits die US-Präsidentschaftswahl von 1948 bescherte den Demoskopien eine herbe Niederlage, denn die größten Institute (Crossley, Gallup, Roper) sagten einheitlich einen Wahlsieg Thomas Dueweys über Harry Truman voraus, der bekanntlich nicht eintrat. Auch wenn dieses Scheitern medialen Wirbel auslöste (Eberlein 2001: 67ff.) und zur Einsetzung einer staatlichen Kommission zur

5 Der Präsident der Vereinigten Staaten wird von einem Wahlmännergremium gewählt, welche direkt gewählt werden. Die Bezeichnung *popular votes* bezieht sich auf die Stimmen, die von den Wählern den jeweiligen Parteien gegeben wurden (vgl. dazu bspw. Lewis-Beck/Rice 1992: 21ff.).

Erforschung der Ursachen der Fehlprognose und Entwicklung von Maßnahmen zur künftigen Verhinderung derselben führte (Moeller et al. 1949; Mosteller et al. 1949), konnte dies den rasanten Bedeutungszuwachs von politischen Meinungsumfragen nicht aufhalten. Der Bericht dieser Kommission stellt die erste Analyse von Einflussfaktoren auf die Genauigkeiten von Voraussagen auf Basis von Wahlabsichtsfragen dar, welche in Abschnitt 3 analysiert werden.

Als Gründe für das Weiterbestehen und den zunehmenden Erfolg der Demoskopie wird gemeinhin die grundsätzlich wissenschaftliche Fundierung der angewendeten Methoden und die Möglichkeiten diese weiterzuentwickeln angesehen (Keller 2001: 36ff.) sowie die schon bis dahin bestehende Institutionalisierung in vielen Staaten (Noelle-Neumann 1988). Kruke (2007: 35) hebt in diesem Kontext die Bedeutung der sich entwickelnden Medienwissenschaften hervor, welche durch ihr Interesse an Rundfunk-Hörerforschung wesentliche Impulse für die Verbreitung demoskopischer Methoden aussandte. Generell hat sich bis dahin die Meinungsforschung auf viele Bereiche ausgedehnt und der Bereich der politischen Meinungsforschung und insbesondere der Wahlprognosen stellte nur einen geringen Teil des Betätigungsfelds von Demoskopern dar. Für das anhaltende Vertrauen in die Meinungsforschung in diesem Bereich ist jedoch auch die Rolle der US-amerikanischen wissenschaftlichen Wahlforschung zu betonen.

Angus Campbell hatte mit seinen Kollegen auf Grundlage nationaler Zufallsstichproben für die Wahl 1948 den Sieger korrekt vorhergesagt (vgl. Kaase 1999a). Auch weitere für die Etablierung der empirischen Wahlforschung maßgebliche Studien auf Grundlage von Individualdaten wurden zu dieser Zeit vorgelegt. Zum einen sind dies die regional begrenzten Untersuchungen von Paul F. Lazarsfeld und seinen Kollegen an der *University of Columbia* (Berelson et al. 1986; Lazarsfeld et al. 1955) und die an der *University of Michigan* von Campbell und Kollegen (1954, 1964) durchgeführten Studien. Damit legten sie den Grundstein der Entwicklung der modernen wissenschaftlichen Wahlforschung als ebenso empirisch wie mikrotheoretisch ausgerichtete Disziplin. Dass sich letztlich die Forschergruppe in Ann Arbor nachhaltiger durchsetzte und auch institutionell das Zentrum der internationalen Wahlforschung werden konnte, führt Kaase (1999a) nicht zuletzt auf den prognostischen Erfolg 1948 zurück. In der Folge gelang Campbell und seinen Kollegen die Etablierung nationaler Wahlstudien am *Survey Research Center* der *University of Michigan*, was die bis

heute anhaltende Dominanz mikrotheoretisch fundierter Wahlforschung in den USA begründet.

In Deutschland wurde dieser Ansatz nicht von der sich in den 1950er Jahren unter US-amerikanischer Protektion aufgebauten Politikwissenschaft aufgegriffen, sondern zunächst von Umfrageinstituten (vgl. Klingemann 2002). Bereits 1945 etablierte das *Office of the Military Government United States* (OMGUS) eine spezielle Abteilung für Meinungsumfragen (*Opinion Surveys Section*), die im Laufe der Zeit stetig ausgebaut wurde und schließlich zur Gründung des Deutschen Instituts für Volksumfragen (DIVO) führte. Die US-Amerikaner beschäftigten von Beginn an deutsche Interviewer und forcierten die Etablierung von Meinungsumfragen auch durch die gezielte Schulung des deutschen Personals durch US-amerikanische Wissenschaftler.⁶ Das DIVO hat wesentlich zur Verbreitung methodischer Kompetenzen im Bereich der Meinungsforschung beigetragen.

Dies hat in der Folge auch zur Gründung privater Demoskopieinstitute geführt. Zu nennen sind hier unter anderen die Gründungen des Instituts für Demoskopie Allensbach (IfD) durch Elisabeth Noelle und Erich Peter Neumann sowie von Emnid durch Karl-Georg von Stackelberg.⁷ Das IfD Allensbach publizierte bereits zur ersten Bundestagswahl 1949 eine Prognose auf Basis einer – von Gallup inspirierten – Quotenstichprobe und sowohl das IfD Allensbach als auch Emnid begleiten seitdem die Wahlvor- und -nachberichterstattung in Deutschland. Diese frühe Adaption der politischen Meinungsforschung in Deutschland wurde auch seitens der Politik aufgenommen und bereits früh entstanden enge Verbindungen zwischen einzelnen Instituten und Parteien. So ließ sich die CDU bereits in den frühen 1950er Jahren vom IfD Allensbach beraten und Elisabeth Noelle-Neumann stieg unter Helmut Kohl zu einer seiner wichtigsten Beraterinnen auf. Die SPD zog nach Adenauers intensiver Nutzung der Dienste von Demoskopien 1961 nach und engagierte sich sogar direkt finanziell bei der Gründung des Instituts für angewandte Sozialwissenschaft (Infas) von Klaus Liepelt, welches auch bis in die 1990er Jahre die Wahlberichterstattung der ARD begleitete

6 Die Arbeiten des DIVO sind dokumentiert bei Merritt und Merritt (1970, 1979), Zeitzeugenberichte liefern Schaefer und Miller (1998) sowie Noelle-Neumann (1992).

7 Emnid wurde 1965 von Georg von Stackelberg 1965 verkauft und gehört seit 1997 zur Unternehmensgruppe Taylor Nelson Sofres (TNS) – wie im Übrigen auch Infratest dimap (TNS Emnid 2009).

(für detailliertere Darstellung der historischen Entwicklung vgl. Gallus 2002; Kruke 2007: 61ff.; Noelle-Neumann 1983).

Die deutsche Politikwissenschaft hingegen wurde nach dem Zweiten Weltkrieg als normative und moralische Staatswissenschaft aufgebaut (vgl. Bleek 2001: 265ff.), die gerade gegenüber der Umfrageforschung große Vorbehalte hatte. Diese Vorbehalte wurden zudem auch in anderen Disziplinen geteilt und selbst grundsätzlich der empirischen Sozialforschung offen gegenüber stehende Wissenschaftler äußerten ihre Bedenken. Diese wurden bereits 1951 im Rahmen einer gemeinsamen Tagung akademischer und privatwirtschaftlicher Sozialforscher in Weinheim deutlich (vgl. Institut für öffentliche Angelegenheiten 1952), im Rahmen derer Theodor W. Adorno (1952) zwar zum einen das demokratische Potenzial von Meinungsumfragen betonte, zugleich jedoch vor deren Missbrauch warnte. Da sich die Kritik an der Meinungsforschung häufig auf Wirkungen von aus Umfragen gewonnenen politischen Voraussagen bezieht, wird sie in Abschnitt 2.1.2 detaillierter aufgegriffen und erörtert.

Entsprechend dieser Ausrichtung der ersten politikwissenschaftlichen Lehrstühle in Deutschland konnte sich eine „umfragebasierte Mikroanalyse“ (Klingemann 2002: 51) zunächst auch trotz der schnellen Expansion des Faches nicht durchsetzen (Bleek 2001: 308ff.). So lassen sich größere empirische Forschungsansätze, welche makroanalytisch-geschichtlich geprägt waren, erst zu Ende der 1950er Jahre von der ersten Schülergeneration ausmachen, die sich in den 1960er Jahren bereits durch eine „Kritische Politikwissenschaft“ (Klingemann 2002: 56) herausgefordert sah. Die mikrofundierte empirische Wahlforschung, wie sie in den Vereinigten Staaten von Amerika dominierte, konnte erst mit der Kölner Wahlstudie von 1961, die wesentlich von Scheuch und Wildenmann (1965) initiiert wurde, in Deutschland Fuß fassen (vgl. Kaase 2000).

Dabei gingen wichtige Impulse von Soziologen aus, denn die empirische Wahlforschung wurde im Wesentlichen als Teilgebiet der Politischen Soziologie im Nachkriegsdeutschland aufgebaut (vgl. zur Entwicklung der Soziologie im Nachkriegsdeutschland Scheuch 1990, 2002). Dies ist interessant, zumal sich die Re-Etablierung der Soziologie in Deutschland wesentlich schwieriger gestaltete als die der Politikwissenschaft. Während die Einrichtung neuer politikwissenschaftlicher Lehrstühle und Institute als Versuch verstanden wurde, eine Demokratisierung Deutschlands zu forcieren, musste die Soziologie ihre Legitimation zwischen der aufstrebenden Politologie und der etablierten Nationalökonomik

suchen. Die empirische Ausrichtung der Soziologie ist nicht zuletzt auf den Einfluss einiger Re-Emigranten zurückzuführen. So ist die Institutionalisierung empirischer Forschungsinstitute in Form des Zentralarchivs in Köln (ZA) und des Zentrums für Umfragen, Methoden und Analysen (ZUMA) in Mannheim maßgeblich von Soziologen wie M. Rainer Lepsius und Erwin K. Scheuch mitgeprägt worden und auch in der soziologischen Ausbildung wurde schneller und konsequenter als in der Politologie die Methodenausbildung im Curriculum verankert.

Trotz dieser Institutionalisierung fristete die empirische Sozialforschung in Deutschland bis in die 1970er Jahre hinein ein Nischendasein. Dies bot möglicherweise die Chance der Entwicklung eines gemeinsamen Forschungsprogramms der empirischen Wahlforscher, was Naschold (1971) zu dem Urteil führte, dass es sich um eine der am weitest entwickelten Disziplinen der Sozialwissenschaften handele – ein Urteil, das bis heute Bestand hat (vgl. bspw. Eith 2008). Dieses einheitliche Forschungsprogramm der Wahlforschung führte in den 1990er Jahren jedoch zur Kritik mangelnder Innovationsfreudigkeit (Schultze 1991).

Auch wenn diese Kritik in der Disziplin nicht geteilt wird (Falter/Rattinger 1992), kann man konstatieren, dass die deutsche Wahlforschung über viele Jahre im Wesentlichen von einem Netzwerk aus Wissenschaftlern aus der Zeit der Kölner Wahlstudie geprägt wurde, die sehr erfolgreich ihre Forschungsschwerpunkte auch über ihre Schüler in der deutschen Wahlforschung verankerten. Damit einher ging auch ein zunächst relativ enges Forschungsprogramm, das sich im Wesentlichen an der *Michigan School* (Campbell et al. 1954, 1964) orientierte.

Erst in jüngerer Zeit scheint sich eine Pluralisierung der Wahlforschung in Deutschland durchzusetzen. Dies ist unter anderem daran abzulesen, dass die Monopolstellung der *Blauen Bände*⁸ als – über viele Jahre – zentrales Wahlana-

8 Die wegen ihres (zumeist) blauen Einbandes so genannte Reihe etablierten Max Kaase und Hans-Dieter Klingemann als zentrale Publikation der Wahlanalyse mit dem ersten Band zur Bundestagswahl 1980 (Kaase/Klingemann 1983). Die Reihe wurde bis zur Bundestagswahl 1998 von diesen beiden betreut (Klingemann/Kaase 2001) und wird aktuell von Gabriel, Falter und Weßels (Falter et al. 2005; Gabriel et al. 2009) weitergeführt. Den Blauen Bänden voraus gingen zwei Aufsatzsammlungen mit systematischen Analysen der Bundestagswahlen von 1972 und 1976, die jeweils in Schwerpunktheften der Politischen Vierteljahresschrift erschienen (Kaase 1973, 1977). Zu früheren Bundestagswahlen liegen nur Einzelstudien vor; die erwähnte Kölner Wahlstudie (Scheuch/Wildenmann 1965) war mit Abstand die einflussreichste

lysewerk der Bundestagswahlen Konkurrenz bekommen hat (etwa durch Brett-schneider et al. 2004, 2007). Auch löst sich die Konzentration auf die klassischen Erklärungsansätze (makro- und mikrosoziologischer sowie sozialpsychologischer Ansatz) zunehmend auf und es werden neue theoretische Perspektiven, wie sie etwa Merrill III und Grofmann (1999) vorschlugen, berücksichtigt (bspw. Debus 2007). Auch werden in verstärktem Maße neue methodische Wege beschritten, die für Impulse sorgen.

Zum einen sind hier durch technische Entwicklungen nun leichter anwendbare rechenintensive statistische Verfahren wie Mehrebenenanalysen zu nennen, die es ermöglichen, Einflüsse von Kontexteffekten auf das Wahlverhalten zu untersuchen (z. B. Klein/Plötschke 2000). Ebenfalls häufiger findet man Zeitreihenanalysen (Pappi/Bytzek 2007; Schaffer/Schneider 2005). Eine weitere methodische Innovation bezieht sich auf die Anwendung neuer Designs. Im Rahmen der Surveyforschung werden etwa verstärkt faktorielle Surveys eingesetzt (Klein/Rosar 2005a, 2005b, 2009) und auch Experimente gewinnen an Bedeutung (Gschwend/Hooghe 2008).

In thematischer Hinsicht ist ebenfalls eine Verbreiterung des Spektrums zu beobachten. So nimmt die zunächst nur von Kirchgässner (1976, 1984) aufgegriffene Diskussion um das *economic voting* einen größeren Stellenwert ein (vgl. Feld/Kirchgässner 2000; Fröchling 1998; Kellermann/Rattinger 2007; Kirchgässner 2009; Pappi/Bytzek 2007; Rattinger/Faas 2001). Mit *economic voting* ist eine Forschungsrichtung angesprochen, welche einen dominanten Einfluss wirtschaftlicher Faktoren wie Inflation, Arbeitslosigkeit, Wirtschaftswachstum, aber auch die subjektive Leistungsbewertung der Wirtschaftspolitik der Parteien und/oder Kandidaten auf das Wählerverhalten postuliert.⁹ Anschließend an die Arbeiten von beispielsweise Downs (1968), Fair (1978), Key (1960), Kramer (1971) und Tufte (1978) wird hier postuliert, dass wirtschaftliche Prosperität im

dieser Arbeiten (für einen knappen Überblick über die einschlägigen deutschen Wahlstudien siehe Roth 1998: 159ff.).

9 Eine kompakte Einführung liefern Lewis-Beck und Paldam (2000) sowie Anderson (2007); einen Überblick über den Forschungsstand in verschiedenen Ländern kann man durch die Beiträge im Sammelband von Dorussen und Taylor (2002) gewinnen. International vergleichende empirische Befunde haben Chappell und Veiga (2000) sowie Lewis-Beck und Stegmaier (2000) zusammengetragen. Auch für Deutschland liegen verschiedene Befunde zum Einfluss ökonomischer Variablen vor. Einen aktuellen Überblick liefern Pappi und Shikano (2007: 59ff.).

Allgemeinen den amtierenden Regierungsparteien nutzt, während die Opposition von wirtschaftlichen Krisen profitiert. Aktuelle Diskussionen drehen sich dabei um Fragen, unter welchen Bedingungen retrospektive oder prospektive Leistungsbewertungen beziehungsweise soziotropisches oder egozentrisches Wählerverhalten relevant sind.

Weiterhin werden verstärkt Phänomene wie Stimmensplitting, strategisches und koalitionsorientiertes Wählen berücksichtigt (Bytzek 2008; Gschwend 2003, 2007; Hilmer/Schleyer 2000; Linhart 2007; Pappi/Thurner 2002; Schoen 1999; Thurner/Pappi 1998). Für diese Phänomene stellt Deutschland aufgrund der zwei zu vergebenden Stimmen bei Bundestagswahlen, der Fünf-Prozent-Hürde sowie der zumeist notwendigen Koalitionen für die Regierungsbildung einen besonders interessanten Fall dar. Auch für die Prognose von Wahlergebnissen ist diese Regelung entscheidend, denn dadurch gewinnt die korrekte Vorhersage der Stimmenanteile kleiner Parteien bedeutend an Gewicht. Ob kleine Parteien die Fünf-Prozent-Marge übertreffen oder nicht, beeinflusst maßgeblich die Zusammensetzung des Parlaments und damit die Mehrheitsverhältnisse und mögliche Regierungskoalitionen.

Dennoch ist ebenso festzustellen, dass – anders als etwa in den USA oder auch anderen europäischen Staaten wie Großbritannien oder Frankreich – bisher kaum Versuche seitens der deutschen wissenschaftlichen Wahlforschung bestehen, Wahlergebnisse vorherzusagen (für eine Ausnahme siehe Abschnitt 2.3.2). Vielmehr beschränkt man sich auf die Beschreibung und insbesondere Erklärung von Wahlverhalten und überlässt – vielleicht auch aufgrund der gewachsenen und besonderen Verzahnung wissenschaftlicher und privatwirtschaftlicher Wahlforschung – die Voraussage den Meinungsumfrageinstituten, wenngleich diesen Versuchen auch seitens der Wissenschaft immer wieder kritisch begegnet wird. Diese Einwände werden im Folgenden thematisiert.

2.1.2 Kritik an der Meinungsforschung

Die Meinungsforschung wie auch die mikrofundierte empirische Forschung generell sah sich insbesondere nach ihrer Etablierung Kritik ausgesetzt. In Deutschland bezog sich diese insbesondere auf die privatwirtschaftlichen Meinungsforscher, die sich – wie geschildert – einer überwiegend normativ geprägten

Politikwissenschaft gegenüber sahen und damit auf keine argumentative Unterstützung seitens der Wissenschaft hoffen konnten. Zudem banden sich diese sehr früh eng an politische Parteien, was weitere Angriffsfläche für Kritik bot. Schließlich bestand von Beginn an ein Interesse der politischen Meinungsforschung an den Möglichkeiten der Voraussage von Wahlergebnissen auf Grundlage von Umfragen. Auch dies implizierte eine lebhaft, bis heute nicht abgeschlossene Diskussion, welche vielleicht die größten Auswirkungen auf die Praxis der politischen Meinungsforschung ausübte. Mit der zunehmenden Etablierung einer mikrofundierten empirischen Wahlforschung in der deutschen Hochschullandschaft ließ auch die geistesgeschichtlich-normativ geprägte Kritik an der Meinungsforschung als solcher nach. An die Stelle dieser nicht selten fundamentalen Kritik, welche die Existenzberechtigung quantitativer Individualdatenanalyse in Frage stellte, ist in jüngerer Zeit eher eine Debatte über methodische Aspekte der Meinungsforschung sowie deren Umgang mit der Berichterstattung ihrer Ergebnisse getreten. An dieser Stelle werden sowohl die aus heutiger Sicht als historisch zu bezeichnende Debatte um die Meinungsforschung per se nachgezeichnet als auch die aktuellen Diskussionspunkte aufgegriffen.

Hierbei wird nicht chronologisch vorgegangen, da sich die unterschiedlichen Kritiken auch zeitlich überschneiden, sondern es werden vier zentrale Aspekte aufgegriffen und erörtert. An dieser Stelle kann diese vielschichtige Diskussion nur in ihren Grundzügen nachgezeichnet werden und es werden darüber hinaus Einwände, die empirische Individualdatenanalyse per se ablehnen ebenso wenig betrachtet (vgl. hierzu Keller 2001: 259ff.) wie die mannigfaltige Kritik am Begriff der öffentlichen Meinung (klassisch hierzu: Blumer 1948; Bourdieu 1993; die Gegenperspektive liefert Noelle 1967). Kritikpunkte, welche sich explizit auf die Praxis von Wahlabsichtsfragen beziehen, werden in Abschnitt 2.1.3 aufgegriffen.

Die aus der Debatte als zentral herausgegriffenen Punkte betreffen erstens die mögliche Beeinflussung des Meinungsklimas und der Wählerentscheidung durch Meinungsumfragen. Zweitens wird die postulierte Beeinflussung politischer Entscheidungen, drittens die vermuteten Auswirkungen auf den verfassungsmäßigen Staatsaufbau und viertens die Gefahren im Umgang mit den Ergebnissen der Meinungsforschung angesprochen.

2.1.2.1 Beeinflussung des Wählers

Ein Strang der Kritik richtet sich gegen die missbräuchliche Verwendung des Instrumentariums der Meinungsforschung und/oder der nicht-intendierten Wirkung demoskopischer Ergebnisse auf die Wählerschaft. Ausgehend von Zweifeln an der Gültigkeit der generierten Aussagen wurde die Produktion von gefährlichem „Herrschaftswissen“ (Wildenmann 1992: 62) postuliert und Manipulation vermutet. Albrecht (1988: 45) etwa wirft der Branche vor, sie habe „eine handfeste Aufgabe (...) [, nämlich] das sogenannte Meinungsklima selbst in der einen oder anderen Richtung zu beeinflussen“. Hintergrund dieser – häufig scharf formulierten – Kritik ist das Postulat, dass sich Wähler in ihrer Entscheidung durch die Publikation von Wahlabsichtsfragen auch aufgrund der eigenen Uninformiertheit und Unselbständigkeit beeinflussen lassen. Dies würde von den Meinungsforschungsinstituten, welche nicht selten als Handlanger ihrer politischen Auftraggeber aufträten, zumindest in Kauf genommen (etwa Albrecht 1988; Rückmann 1972; Scharp 1958).

In diesem Kontext werden jenseits normativer Einwände verschiedene mögliche Wirkungen von publizierten Meinungsumfragen auf die Wählerentscheidungen diskutiert. Am prominentesten sind dabei der *bandwagon*- sowie der *underdog*-Effekt. Mit ersterem ist ein Mitläufereffekt angesprochen, wonach Wähler gerne auf den Sieger setzen und somit wird durch die Publikation insbesondere der Sonntagsfrage die jeweils stärkste Partei begünstigt. Mit dem *underdog*-Effekt wird eine mögliche gegenteilige Wirkung postuliert (auch Mitleidseffekt genannt, bei Hartenstein (1967: 288): „Bürgerschreck-Effekt“). Demnach wird vermutet, dass gerade die in Umfragen schwächer abschneidenden Parteien am Wahltag profitieren, da so die Stammwähler mobilisiert werden würden (vgl. Simon 1954).

Entgegen dieser Position haben Kirchgässner und Wolters (1987) in Erweiterung des Downsschen (1968: 35ff.) Ansatzes des Wählerverhaltens ein Modell vorgelegt, welches für Mehrparteiensysteme mit Fünf-Prozent-Klausel zeigt, dass weder *bandwagon*- noch *underdog*-Effekte theoretisch zu erwarten sind. Vielmehr begünstigen „wiederholte Meinungsumfragen“ (Kirchgässner 1986: 246) rationales Wählen, weshalb die Förderung ihrer Veröffentlichung gefordert wird.

Mit Blick auf das deutsche Wahlrecht werden zwei weitere Effekte diskutiert: der Fallbeileffekt sowie der Leihstimmeneffekt. Diese Effekte hängen wie-

derum eng mit theoretischen Überlegungen aus *Rational-Choice*-Perspektive über das Wählerverhalten in Mehrparteiensystemen zusammen (für eine Übersicht vgl. Mueller 2003: 264ff.). So gründet der Fallbeileffekt auf der Annahme, dass sich die Wähler möglicherweise für eine alternative Partei entscheiden, sofern die eigentlich meist präferierte Partei an der Fünf-Prozent-Hürde zu scheitern droht. Der Wähler, so das Argument, würde seine Stimme lieber einer potenziell erfolgreichen und in diesem Fall auch möglicherweise inhaltlich weiter entfernten Partei geben, als diese zu verschenken. Demnach versucht der Wähler so genannte *wasted votes* zu vermeiden (für Deutschland vgl. hierzu Behnke 2007: 207ff.; Schoen 1999).

Mit dem Leihstimmeneffekt wird ebenfalls das Phänomen des strategischen Wählens im deutschen Wahlsystem angesprochen durch die darin verankerte Möglichkeit des Aufteilens der Erst- und Zweitstimme auf unterschiedliche Parteien. Die Idee hier ist, dass etwa bestimmte Koalitionen von Parteien bevorzugt werden, für die dann entsprechend gestimmt wird – so ist eine Vermutung, dass kleine Parteien eher Zweitstimmen von Anhängern großer Parteien „geliehen“ bekommen, während große Parteien eher durch Erststimmen für ihre Wahlkreis-kandidaten von Anhängern der kleinen Parteien profitieren.

Die empirische Evidenz hinsichtlich der postulierten Effekte fällt gemischt aus. Zwar belegen einige frühe Laborexperimente das Vorliegen eines *bandwagon*-Effekts (vgl. bspw. Sorokin/Boldyreff 1932), doch deren externe Validität bei Wahlen wird durch natürliche Experimente in Frage gestellt. Aufgrund der Zeitverschiebungen in den USA haben bei bundesweiten Wahlen an der Westküste die Wahllokale noch geöffnet, während an der Ostküste bereits die ersten Hochrechnungen vorgestellt werden. Dieses Phänomen des *Western Voting* ermöglicht gute Testmöglichkeiten des Vorliegens von Effekten der Veröffentlichung von Ergebnissen von Wahlumfragen auf das Wahlergebnis. Entsprechende Studien finden jedoch weder Belege für das Wirken des *bandwagon*-noch des *underdog*-Effekts (für einen Überblick vgl. Donsbach 1984). Gallup (1965: 546) spricht deshalb auch vom „bandwagon myth“.

In einer vergleichbaren Situation zum *Western Voting* im Rahmen einer Nachwahl zur Bundestagswahl 1965 konnten entsprechende Effekte ebenfalls nicht gefunden werden (Hartenstein 1967). Allerdings war die Zahl der Stimmensplitter bei der Nachwahl 1965 in den beiden Wahlkreisen ungewöhnlich hoch, was man in Richtung einer Begünstigung strategischen Wählens durch die

Veröffentlichung von Meinungsumfragen deuten könnte und damit zumindest als Hinweis auf eine Bestätigung des Modells von Kirchgässner und Wolters (1987) aufgefasst werden könnte.

Auf Individualdatenbasis haben für Deutschland Lupri (1969), Donsbach (1984), Brettschneider (1991, 1992, 2000), Schmitt-Beck (1996) sowie Faas und Schmitt-Beck (2007) einschlägige Studien zum Thema vorgelegt.¹⁰ Als Desiderat dieser Studien ist festzuhalten, dass Meinungsumfragen zunehmend von den Wählern wahrgenommen werden und sich mit der Kenntnis der Ergebnisse der Vorwahlumfragen auch die Erwartungen hinsichtlich der antizipierten Koalitionsregierung verändern. Sowohl die Wahrnehmung als auch die Erwartungsanpassung des Wahlausgangs variiert dabei mit soziodemographischen Merkmalen (Geschlecht, Alter, Beruf) und insbesondere mit der politischen Involvierung: Politisch Interessierte nehmen Wahlumfragen eher wahr und können in Umfragen eher deren Voraussagen replizieren. Das politische Interesse ist wiederum stark mit soziodemographischen Merkmalen, insbesondere Bildung konfundiert. Interessant ist darüber hinaus, dass ebenfalls gefunden wurde, dass FDP-Anhänger (unter Kontrolle auf die genannten Variablen) eher Wahlumfragen wahrnehmen als die Anhänger anderer Parteien – dies könnte man aufgrund der häufig unsicheren Situation der FDP im Hinblick auf den Einzug in den Bundestag als weiteres Indiz dafür werten, dass Vorwahlumfragen eher rationales beziehungsweise strategisches Wählen begünstigen. Auch im internationalen Kontext kommen Blais et al. (2006) zu ähnlichen Befunden: Sie zeigen anhand Auswertungen kanadischer Wahlstudien, dass Wahlabsichtsfragen die Erwartungen der Wähler bezüglich des Wahlausgangs beeinflussen.

Befunde zu den direkten Einflüssen auf das Wahlverhalten sind spärlich und zum Teil methodisch eher problematisch (etwa selbst berichtete Einschätzung über die Beeinflussung durch Wahlumfragen bei der eigenen Wahlentscheidung), kommen jedoch übereinstimmend zum Schluss, dass für die postulierten Effekte kaum empirische Evidenz vorliegt und die Einflüsse situationsspezifisch variieren (Brettschneider 1992: 66). Auch aus anderen Ländern werden widersprüchliche Befunde berichtet. Während West (1991) Hinweise auf einen *underdog*-Effekt findet, welcher jedoch bei US-Präsidentenwahlen seinen Analysen

10 Überblicke über den Forschungsstand liefern Gallus und Lühe (1998) sowie Schoen (2002).

zufolge äußerst schwach ist, belegen Skalabans (1988) Auswertungen die Existenz eines *bandwagon*-Effekts im Rahmen der US-Präsidentenwahlen 1980 und 1984. Auch für Großbritannien gehen McAllister und Studlar (1991) von der Existenz des *bandwagon*-Effekts aus. Entgegen dieser Befunde können Gallup und Rae (1940) keine Hinweise auf entsprechende Effekte finden und auch die jüngeren Studien von Ansolabehere und Iyengar (1994) sowie Blais et al. (2006) negieren entsprechende Effekte. Vor dem Hintergrund der insgesamt immer noch dürftigen empirischen Basis, fordert Gallus (2003: 135) eine „systematische Demoskopiewirkungsforschung“.

Weiterhin werden Effekte von Vorwahlumfragen auf die Wahlbeteiligung diskutiert. So wird ein Defätismuseffekt postuliert, welcher aussagt, dass die potenziellen Wähler der schwächeren Partei, ob der aussichtslosen Lage lieber gleich der Wahlurne fern bleiben. Weiterhin ist ein Lethargieeffekt denkbar, welchem zufolge sich die Wähler der stärkeren Partei siegessicher wähnen und folglich den Wahlausgang von zuhause verfolgen, ohne selbst ihre Stimme abzugeben. Der Bequemlichkeitseffekt hingegen vermutet, dass bei sicherem Wahlausgang die Unentschiedenen sich nicht an der Wahl beteiligen werden.

Anzunehmen ist, dass das Ausmaß dieser genannten Effekte maßgeblich mit der erwarteten Knappheit des Wahlausganges zusammenhängt. Diese Hypothese geht auf das erstmals von Downs (1968: 255ff.) beschriebene Wahlparadoxon zurück. Unterstellt man rationale Wähler, dann ist theoretisch nicht zu erwarten, dass diese sich an einer Wahl beteiligen werden, da die einzelne Stimme zu unbedeutend für den Wahlausgang ist und damit die Kosten des Urnengangs schnell dessen Nutzen übersteigen. Da der eigene Einfluss der Stimme von Wählern systematisch überschätzt wird, ist ein Effekt der erwarteten Knappheit des Wahlausgangs auf die Wahlbeteiligung in der Form zu erwarten, dass mit zunehmender Knappheit des Wahlausgangs die Wahlbeteiligung steigt (für einen Überblick vgl. Aldrich 1993; Mueller 2003: 303ff.). Mit steigender Bedeutung dieses Mobilisierungseffekts sollte auch die Abnahme der Relevanz der beschriebenen Effekte einhergehen.

Auch wenn es eine rege Forschung zur Ergründung der Ursachen für unterschiedliche Wahlbeteiligungshöhen gibt (für einen Überblick in international vergleichender Perspektive vgl. Blais 2006), sind die empirischen Befunde im Hinblick auf die angesprochenen Hypothesen – vor allem für Deutschland – spärlich gesät. Hartenstein (1967: 296f.) zeigt, dass bei der Nachwahl in zwei Wahl-

kreisen bei der Bundestagswahl 1965 der Rückgang der Wahlbeteiligung in diesen von 1961 auf 1965 deutlich größer ausgefallen ist als im Bundesvergleich. Brettschneider (2000: 492) sieht in einer Zusammenfassung des Forschungsstandes keine Anzeichen für eine Bestätigung auch nur eines der postulierten Effekte; auch für Deutschland vermutet er keine Auswirkungen von Vorwahlumfragen auf die Höhe der Wahlbeteiligung – auch wenn er diesbezüglich keine entsprechenden Befunde anführt (ebd.). Kirchgässner und Meyer zu Himmern (1997) können mit ihren Analysen einen Mobilisierungseffekt für die Bundestagswahl 1987 finden, nicht jedoch für die Wahl 1983. Für die ersten gesamtdeutschen Wahlen 1990 und 1994 zeigt sich für die alten Bundesländer wiederum ein signifikant positiver Effekt der erwarteten Knappheit auf die Wahlbeteiligung, während sich in den neuen Bundesländern ein signifikant negativer (1990) beziehungsweise kein (1994) Effekt feststellen lässt. Auch im internationalen Kontext geht man, wenn, dann von eher schwachen Effekten aus. So folgert Sudman (1986: 331) in seinem Forschungsüberblick für die USA, dass „exit polls appear to cause small declines in total voting in areas where the polls close late for those elections where the exit polls predict a clear winner when previously the race had been considered close“. Insofern ist auch in Bezug auf den Kenntnisstand dieser Effekte zu konstatieren, dass weitere Forschung dringend geboten scheint.

Insgesamt ist bezüglich der Diskussion um die Wirkung demoskopischer Ergebnisse auf die Wählerschaft festzuhalten, dass solche Einflüsse nach wie vor häufig postuliert werden, obwohl bisher keine überzeugende empirische Evidenz für das Vorliegen entsprechender Effekte gefunden wurde. Hierbei wird oft normativ argumentiert und es werden häufiger Regelungen bis hin zu einem Verbot der Publikation von Meinungsfragen im Vorfeld von Wahlen gefordert. Dies ist im Übrigen keineswegs unüblich. So gibt es entsprechende Regelungen beispielsweise in Österreich, Polen, Portugal, der Schweiz, Spanien und Ungarn (Gallus 2002: 34). Die deutsche Diskussion hierzu wurde insbesondere in den 1970er Jahren geführt (Benda 1974, 1981; Dach 1997; Hartenstein 1969; Loewenstein 1971; Neeff 1971), jedoch gibt es bis heute immer wieder entsprechende Vorstöße – zuletzt im Rahmen des Wahlkampfes zur Bundestagswahl 2002 durch den SPD-Politiker Wolfgang Thierse (vgl. Gallus 2003: 130). Aufgrund mangelnder empirischer Evidenz für das Vorliegen der postulierten Effekte und dem nach wie vor unbefriedigenden Forschungsstand in dieser Hinsicht wurden in Deutschland Verbotsforderungen jedoch zurückgewiesen.

2.1.2.2 Beeinflussung politischer Entscheidungen

Eine weitere Gefahr von Meinungsumfragen wird häufig darin gesehen, dass politische Entscheidungen beeinflusst werden (vgl. Kruke 2007: 452f.). Diese ist dabei, dass die fortwährende Untersuchung der Mehrheitsmeinung dazu führe, dass Politiker ihre Entscheidung nicht mehr souverän treffen würden. Grundlage dieser Kritik ist nicht zwingend die Auffassung, dass die Meinungsforschung per se versucht politische Entscheidungen bewusst zu beeinflussen, als vielmehr die Vermutung, dass Umfragen von politischen Entscheidungsträgern als Volkswille aufgefasst werden. Hennis (1957: 42), der diese Kritik erstmals formulierte, sah die Gefahr der Meinungsumfragen in der Folge darin, „dass sie zu einem Abbau der staatsmännischen Führungs- und Verantwortungsbereitschaft beitragen“.

Vertreter der großen Meinungsforschungsinstitute entgegneten zum einen, dass die Verantwortung der Verwendung ihrer Ergebnisse nicht bei ihnen läge, da sie lediglich die Informationen liefern würden. Zum anderen wurden gerne zahlreiche Beispiele dafür angeführt, in welchen konkreten Sachfragen sich Politiker über die qua Demoskopien berichtete Mehrheitsmeinung hinweg gesetzt hätten (vgl. Schmidtchen/Noelle-Neumann 1963).

Diese Debatte weist eher prosaische Züge auf und mündete insbesondere in eine normative Diskussion über die Bewertung der Meinungsforschung aus demokratietheoretischer Sicht (dazu der nächste Punkt). Empirisch belastbare Studien, die den direkten Einfluss von Meinungsumfragen auf politische Entscheidungen thematisieren, liegen nicht vor. Einzig Fuchs und Pfetsch (1996) haben eine entsprechende Arbeit vorgelegt, die aus system- sowie demokratietheoretischer Perspektive der Frage nachgeht, inwieweit das Regierungssystem auf die öffentliche Meinung reagiert und sich an ihr orientiert. Ihre Ausführungen sind zum einen rein theoretischer Art und zum anderen stützen sie sich auf eine Befragung von 38 Personen – das Ergebnis, „daß im Vorstellungsbild der Akteure des Regierungssystems sowohl die Responsivität als auch die Beeinflussbarkeit [durch die Umfragen, Anmerkung des Autors] enthalten sind“ (ebd.: 128), wird aufgrund der niedrigen Fallzahl sowie dem nicht näher erläuterten Verfahren der Stichprobenziehung als nicht belastbar eingestuft.

Im Ergebnis wurde den – empirisch nicht abgesicherten – Einschätzungen von Friedeburg (1961), Furmaniak et al. (1975) und auch Bretschneider (1985) gefolgt, die keine direkte Beeinflussung der Wahlentscheidungen sehen, hinge-

gen die Effekte auf die politische Propaganda beziehungsweise Wahlkampfführung und Öffentlichkeitsarbeit als relevant betrachten.

Der Einfluss von Meinungsumfragen auf die Ausrichtung von Wahlkämpfen ist gut dokumentiert und auch die Wahlkampfwirkungsforschung erlebt in jüngerer Zeit verstärkte Aufmerksamkeit (Schoen 2005a). Hierbei stehen zum einen die Wahlkampagnen selbst im Fokus des Interesses. So liegen beispielsweise Analysen des Wahlkampfes der SPD im Vorfeld der Bundestagswahl 1998 (Schmitt-Beck 2001) sowie der Bundestagswahl 2005 (Schmitt-Beck/Faas 2007) vor. Zum anderen ist die Frage nach der Wirkung dieser Wahlkämpfe auf die Wählerschaft relevant. Insbesondere im Rahmen des mikrosoziologischen Ansatzes (Berelson et al. 1986: 118ff.; Lazarsfeld et al. 1955: 73ff.) wird versucht die Wirkung von Wahlkämpfen zu erfassen. Demnach ist zu erwarten, dass Wähler, welche *cross-pressure* ausgesetzt sind, als Zielgruppe für Wahlkämpfe interessant sind. Grundsätzlich sind Wahlkämpfe als Teil des Einflusses interpersonaler Kommunikation und der Medienberichterstattung aufzufassen, welche diesem Ansatz zufolge insbesondere als Aktivator und/oder Verstärker politischer Prädispositionen wirken. Der sozialpsychologische Ansatz liefert ebenfalls Hinweise auf mögliche Wirkungsweisen von Wahlkämpfen (Campbell et al. 1964: 285f.). Hiernach wirken diese auf die Themen- und Kandidatenorientierung der Wähler. Damit ist eher eine nachhaltige Wirkung von Wahlkämpfen auf Wähler mit wenig verfestigten Parteibindungen und kurzfristig vor einer Wahl zu erwarten. Problematisch indes ist, dass in beiden Ansätzen eine Isolierung der Wahlkampfwirkung von anderen kommunikationsbedingten Einflüssen schwierig erscheint.

Entsprechend ist die empirische Evidenz zur Wirkung von Wahlkämpfen auf die Wählerschaft eher zwiespältig. Bereits Lazarsfeld et al. (1955: 40ff.) belegen, dass nur ein kleiner und selektiver Teil der Wählerschaft überhaupt Wahlkämpfe rezipiert – eine notwendige Voraussetzung zur Entfaltung einer etwaigen Wirkung. Für Deutschland kann Schmitt-Beck (2002) diesen Befund bestätigen. In Bezug auf die Wirkungen des Wahlkampfes auf diese Wählergruppe stellen Lazarsfeld et al. (1955: 73ff.) fest, dass insgesamt eher Effekte der Stabilisierung und Aktivierung festzustellen sind. Demnach bewirken Wahlkämpfe eher eine Mobilisierung der Stammwähler als einen Umschwung von Wählern zu Parteien, welche der eigenen Parteibindung weniger entsprechen. Diese Befunde konnten

in der Folge bestätigt werden (Schoen 2005a: 537) und auch für Deutschland liegen konforme empirische Resultate vor (z. B. Finkel/Schrott 1995).

Unklar bleibt allerdings, inwiefern hierfür Ergebnisse von Wahlumfragen bedeutsam sind. Einflüsse von Wahlumfragen auf die Wahlkampfführung sind belegt und es konnte auch vielfach gezeigt werden, dass zumindest ein Teil der Wähler Wahlkämpfe intensiv verfolgt und ihr Wahlverhalten hiervon beeinflussen lassen, auch über eine Modifizierung ihrer Einstellungen zu Themen und/oder Kandidaten. Befunde zu den letztlich wirkenden Meinungsfragen über die Wahlkämpfe auf die Wählerschaft liegen jedoch nicht vor und sind – angesichts der jeweils eher geringen Effekte und zahlreicher anderer Einflüsse – allenfalls als gering einzustufen.

2.1.2.3 Beeinflussung der staatlichen Konstitution

Als weiterer Diskussionsstrang auszumachen ist der Disput über die potenziellen strukturellen Veränderungen, die Meinungsumfragen letztlich bewirken könnten. Kritiker (etwa Hennis 1957, 1983; Sontheimer 1964; Thiemer 1973) postulierten die Unterminierung der parlamentarischen Demokratie, da durch den massiven Rückgriff auf die Ergebnisse der Meinungsforschungsinstitute plebiszitäre Elemente – entgegen dem Willen der Verfassungsväter – gestärkt würden. Letztlich ist die Demoskopie dieser Position zufolge als „demokratiefeindliches Instrument“ (Hennis 1957: 37) einzustufen.

Insbesondere führende Mitglieder von Meinungsforschungsinstituten sehen hingegen die Demoskopie als „demokratische Methode“ (Neumann 1952: 44) an, der gar emanzipatorische Wirkungen zugesprochen werden (bspw. Crespi 1952; Schmidtchen/Noelle-Neumann 1963: 191). So kontrovers auch diese Debatte geführt wurde (vgl. etwa die Beiträge der in Kaase et al. 1983 dokumentierten Podiumsdiskussion), muss aus heutiger Sicht festgehalten werden, dass diese im Sand verlaufen ist, und der Auffassung des Autors zufolge ein Prozess des „Abbaus der tragenden Grundlagen der repräsentativen Demokratie“ (Hennis 1957: 46) aufgrund zunehmender Meinungsumfragen nicht festzustellen ist. Zumindest ist in Bezug auf das damit zumeist einhergehende Postulat ausufernder direktdemokratischer Elemente auch aus empirischer Sicht zu konstatieren, dass ein Ausbau direktdemokratischer Beteiligung auf Bundesebene trotz entsprechender

Diskussionen insbesondere nach der Wiedervereinigung nicht festzustellen ist. Auch auf Bundesländerebene sind nach wie vor direktdemokratische Elemente wenig ausgebaut (Hartleb/Jesse 2005; Jesse 2008: 283ff.; Rehmet 2002; Ritter 1992: 42ff.; Schiller/Mittendorf 2002).

2.1.2.4 Verwendung der Ergebnisse von Meinungsumfragen

Schließlich wurde immer wieder auf die Gefahren im Umgang mit den Ergebnissen der Meinungsforschung hingewiesen. Diese Kritik zielt größtenteils weniger direkt auf die Meinungsforschung als solche ab als vielmehr auf die Medien, welche die Ergebnisse publizieren. Betont wird hierbei die Verantwortung der Medien, aber auch der Umfrageinstitute, in Bezug auf die Praxis der Berichterstattung. Hier wird beispielsweise darauf verwiesen, die jeweilige Leserschaft über methodische Details zu informieren, und auch eine umsichtige Interpretation der Befunde wird gefordert, welche nicht über die Belastbarkeit der Ergebnisse hinausgehen soll (vgl. hierzu z. B. Atteslander 1980; Decker/Zadra 2001; von Friedeburg 1961).

Auch wenn diese Einwände seitens der kommerziellen Meinungsforschung durchaus ernst genommen werden und Richtlinien hinsichtlich der Verwendung und Kommunikation von Umfrageergebnissen erarbeitet wurden (Kaase 1999b; Sahner 1998), erscheint diese Forderung bis heute aktuell. Ein zunehmendes Interesse an der Analyse des Verhältnisses zwischen Demoskopie und Medien ist darüber hinaus zu bemerken (Brettschneider 1996, 2002b, 2005b; Dahlem 2001; Decker/Zadra 2001; Donovitz 1999; Gallus/Lühe 1998).

Aus heutiger Perspektive zeigt sich, dass insbesondere die bis in die 1980er Jahre normativ geprägte Auseinandersetzung um die Auswirkungen der Meinungsforschung auf die Demokratie nicht mehr geführt wird. Es muss konstatiert werden, dass sich hier die Position der Meinungsforscher behauptet hat, wenngleich sie weder mit ihren Argumenten die Kritiker überzeugen, noch überzeugende empirische Belege in nennenswertem Umfang für ihre Positionen vorlegen konnte. Kruke (2007: 465) konstatiert als Ergebnis der Kritik an der Demoskopie und deren Reaktionen darauf, dass letzten Endes die Meinungsforschung akzeptiert wurde und deren „Mechanik in der Beschreibung öffentlicher Meinung“ nicht mehr in Frage gestellt wurde.

Bis heute virulent ist jedoch die Diskussion über die Einflüsse von Ergebnissen von Meinungsumfragen auf individuelle Wahlentscheidungen oder ihre indirekten Wirkungen über Wahlkämpfe und Medienvermittlung. Während hierüber bisher normativ und ergebnislos gestritten wurde, ist anstelle dessen das Bemühen getreten, diese möglichen Einflüsse empirisch zu untersuchen. Die Beantwortung dieser Fragen ist überwiegend offen, jedoch sind zunehmend Anstrengungen zu erkennen, die Wechselwirkungen zwischen kommerzieller Meinungsforschung und Medien zu untersuchen.

2.1.3 *Wahlabsichtsfragen*

Die Wahlabsichtsfrage ist fest etablierter Bestandteil, gar „Kernindikator“ (Wüst 1998: 325) der politischen Dauerbeobachtung durch die kommerzielle Meinungsforschung. Dabei hat sich die so genannte Sonntagsfrage als Standardinstrument durchgesetzt, die in ähnlicher Form von allen großen Meinungsforschungsinstituten eingesetzt wird. Als Fragestellung hat sich folgende Version etabliert: „Wenn am nächsten Sonntag Bundestagswahl wäre, welche der folgenden Parteien würden Sie dann wählen?“.

Diese Frage ist Teil praktisch jeder Befragung zur Erfassung der politischen Stimmung und wird gleichermaßen von Wissenschaft wie Meinungsforschungsinstituten verwendet. Erstere ziehen die Wahlabsicht öfter als Proxy für tatsächliches Wahlverhalten heran, insbesondere um Hypothesen hinsichtlich kausaler Mechanismen der Wählerentscheidung zu testen (ein Beispiel hierfür: Schoen 2004a), letztere nutzen sie, um „aktuelle Wahlneigungen“ (Infratest dimap 2008) und „Projektionen“ (Forschungsgruppe Wahlen 2008a) von Wahlergebnissen auf deren Basis zu ermitteln. Dieser Verwendungszweck steht im Mittelpunkt der weiteren Ausführungen.

Zur Erstellung entsprechender Voraussagen von Wahlergebnissen werden als Basis die aggregierten Wahlabsichtsäußerungen der Probanden herangezogen. Allerdings bestehen hinsichtlich des genauen Prozedere institutsspezifische Unterschiede, die zum einen methodische Aspekte des Studiendesigns betreffen und zum anderen den Umgang mit der Datenaufbereitung. Da die publizierten Sonntagsfragen die mit Abstand am weitest verbreitete Form der Wahlprognose in Deutschland darstellen und diese in Abschnitt 3 empirisch

detailliert analysiert werden, wird auf diese institutsspezifischen Praktiken ausführlicher eingegangen als beispielsweise im Weiteren auf die weniger bedeutenden Wahlerwartungsfragen.

Das grundsätzliche *Erhebungsdesign* weicht heutzutage wesentlich nur beim IfD Allensbach ab. Entgegen der sonst etablierten Praxis der Verwendung von Zufallsstichproben gemäß der Empfehlungen des Arbeitskreises Deutscher Markt- und Sozialforschungsinstitute e. V. (ADM 1979), verwendet das IfD Allensbach Quotenstichproben (vgl. dazu die methodischen Erläuterungen in den Jahrbüchern des IfD: Noelle-Neumann 1976; Noelle-Neumann/Köcher 2002a; Noelle/Neumann 1956, 1957, 1965, 1967, 1974). Auch hinsichtlich der Befragungsart weicht das IfD Allensbach von der mittlerweile üblichen Praxis ab: Nach wie vor werden hier persönliche Interviews geführt, während die anderen Institute computergestützte telefonische Interviews (CATI) durchführen, wobei die Stichprobenziehung häufig nach einem *Random-Digit-Dialing*-Verfahren erfolgt (bei Infratest dimap etwa *Last-Two-Digit-Dialing*-Verfahren, vgl. Hilmer/Hofrichter 2002: 34).

Es ist davon auszugehen, dass die Ausschöpfungsquoten über die Institute hinweg und auch im Laufe der Zeit variieren. Wüst (1998) weist darauf hin, dass diese ein maßgebliches Qualitätskriterium einer Umfrage seien, noch wichtiger als die Höhe der Fallzahl (ähnlich: Neller 2005). Eine Angabe der Ausschöpfungsquoten seitens der Institute vermisst man zwar schmerzlich, doch über den betrachteten Zeitraum von 1949 bis 2005 ist von einer nicht unbeträchtlichen Variation auszugehen.

Vermutet werden kann dabei insbesondere ein stetiges Sinken der Ausschöpfungsquoten im Zeitablauf, auch wenn die Diskussion über die Entwicklung der Ausschöpfungsquoten im Zeitverlauf durchaus kontrovers geführt wird. Dies liegt nicht zuletzt an mangelnden Forschungsergebnissen hierzu oder auch differierenden Definitionen und jeweils unterschiedlich betrachteter Erhebungsmethoden. Anders (1985) sowie Converse und Traugott (1986) berichten bereits Mitte der 1980er Jahre aus Sicht der kommerziellen Meinungsforschungsinstitute im Rahmen von vorwiegend persönlichen Interviews von sinkenden Ausschöpfungsquoten und auch von der Heyde (2002) postuliert eine entsprechende Tendenz. Porst (1996) hingegen findet hierfür zumindest für die von der ZUMA betreuten Befragungen keine empirische Evidenz. Schnell (1997: 71ff.) wiederum belegt zum einen die allgemeine Tendenz sinkender Ausschöpfungsquoten

und großer Variationen über verschiedene Auftraggeber hinweg, kann aber zum anderen in Bezug auf telefonische Umfragen die sinkende Tendenz für die Politbarometerumfragen nicht belegen (ebd.: 116ff.). Neuere Studien weisen jedoch auch eher auf ein Absinken der Ausschöpfungsquoten hin (etwa Couper/De Leeuw 2003; Steeh et al. 2001).

Neben diesen stichprobenspezifischen Unterschieden sind im Hinblick auf die erhebungsbezogenen Unterschiede variierende *Frageformulierungen und -positionierungen* zu nennen. Im Politbarometer der Forschungsgruppe Wahlen wird zunächst die Absicht der Wahlteilnahme erfasst, dann werden diejenigen, welche eine Wahlbeteiligung nicht ausschließen, nach ihrer Wahlabsicht befragt. Im Forsa-Bus hingegen werden die Probanden zunächst nach ihrem vergangenen Wahlverhalten befragt und dann nach ihrer Wahlabsicht bei der zukünftigen Bundestagswahl, wobei hier keine explizite Frage zur Wahlbeteiligungsabsicht vorgeschaltet ist.¹¹ Die genaue Frageformulierung der anderen Institute ist nicht nachzuvollziehen, da diese ihre Fragebögen nicht veröffentlichen, es ist jedoch anzunehmen, dass die Praxis nicht wesentlich abweicht. Die Platzierung der Wahlabsichtsfrage innerhalb der Befragungen ist ebenfalls nur vereinzelt nachvollziehbar. Im Forsa-Bus 2005 sind den Codebuch-Angaben zufolge erst für den Befragten zentrale Themen und Probleme sowie die Partei, die seiner Mei-

11 Die genauen Frageformulierungen im Politbarometer lauten: „Wenn am nächsten Sonntag Bundestagswahl wäre, würden Sie dann zur Wahl gehen?“ – die möglichen Antwortkategorien waren: Ja, nein, weiß nicht bzw. auf jeden Fall zur Wahl gehen, wahrscheinlich zur Wahl gehen, wahrscheinlich nicht zur Wahl gehen, auf keinen Fall zur Wahl gehen oder habe bereits Briefwahl gemacht (07-09/94, 09-11/98, 09-10/02) und „Welche Partei würden Sie dann wählen?“ Die Liste der dann verlesenen Antwortmöglichkeiten variiert von Wahl zu Wahl, aber die Kernparteien werden in folgender Reihenfolge abgefragt: CDU (in Bayern CSU vorlesen), SPD, FDP, GRÜNE (ab 07/79) bzw. Bündnis 90/Grüne (ab 94), Republikaner (ab 05/88), PDS (ab 08/90), keine eindeutigen Nennungen werden folgenden Kategorien zugeordnet: weiß nicht (01/77-03/78, 06/80), keine Angabe/verweigert, keine Angabe/Befragter kann keine konkrete Partei nennen (94,02) oder trifft nicht zu (sofern zuvor Wahlteilnahme ausgeschlossen wurde) (Forschungsgruppe Wahlen o.J.).

Im Forsa-Bus 2005 lautete die Frage: „Und welche Partei würden Sie wählen, wenn am kommenden Sonntag Bundestagswahlen wären?“ Die Antwortkategorien (neben weiteren Parteien) waren: CDU, CSU, SPD, FDP, Grüne/Bündnis 90, PDS, Republikaner, bin nicht wahlberechtigt, gehe nicht wählen, weiß nicht, keine Angabe (Forsa o.J.).

Infratest dimap (2008) gibt als Frageformulierung nach der Wahlabsicht Folgendes an: „Welche Partei würden Sie wählen, wenn am kommenden Sonntag Bundestagswahl wäre?“. Jedoch bleiben die möglichen Antwortkategorien ebenso unklar wie die genaue Positionierung der Frage im Fragebogen und die Erfassung der Wahlbeteiligungsabsicht.

nung nach diese Probleme am ehesten lösen könne, und seine Kanzlerpräferenz erfasst worden, darauf hin das vergangene Wahlverhalten und dann erst die Wahlabsicht. Im Politbarometer wird zunächst das Bundesland, in dem der Befragte wahlberechtigt ist, erfasst sowie die Gemeindegröße, dann die vom Befragten als zentral erachteten Probleme, hierauf folgend die Wahlbeteiligungssowie Wahlabsicht und im Anschluss erfolgt die Erfassung des vergangenen Wahlverhaltens.

Diese Unterschiede in der Fragebogengestaltung können mitunter erhebliche Auswirkungen auf die Ergebnisse haben (vgl. etwa Atteslander 1980: 136ff.). Eine systematische Analyse von Fragereihenfolgeeffekten für Wahlabsichtsfragen im Kontext von Bundestagswahlen hat bisher nur Gibowski (1973) vorgelegt. Auf Grundlage eines *split-half*-Verfahrens mit Paneldaten, in denen sowohl die Wahlabsicht als auch das spätere Wahlverhalten erfasst wurde, belegt er, dass ein systematischer Zusammenhang zwischen der Platzierung der Wahlabsichtsfrage und den erhaltenen Stimmenanteilen für die CDU und SPD besteht und zwar in der Form, dass der Anteil an CDU-Nennungen mit einer späteren Fragestellung der Wahlabsicht zurückgeht. Weiterhin zeigt er, dass das politische Interesse eine maßgebliche Rolle spielt: Je später die Frage gestellt wird, desto größer sind die Abweichungen bezüglich der späteren Wahlergebnisse bei den politisch weniger Interessierten. Die praktische Folgerung aus Gibowskis (1973) Analyse wäre die Wahlabsichtsfrage möglichst früh zu stellen – die gängige Praxis, zumindest sofern die Reihenfolge der Fragen im Codebuch mit der tatsächlichen Fragereihenfolge übereinstimmt, weicht sowohl im Politbarometer als auch im Forsa-Bus hiervon ab. Allerdings betont Gibowski (2000), dass seinen Analysen zufolge zumindest die Rückerinnerungsfrage im Politbarometer immer nach der Wahlabsicht gestellt werde.

Juhász (1993) bemängelt entsprechend die fehlende methodische Forschung zu politischen Umfragen und vor allem zur Validität der Wahlabsicht und der Rückerinnerung sowie den bestehenden Wechselwirkungen. Er selbst kann auf Basis von Paneldaten der Wahlstudien zu den Bundestagswahlen 1983, 1987 und 1990 nur Teilantworten hierzu liefern, indem er der Frage nachgeht, inwieweit Wahlabsicht und Rückerinnerung wirklich das messen, was sie messen möchten. Er zeigt auf, dass im Durchschnitt lediglich 80 Prozent der Befragten stabile Angaben hinsichtlich ihrer Angaben zur Wahlabsicht und der Rückerinnerung an ihr vergangenes Wahlverhalten aufweisen. Die weiterführenden Analysen zeigen,

dass „Antworten auf die Wahlabsichts- und Rückerinnerungsfrage bei Wiederholungsbefragungen nicht zuletzt eine Funktion der Polarisierung der Bewertungen der politischen Parteien und der Veränderung dieser Einschätzungen zwischen den Befragungszeitpunkten“ (ebd.: 48) sind.

Hinsichtlich der Erhebungspraktiken der Wahlabsicht ist festzuhalten, dass Informationen, die über grundsätzliche Informationen über Stichprobenverfahren und die Fallzahl hinausgehen, unzulänglich dokumentiert sind. Wenig überraschend ist, dass zu den über das Zentralarchiv für empirische Sozialforschung zu beziehenden Dauerbeobachtungsstudien (Politbarometer und Forsa-Bus und seit Mai 2009 auch der DeutschlandTrend von Infratest dimap) die ausführlichsten Angaben vorliegen. Zumindest sind hier präzise Angaben zu den Stichprobenverfahren, Fallzahlen und auch Fragestellungen in den einzelnen Erhebungswellen zu finden. Gleichwohl ist erstaunlicherweise festzustellen, dass Forschungsergebnisse, die sich auf methodische Aspekte von Wahlabsichtsumfragen beziehen, kaum vorliegen, wengleich immer wieder mögliche Einflussfaktoren auf deren Güte angeführt werden (vgl. etwa Maurer 2008; Witterauf 2008). Diese Faktoren werden in Abschnitt 3 diskutiert.¹²

Neben diesen methodischen Differenzen der Erhebungspraktiken bestehen beträchtliche Unterschiede der Institute im Umgang mit der *Veröffentlichung der Daten* der Sonntagsfrage. Eng mit dieser Praxis ist die Frage danach verbunden, was mit der Wahlabsicht gemessen wird, und es ist festzustellen, dass sich je nach Institut ein erheblich unterschiedlicher Sprachgebrauch etabliert hat. Der Disput kreist dabei um die Frage, ob die Sonntagsfrage als „Prognose“ eines zukünftigen Wahlergebnisses interpretiert werden darf. Während in den Anfängen der politischen Meinungsforschung in Deutschland hierüber kaum Uneinigkeit bestand und auch seitens der Demoskopien von „Prognosen“ gesprochen wurde (vgl. etwa hierzu den Sprachgebrauch im Rahmen der angesprochenen Verbots-

12 Ein Aspekt, der nicht weiter aufgegriffen wird, da er in der deutschen Wahlforschung bisher nicht behandelt wurde, wird von Bassili (1993, 1995; Bassili/Bors 1997) untersucht. Er fand heraus, dass in Telefonsurveys die Reaktionszeit auf die Wahlabsichtsfrage als Indikator für die Stärke der Wahlabsicht aufgefasst werden kann, und er zeigt zudem, dass diese in Verbindung mit der Wahlabsicht für die Verbesserung von Voraussagen des späteren tatsächlichen Wahlverhaltens genutzt werden kann. Für die deutsche Forschung zu Antwortlatenzzeiten vgl. Urban und Mayerl (2007).

diskussion Blecha 1969; Fessel 1969; Hartenstein 1969), hat sich im Zeitablauf der Sprachgebrauch gewandelt.

Als ein Grund hierfür kann zunächst das medial heftig diskutierte „Wahlvorhersagedebakel“ (Eberlein 2001: 85) vor der Bundestagswahl 1965 angeführt werden. Vor der Wahl wurde von den Meinungsforschern lange ein Kopf-an-Kopf-Rennen von CDU/CSU und SPD gesehen. Die Ergebnisse der jeweils letzten Wahlabsichtsumfragen hinterlegten Emnid¹³ und das IfD Allensbach bei einem Notar, um Zweifel an „wahren“ Prognosen aufzuheben. Diese Voraussetzungen wurden erst am Wahlabend geöffnet. Entgegen aller vorherigen Äußerungen – auch in Interviews von Noelle-Neumann selbst – prognostizierte das IfD Allensbach einen deutlichen Vorsprung von CDU/CSU, während Emnid weiterhin ein Kopf-an-Kopf-Rennen sah. Dies führte zu einer anhaltenden Debatte insbesondere über Manipulationsabsichten seitens des IfD Allensbach im Vorfeld der Wahl (aus Emnid-Perspektive vgl. Eberlein 2001: 341ff.; zur Sicht des IfD vgl. Noelle-Neumann 1983; Noelle-Neumann 2001: 14ff.).

Als weiterer Grund kann die massive Zunahme an Meinungsumfragen gesehen werden und damit einhergehend die Etablierung einer Dauerbeobachtung politischer Meinungen (z. B. Politbarometer, DeutschlandTrend). Gerade in den Zeiten zwischen zwei Bundestagswahlen werden zwar regelmäßig Ergebnisse von Wahlabsichtsumfragen berichtet, diese können jedoch kaum als Prognosen aufgefasst werden. Da sich die Darstellung sowie das Instrument jedoch in der Vorwahlkampfzeit nicht ändern, wird eine Trennung zwischen Meinungserfassung und Voraussage zunehmend schwierig. So betonen Hilmer und Hofrichter (2002: 33), dass mit der Sonntagsfrage lediglich die „aktuelle politische Stimmung“ beziehungsweise die „aktuelle Stärke der Parteien“ zu erfassen sei. Als „Prognosen“ bezeichnet Infratest dimap explizit nur die für die ARD erstellte Voraussage, welche auf *exit-poll*-Befragungen beruht (ebd.: 35). Auch Liepelt (1966, 1996), Gibowski (2000) sowie Plischke und Rattinger (2009) heben hervor, dass erfasste Wahlabsichten nicht als Prognose aufgefasst werden können.

Da unstrittig ist, dass die erfassten Rohdaten der Wahlabsicht keine brauchbare Abschätzung des Wahlergebnisses in der Zukunft widerspiegeln, hat sich in

13 Der ehemalige Leiter des Emnid-Instituts für Politische Meinungsforschung distanziert sich nachdrücklich davon, dass Emnid Wahlprognosen erstellt hat – Eberlein (2001: 342) stellt dies als „Privatvergnügen“ des Emnid-Gründers von Stackelberg dar.

Deutschland die Praxis der *politischen Gewichtung* etabliert. Neben einer möglicherweise designbedingten Gewichtung (etwa aufgrund der Überrepräsentation der neuen Bundesländer oder der Verwendung geschichteter Stichproben), veröffentlichen alle Institute die Ergebnisse der Sonntagsfrage nur unter Einbezug weiterer Gewichtungsfaktoren. Diese tragen der Tatsache Rechnung, dass zahlreiche Probanden zum Zeitpunkt der Befragung noch unentschlossen sind bezüglich ihrer Wahlentscheidung und dass sich die Gruppe der Unentschlossenen nicht gemäß der bereits Entschlossenen auf die einzelnen Parteien verteilen. Weiterhin sollen durch die Umgewichtung bestenfalls alle weiteren denkbaren Effekte auf die Diskrepanz zwischen Wahlabsicht und späterem tatsächlichen Wahlverhalten berücksichtigt und entsprechend korrigiert werden.

Die genaue Vorgehensweise bei dieser „Veredelung“ (Liepelt 1996: 110) der Sonntagsfrage unterliegt dem Institutsgeheimnis und zu den zur Anwendung kommenden Verfahren gibt es nur vage Hinweise. Liepelt (ebd.) etwa unterscheidet danach, ob von der „intelligenten Umrechnung“ nur die Unentschlossenen oder das gesamte Sample betroffen ist. Ihm zufolge beziehen Infas und die Forschungsgruppe Wahlen nur die zum Zeitpunkt der Befragung Unentschlossenen mit ein, während das IfD Allensbach und Emnid alle Fälle politisch nachgewichten.

Eine zentrale Variable, die von allen Instituten bei dieser Gewichtung berücksichtigt wird, ist die Rückerinnerung an das Wahlverhalten bei der letzten entsprechenden Wahl (Krauß 1994). Ulmer (o.J.) behauptet in diesem Zusammenhang gar, dass die veröffentlichten Sonntagsfragen nicht mehr seien als eine Anpassung der „Resultate der letzten Wahlen Pi mal Daumen an die aktuelle Situation“.

Zumindest in jüngerer Zeit mit steigender Volatilität des Wahlverhaltens ist anzunehmen, dass die Bedeutung weiterer Faktoren wie Parteineigungen, Kandidatenpräferenzen oder die Leistungsbewertung einzelner Parteien bei der Umrechnung zunimmt. Inwieweit von welchen Instituten neben der Rückerinnerung weitere politische Variablen herangezogen werden, ist ebenso unbekannt wie die Kenntnis darüber, wie sich die Gewichtungsverfahren im Zeitverlauf gewandelt haben. Auch nicht abschließend zu beurteilen ist, ob die angewendeten Umrechnungsprozeduren theoretisch fundiert oder eher erfahrungsbasiert sind, auch wenn mit Furmaniak et al. (1975: 571) vermutet werden kann, dass die vorgenommenen Operationen vorwiegend auf „Faustregeln“ beruhen.

Sicher ist allerdings, dass alle Institute vor der Veröffentlichung entsprechende Nachgewichtungsverfahren anwenden, die sie nicht im Detail publik machen. Einzig von der Forschungsgruppe Wahlen ist bekannt, dass sie jeweils eine „Stimmung“ und eine „Projektion“ berichtet (Plischke/Rattinger 2009; Wüst 2003). Erstere gibt die gemessene Wahlabsicht wieder, während letztere Ergebnis umfassenderer politischer Nachgewichtung ist. Da mit dem Politbarometer zumindest für eine bedeutende Datenquelle, aus der sich publizierte Sonntagsfragen speisen, Rohdaten vorgelegt werden, kann immerhin für dieses Institut der Einfluss der Gewichtung auch im Zeitverlauf dargestellt werden (vgl. Abschnitt 3.4.1).

Angesichts der umfangreichen „geheimgehaltenen Kosmetik“ (Ulmer 1987a: 16) und des zumeist verschleierte Anteils Unentschlossener muss konstatiert werden, dass trotz aller Distanzierung der Institute die Sonntagsfrage als Prognoseinstrument gesehen werden muss und so medial kommuniziert wird. Glaubt man inferenzstatistischen Gesetzmäßigkeiten, wie es alle Institute vorgeben, so ist die Erfassung der aktuellen Stimmung das Ergebnis der Rohdaten (allenfalls gewichtet nach Designspezifika) auf Grundlage von Zufallsstichproben. Dies ist die einzige Möglichkeit der Schätzung von Populationsparametern innerhalb bestimmter, angebbarer Fehlerintervalle. Und sind zu einem bestimmten Zeitpunkt vor der Wahl ein Drittel der Befragten unentschlossen, so entspricht dies der aktuellen Stimmung. Auch Argumente, die darauf hinauslaufen, dass die gewichteten Sonntagsfragen eine Prognose für eine hypothetische Wahl am jeweils kommenden Sonntag seien und so gewählt werden würde, wenn an diesem unmittelbar bevorstehenden Wochenende Wahl wäre, laufen fehl, denn eine – auch von den Demoskopern immer wieder eingeforderte – Überprüfungsmöglichkeit ihrer Ergebnisse ist mit dieser Immunisierungsstrategie nicht möglich.¹⁴

14 Auf diesen Umstand weisen auch Plitschke und Rattinger (2009: 504) hin, wenngleich sie dennoch den Terminus „Prognose“ für die veröffentlichten Sonntagsfragen ablehnen. Sie unterscheiden – gemäß einer ASI-Richtlinie zur Veröffentlichung von Ergebnissen der Wahlforschung (Sahner 1998) – zwischen politischer Stimmung, Projektion und Prognose. Die politische Stimmung entspricht den Rohdaten der Wahlabsichtsumfrage. Als Projektionen sind demzufolge die veröffentlichten Sonntagsfragen zu bezeichnen. Hierunter sind „aktuelle Wählerpotentiale“ (ebd.: 99) zu verstehen, die aus einer politischen Gewichtung der geäußerten Wahlabsichten resultieren. Prognosen liegen der ASI zufolge nur vor, wenn Ergebnisse von Umfragen, welche „wenige Tage vor der Wahl“ durchgeführt wurden, „durch die Hinzuziehung externer Daten mittels einer Modellrechnung“ für die „Vorhersage zum Ausgang der bevorstehenden Wahl“ herangezogen werden (ebd.). Dieser Unterscheidung zwischen Projektionen und Prognosen wird hier aus den im Text genannten Gründen nicht gefolgt. Im Übrigen ist die Unterscheidung

Weiterhin erscheint aufgrund der gängigen Praxis die Angabe von Fehlergrenzen unsinnig, denn die berichteten Ergebnisse beruhen nicht auf den Rohdaten und erfüllen demnach nicht mehr die Voraussetzungen, um inferenzstatistische Verfahren anwenden zu können (vgl. Ulmer 1987a: 11). Diese Perspektive wird im Übrigen auch in den USA geteilt:

„By the end of an election campaign, published polls can be thought of as forecasts. If polls cannot be considered in this way, they would lack even entertainment value. If late-campaign polls are forecasts, then important questions are, how well do the polls do, and why do some polls forecast better than others?“ (Pickup/Johnston 2008: 270).

Neben der politischen Gewichtung hat vor allem deren Geheimhaltung Kritik hervorgerufen. So weisen etwa Voss et al. (1995) darauf hin, dass auch in den USA üblicherweise Gewichtungsverfahren eingesetzt werden, welche politische Variablen zur Umlegung der Unentschiedenen auf die einzelnen Parteien und/oder Kandidaten heranziehen. Allerdings seien diese Verfahren bekannt. Genau dies hat im Übrigen auch zur Kritik des Gallup-Instituts an der Praxis der deutschen politischen Meinungsforschung geführt (Faz.Net 2005).

Hinsichtlich der Praxis von Wahlabsichtsfragen bleibt festzuhalten, dass einhellig die Auffassung vertreten wird, dass erfasste intendierte Verhaltensabsichten keine adäquate Annäherung an tatsächliches Verhalten sein kann – schon gar nicht, um Wahlergebnisse vorauszusagen. Dennoch hat sich die Sonntagsfrage als zentrales Prognoseinstrument – unabhängig von der semantischen Beschreibung der Meinungsforschungsinstitute – in der Öffentlichkeit durchgesetzt. Der Grund hierfür ist nicht zuletzt in der Praxis der politischen Gewichtung der Rohdaten durch die kommerziellen Meinungsforschungsinstitute zu sehen. Bezweifelt werden kann zudem nicht, dass entsprechende Umrechnungen im Hinblick auf die Güte der Sonntagsfrage als Prognoseinstrument sinnvoll sein können – eine systematische Analyse dessen steht jedoch noch aus und wird in Abschnitt 3 präsentiert.

dung semantischer Natur, denn auch Plischke und Rattinger (2009) „prüfen“ die Plausibilität der „Projektionen“ anhand des tatsächlichen Wahlergebnisses.

2.1.4 *Wahlerwartungsfragen*

Mit Wahlerwartungsfragen soll in Surveys die Vermutung der Probanden erfasst werden, wie die einzelnen Parteien abschneiden. Anstatt nach dem eigenen beabsichtigten Handeln zu fragen, wird eine Einschätzung darüber eingeholt, welchen Stimmenanteil einzelne Parteien nach Auffassung des Probanden erreichen werden, welcher Kandidat Kanzler wird, welche Koalition die Regierung stellen wird etc. In Deutschland wurden derartige Wahlerwartungsfragen zur Voraussage von Wahlergebnissen bisher nicht systematisch eingesetzt. Allerdings hat das IfD Allensbach zumindest in der Vergangenheit immer wieder die Erwartung, welche Partei stärkste Kraft im Bundestag wird, abgefragt (vgl. Noelle-Neumann 1976: 134; Noelle-Neumann/Köcher 2002a: 791; Noelle/Neumann 1956: 176). Die Daten hierzu liegen jedoch nicht vor und es ist auch nicht bekannt, ob diese Wahlerwartungsfragen für Vorhersagezwecke seitens des IfD Allensbach genutzt wurden. Ebenfalls wurden in der German Longitudinal Election Study (GLES) zur Bundestagswahl 2009 im Rahmen der Vorwahlbefragungen die Koalitionserwartungen erhoben (Rattinger et al. 2009).

Idee hinter dem Einsatz von Wahlerwartungsfragen als Vorhersageinstrument ist, dass eine geringere Verzerrung zu erwarten sei, da hypothetisch geäußertes Verhalten, wie im Fall der Sonntagsfrage, keineswegs mit dem später tatsächlich gezeigten Verhalten übereinstimmen muss. Zudem ist zu erwarten, dass auch die zum Zeitpunkt der Befragung Unentschiedenen eine Abschätzung zum Ausgang der Wahl abgeben können. Hiermit würde man folglich zwei zentrale Probleme von Wahlabsichtsumfragen lösen: Deren hypothetischen Charakter und damit ihre Brauchbarkeit erst relativ kurz vor dem Wahltermin sowie die notwendige politische Gewichtung der Daten.

Als zentraler Vorteil der Wahlerwartungsfragen gegenüber den vergleichend betrachteten Wahlabsichtsumfragen wird gesehen, dass die Wahlerwartungsfragen bereits zu einem früheren Zeitpunkt vor dem Wahltermin die Genauigkeit der späteren Wahlabsichtsumfragen aufweisen. Dementsprechend sehen Lewis-Beck

und Tien (1999: 183) Wahlerwartungsfragen als „a promising micro alternative“ insbesondere zu den noch zu erörternden Wahlbörsen (Abschnitt 2.2) an.¹⁵

Bei dieser Einschätzung ist zu berücksichtigen, dass sich die Autoren auf US-Präsidentschaftswahlen beziehen und hier im Allgemeinen nur der Erfolg von zwei Kandidaten abgeschätzt werden muss. Insofern ist die Übertragbarkeit des Verfahrens auf das deutsche Wahlsystem durchaus skeptisch zu sehen. Während die bisher betrachteten Fragen jeweils ausschließlich nach dem erwarteten Wahlsieger fragen – und auch die vom IfD Allensbach eingesetzten Fragen nur die Erwartung, welches die stärkste Partei werden wird, abfragen –, interessieren in Deutschland die jeweiligen Stimmenanteile der einzelnen Parteien. Ob eine Voraussage dieser – eine ungleich komplexere Fragestellung – mit Hilfe dieses Instruments möglich ist, muss an dieser Stelle unbeantwortet bleiben.

2.1.5 Wahltagsbefragungen

Als letzte umfragebasierte Version von Voraussagen sollen noch Wahltagsbefragungen, die so genannten *exit polls* angesprochen werden. Der zentrale Unterschied zu Wahlabsichts- und Wahlerwartungsfragen ist, dass die Wahltagsbefragungen nicht vor, sondern unmittelbar nach dem Urnengang durchgeführt werden, und demnach tatsächliches Verhalten anstatt prospektiver Verhaltensabsichten oder Erwartungen erfassen. Wahltagsbefragungen bilden in Deutschland die Grundlage der „18 Uhr-Prognosen“ und stellen ein zentrales Element der Berichterstattung am Wahlabend dar.

Die Geschichte der Wahltagsbefragungen ist deutlich kürzer als etwa die der Sonntagsfrage. Der erste bekannte *exit poll* wurde vom Nachrichtensender CBS im Rahmen der Gouverneurswahl in Kentucky, USA im Jahr 1967 durchgeführt (Levy 1983). Allerdings werden in den USA die Ergebnisse Levy (ebd.: 65) zufolge weniger als Prognoseinstrument genutzt als vielmehr zur detaillierten Analyse des Wählerverhaltens. In Deutschland hingegen wird das Instrument der Wahltagsbefragung vornehmlich zu Prognosezwecken eingesetzt, wenn auch

15 Möglicherweise eröffnet die Anwendung von Wahlerwartungsfragen in Panelstudien, welche Vor- und Nachwahlbefragungen umfassen zudem einen Ansatzpunkt zur Untersuchung der bereits angesprochenen *bandwagon*-, *underdog*- oder Defätismus-Effekte (vgl. Abschnitt 2.1.2.1).

Hilmer (2008: 105) die Wichtigkeit der erhobenen Daten für die deskriptive Analyse des Wahlverhaltens hervorhebt. Erstmals erprobt wurde es durch Infas im Jahr 1976, die erste Veröffentlichung von Voraussagen von Wahlergebnissen auf Grundlage von *exit polls* erfolgte ebenfalls durch Infas im Jahr 1978 im Rahmen der Landtagswahlen in Hamburg und Niedersachsen (Hilmer/Hofrichter 2001: 445).

Ebenso wie bei der Sonntagsfrage schloss sich auch an die Praxis der Wahltagsbefragung eine rege Debatte an, die insbesondere um die Frage kreiste, ob möglicherweise – trotz eines Veröffentlichungsverbot der Ergebnisse bis zur Schließung der Wahllokale – das Abstimmungsverhalten der Wähler durch ein Bekanntwerden der Umfrageergebnisse vor 18 Uhr beeinflusst werden könnte (Schultze 1980: 75). Gefordert wurde von den Kritikern nicht zwingend ein völliges Verbot der Befragungen, jedoch ein Verbot der Auswertung der Fragebögen vor 18 Uhr – ähnlich wie dies für die Auszählung der Stimmen gilt. In Folge dieser Debatte setzte die ARD, welche die Infas-Studien in Auftrag gab, die Verwendung von Wahltagsbefragungen aus und erst mit der Wiedervereinigung konnte sich die Methode letztlich durchsetzen und gehört seitdem zum Höhepunkt des Wahlabends – zumindest aus Sicht der Meinungsforscher. Zurückzuführen ist diese Etablierung im Rahmen der ersten gesamtdeutschen Wahl darauf, dass diese Wahl mit besonderem Interesse verfolgt wurde, jedoch für das ostdeutsche Elektorat kaum wahlsoziologische Informationen vorlagen. Die Wahltagsbefragungen versprachen hier, diese Lücke mit präzisen Daten zu schließen (vgl. Hilmer/Hofrichter 2001).

In methodischer Hinsicht stellen die Wahltagsbefragungen für Umfrageinstitute einen enormen logistischen Aufwand dar, der gerechtfertigt wird durch die Erwartung möglichst präziser Schätzungen der Stimmenanteile der Parteien. Diese Erwartung fußt insbesondere auf drei Merkmalen der Wahltagsbefragung: Erstens werden ausschließlich Wähler befragt, während bei Vorwahlumfragen Wahlberechtigte die Grundgesamtheit darstellen und deshalb besondere Verzerrungen aufgrund der Unsicherheit über die Wahlbeteiligung zu erwarten sind; zweitens wird tatsächliches Verhalten abgefragt, das unmittelbar vor der Befragung vollzogen wurde, und insofern kann mit validen Angaben hierüber gerechnet werden; drittens arbeiten die Institute mit großen Stichproben zur Reduktion des Stichprobenfehlers und können hohe Ausschöpfungsquoten realisieren (Hilmer 2008: 95; Hilmer/Hofrichter 2001: 447).

Die Stichprobenziehung erfolgt in der Regel zweistufig. Zunächst werden nach einem, zumeist geschichteten, Zufallsverfahren Stimmbezirke ausgewählt (etwa 400 bei Bundestagswahlen, 120-200 bei Landtagswahlen). In den ausgewählten Stimmbezirken werden schriftliche Befragungen von zufällig ausgewählten Wählern direkt nach ihrer Stimmabgabe durchgeführt. Die Probanden werden persönlich angesprochen und erhalten einen kurzen Fragebogen (maximal zwei DIN A4-Seiten), den sie in einer „Wahlkabine“ ausfüllen und danach in eine Box werfen können. Insgesamt werden bei Bundestagswahlen circa 20.000, bei Landtagswahlen zwischen 5.000 und 10.000 Interviews durchgeführt. Per Telefon werden von den Interviewern vor Ort die Antworten auf den einzelnen Fragebögen an die Zentrale übermittelt, die alle eingehenden Daten sammelt und auswertet. Der Erstellung der Voraussage geht ein mehrstufiges Gewichtungungsverfahren voraus, das Designaspekte berücksichtigt (unterschiedliche Auswahlwahrscheinlichkeiten der Stimmbezirke und gegebenenfalls der Probanden). Darüber hinaus wird nach soziodemographischen Faktoren auf Grundlage der repräsentativen Wahlstatistik gewichtet und häufig auch nach der Rückerinnerung an das Wahlverhalten bei der letzten Wahl. Schließlich werden die nicht im Sample enthaltenen Briefwähler berücksichtigt. Das genaue Gewichtungsverfahren wird von den Instituten – wie beim *Redressment* der Sonntagsfragen – nicht veröffentlicht (vgl. Hilmer/Hofrichter 2001: 447ff.).

2.1.6 *Zum prognostischen Wert von umfragebasierten Vorhersagen*

In den vorherigen Abschnitten wurden drei verschiedene umfragebasierte Ansätze zur Voraussage von Wahlergebnissen behandelt: Wahlabsichtsfragen, Wahlerwartungsfragen sowie Befragungen am Wahltag zum Wahlverhalten. Diese werden nun vor dem Hintergrund des Prognosebegriffs, wie er in Abschnitt 1.2 präsentiert wurde, bewertet. Demnach sind insbesondere drei Aspekte zu prüfen. Erstens muss die Frage aufgeworfen werden, ob das Explanandum erst nach der Äußerung der Prognose eintritt. Zweitens ist für wissenschaftliche Prognosen zentral, dass auf einen korrekten kausalen Mechanismus verwiesen wird. Schließlich ist drittens zu prüfen, ob die Anfangsbedingungen korrekt erhoben wurden. Im Rahmen der Diskussion des Prognosebegriffs wurde zudem aufgezeigt, dass die Veränderung der Randbedingungen zwischen Zeitpunkt der Prognose und

dem Eintritt des Explanandums korrekt spezifiziert werden muss – es wurde bereits in Abschnitt 1.2 eine Abschwächung dieser Anforderung hin zu konditionalen Prognosen diskutiert, weshalb von diesem Kriterium im Folgenden abgesehen wird.

An dieser Stelle soll noch einmal darauf hingewiesen werden, dass im Rahmen dieser methodologischen Bewertung nur Aussagen darüber getroffen werden, ob die behandelten Ansätze die genannten Kriterien erfüllen und somit als wissenschaftliche Prognose klassifiziert werden können. Über die etwaige empirische Prognosegüte ist damit noch nichts ausgesagt.

Betrachtet man zunächst die *Wahlabsichtsfragen*, dann ist in Bezug auf das erste Kriterium – der Äußerung der Prognose vor Eintritt des Explanandums – festzuhalten, dass die Sonntagsfragen dieses erfüllen. In der Regel werden mittlerweile die Wahlabsichten im Rahmen der politischen Dauerbeobachtung regelmäßig erfasst und zeitnah zur Befragung veröffentlicht. Die publizierte Wahlabsicht bezieht sich hierbei auf die anstehende Wahl.

In Bezug auf die Forderung nach einem Verweis auf einen kausalen Mechanismus ist zunächst festzuhalten, dass die Logik hinter der Stimmenanteilermittlung für die einzelnen Parteien eine einfache Aggregationsregel ist. Allerdings werden die Wahlabsichtsfragen vor ihrer Veröffentlichung nach einem von Institut zu Institut unterschiedlichen und nicht veröffentlichten Verfahren gewichtet (vgl. Abschnitt 2.1.3), um so eine „veredelte intuitive“ (Naschold 1971: 27) Prognose zu erhalten. Zu vermuten ist nach allem, was über die Gewichtungsverfahren bekannt ist, dass hier zumindest auch theoretische Überlegungen mit einfließen und somit möglicherweise tatsächlich auf Mechanismen rekurriert wird. So ist in Bezug auf das individuelle Wahlverhalten festzuhalten, dass zumindest der sozialpsychologische Ansatz und die soziologischen Ansätze empirisch vielfach überprüft wurden und durchaus eine bis heute robuste Erklärungskraft aufweisen. Insofern könnte man diese theoretischen Ansätze durchaus als Ausgangspunkt für die Ableitung von wissenschaftlichen Prognosen heranziehen. Die vorgenommenen Gewichtungen sind jedoch nicht nachvollziehbar. Ein zentrales Postulat wissenschaftlichen Arbeitens ist die Gewährleistung der intersubjektiven Nachprüfbarkeit. Dies ist hier verletzt und somit können trotz der Unsicherheit bezüglich der Erfüllung der methodologischen Anforderungen die Sonntagsfragen in der bisherigen Praxis nicht als wissenschaftliche Prognosen

aufgefasst werden. Es wäre jedoch durchaus denkbar, die erhobene Wahlabsicht in eine Prognose von Wahlergebnissen zu integrieren.

Rattinger und Ohr (1989) haben hierzu einen Vorschlag unterbreitet, welcher darauf abzielt zumindest zwei wesentliche Fehlerquellen von Voraussagen auf Basis von Wahlabsichtsumfragen – Unentschlossene sowie Probanden, die nicht gemäß ihren „wahren“ Präferenzen antworten – mit Hilfe von analytisch hergeleiteten Gewichtungen zu korrigieren. Auch wenn die ersten präsentierten empirischen Befunde für diesen Ansatz sprechen, steht zum einen eine systematische Analyse des „Prognosemodells“ von Rattinger und Ohr (1989: 286) noch aus und zum anderen mangelt es auch ihrem Ansatz an einer theoretischen Fundierung. Diese müsste neben Ansätzen zum Wählerverhalten insbesondere eine – empirisch gut geprüfte – Theorie des Befragtenverhaltens berücksichtigen, um den bestehenden methodischen Problemen (Non-Response, Falschaussagen etc.) über die ad-hoc-Postulate von Rattinger und Ohr (1989) hinaus zu begegnen. Hiermit werden die bei Wahlabsichtsfragen weiterhin bestehenden Probleme hinsichtlich einer korrekten Erfassung der Randbedingungen offensichtlich – zumindest die Wahlabsicht zu dem jeweiligen Befragungszeitpunkt müsste richtig erfasst werden. Aufgrund etwa von systematischem Non-Response ist hiervon jedoch nicht eindeutig auszugehen. Dies wäre grundsätzlich bei Stützung auf Gesetzmäßigkeiten zum Befragtenverhalten jedoch ebenfalls zu berücksichtigen.

Über die Argumente von Rattinger und Ohr (ebd.) hinaus wäre eine theoriebasierte Benennung von Einflussfaktoren und deren Stärke notwendig, welche die Abweichungen von der Verhaltensabsicht zum tatsächlichen Verhalten – in Abhängigkeit des Zeitraumes bis zur tatsächlichen Wahlhandlung – erklären. Da weder eine Theorie des Befragtenverhaltens vorliegt, die empirisch fundiert wäre, noch eine Theorie, die präzise Vorhersagen erlaubt, wann eine Verhaltensabsicht zur tatsächlichen Handlung führt, erscheint es jedoch insgesamt wenig fruchtbar, Wahlabsichtsfragen zu wissenschaftlichen Prognosezwecken einzusetzen.

Im Hinblick auf das dritte Kriterium, der korrekten Erhebung der Randbedingungen, ist auf mit Befragungen einhergehende Probleme zu verweisen. Eine korrekte Erfassung der Wahlabsicht wäre streng genommen nur unter der Abwesenheit von Non-Response, sozialer Erwünschtheit und der gerade im Bereich politischer Fragen als einflussreich identifizierten Reihenfolgeeffekte (Gibowski 1973) gegeben. Insofern ist auch in Bezug auf dieses Kriterium zumindest von Verletzungen der Anforderung auszugehen, weshalb insgesamt Wahlabsichtsfra-

gen im Kontext der bisherigen Praxis aus wissenschaftstheoretischer Perspektive nicht als wissenschaftliche Prognosen aufgefasst werden können.

Wahlerwartungsfragen wurden bisher zwar kaum eingesetzt, jedoch erscheinen sie insgesamt aus wissenschaftstheoretischer Perspektive viel versprechender als Wahlabsichtsfragen. Mit Blick auf die zeitliche Reihenfolge ist festzuhalten, dass dies bei Anwendung einzuhalten wäre, jedoch wie bei Wahlabsichtsfragen keine Hürde darstellt.

Eine theoretische Fundierung erscheint zudem nahe liegender als bei den Wahlabsichten. So wäre zu prüfen, ob Surowieckis (2005) Ansatz geeignet ist, hinreichend präzise Bedingungen zu formulieren, die erfüllt sein müssen, um die „Weisheit der Vielen“ auch für die Prognose von Wahlergebnissen zu nutzen. Demzufolge könnte man erwarten, dass gerade auf Grundlage von Zufallsstichproben durchgeführte Wahlerwartungsfragen besonders erfolgreich sein sollten, da dem Ansatz zufolge eine heterogen zusammengesetzte Gruppe von Personen effizienter als Experten Informationen zu aggregieren vermag. So zeigt sich empirisch, dass insbesondere bei Schätzaufgaben, also bei Vorliegen unvollständiger Informationen, diese Überlegung zutrifft. Bedingungen hierfür sind Surowiecki (ebd.: 47) zufolge, dass jeder Einzelne unabhängig entscheidet und handelt, die Gruppe groß und heterogen ist und schließlich, dass jeder darauf vertrauen kann, dass seine Meinung wirklich zählt.

Zumindest das zweite Kriterium dürfte unstrittigerweise durch entsprechende Umfragen erfüllt werden. Hinsichtlich des ersten Aspekts sind jedoch Zweifel angebracht. Angesichts der medialen Präsenz von Wahlabsichtsumfragen und auch Experteneinschätzungen scheint die wesentliche Bedingung der unabhängigen Einschätzung des Probanden nicht zu gewährleisten zu sein. So hat Brettschneider (1992, 2000) gezeigt, dass Wahlumfragen nicht nur wahrgenommen werden, sondern auch die Koalitionserwartung beeinflussen.

Das IfD Allensbach hat sich mit der Erfassung von Wahlerwartungen auseinandergesetzt, sah deren Nutzen jedoch nicht in der Möglichkeit der Wahlprognose, zumal Noelle-Neumann (2001: 19) darauf hinweist, dass im Aggregat Wahlabsichten deutlich stabiler seien als Erwartungen des Wahlausgangs und die Schätzungen des Meinungsbilds der öffentlichen Meinung oft nicht mit den aus Erwartungsfragen ermittelten Einschätzungen übereinstimmen würde (ebd.: 317). Jedoch könnten die Befragten sehr gut Veränderungen im Meinungsbild wahrnehmen, ob also etwa bestimmte Meinungen zu- oder abnehmen.

Auch der dritte Punkt, die Bedingung der glaubhaften Vermittlung, dass die geäußerte Meinung zählt, ist sicher ebenfalls nicht unproblematisch, könnte jedoch möglicherweise durch eine entsprechende Ansprache im Rahmen der Befragung zumindest näherungsweise sichergestellt werden.

In Bezug auf das Kriterium der richtigen Erhebung der Randbedingungen ist ebenso wie bei den Wahlabsichtsfragen auf die mit der Surveyforschung einhergehenden Probleme zu verweisen. Insgesamt bleibt also festzuhalten, dass Wahlerwartungsfragen vielleicht durchaus ein prognostisches Potenzial zugesprochen werden kann – und auch entsprechende Befunde aus den USA deuten dies an. Insofern wäre es sicher wünschenswert, wenn entsprechende Fragen in Zukunft in den einschlägigen Umfragen getestet und gegebenenfalls berücksichtigt würden.

Den Meinungsforschungsinstituten zufolge stellen *Wahltagsbefragungen* das präziseste Prognoseinstrument von Wahlergebnissen dar. Kennzeichen einer wissenschaftlichen Prognose hierfür ist, dass die Vorhersage vor dem Eintritt des Ereignisses ausgesprochen wird. Dieses Kriterium ist bei *exit polls* verletzt, da die Wahl bereits stattgefunden hat. Die Meinungsforscher nutzen hier nur aus, dass sie ihre Befragungen nach dem tatsächlichen Wahlverhalten bereits auswerten dürfen, während die Wahlbüros noch geöffnet haben und die Stimmauszählung erst mit deren Schließung beginnen darf. Bestünde hier eine andere rechtliche Regelung, könnte man das Wahlergebnis auf Grundlage von Hochrechnungen bereits zu einem früheren Zeitpunkt präzise bestimmen.

Abgesehen hiervon ist der Wert der 18-Uhr-Prognose fraglich. Lewis-Beck (1985) nennt als ein zentrales Kriterium zur Evaluation von Prognosen deren zeitlichen Vorlauf (*lead*). Je früher man vor einer Wahl eine präzise Prognose liefern kann, desto höher sei ihr Informationsgehalt. Dementsprechend dürftig fällt die Bilanz der *exit polls* aus, da sie über einen *lead* von nur wenigen Stunden verfügen. Entsprechend kommt Lewis-Beck (ebd.: 56) in Bezug auf den prognostischen Wert von Wahltagsbefragungen zu dem Schluss: „This is not election forecasting; rather, it is election reporting as it happens“.

Der betriebene Aufwand der Institute im Rahmen der Wahltagsbefragungen ist jedoch dennoch nachvollziehbar, denn zum einen ist die Genauigkeit der jeweiligen „18 Uhr-Prognose“ ein Signal für besonders saubere methodische Arbeit und sie ist wohl ein wesentliches Kriterium bei der Verlängerung der Verträge mit den Fernsehanstalten. Zum anderen liefern die Wahltagsbefragungen

mehr Informationen als nur die Voraussage des Wahlergebnisses. Im Rahmen der Befragung werden auch zentrale soziodemographische Angaben erfasst sowie einige Wahlentscheidungsgründe erfragt. Diese sind Teil der Wahlanalysen, die bereits am Abend vorliegen und für Medien, Politiker, Parteien und Öffentlichkeit gleichermaßen einen großen Informationsgehalt aufweisen.

In Bezug auf die beiden anderen Kriterien ist festzuhalten, dass bei Wahltagsbefragungen tatsächliches Verhalten erfasst wird, welches für die Prognose aggregiert wird. Ebenfalls kommen – ähnlich wie bei der Sonntagsfrage – institutsspezifische und nicht bekannte Gewichtungungsverfahren zum Einsatz, welche möglicherweise auf empirisch bestätigten Gesetzmäßigkeiten beruhen. Verletzt wird aber auch hier die intersubjektive Nachprüfbarkeit des Verfahrens

Aufgrund des hohen methodischen Aufwands und den bekannten Berichten über die Praxis der Wahltagsbefragungen muss in Bezug auf das dritte Kriterium, der korrekten Erfassung der Randbedingungen, von nur geringen Verzerrungen, insbesondere im Vergleich zu Wahlabsichts- und Wahlerwartungsfragen, im Rahmen von Vorwahlbefragungen ausgegangen werden.

Zusammenfassend ist festzuhalten, dass keiner der angesprochenen umfragebasierten Ansätze die Anforderungen an wissenschaftliche Prognosen erfüllt. Während bei Sonntagsfragen insbesondere das Gebot intersubjektiver Nachprüfbarkeit der politischen Gewichtung und damit der Bezug auf etwaige Gesetzmäßigkeiten nebulös bleiben, verletzen *exit polls* die Anforderung der zeitlichen Abfolge von Explanans und Explanandum. Wahlerwartungsfragen stellen in diesem Kontext die vielversprechendste Alternative dar, da mit Surowiecki eine theoretische Fundierung möglich erscheint, jedoch werden Wahlerwartungen bisher nicht zur Vorhersage von Wahlergebnissen eingesetzt.

2.1.7 Vorliegende Evidenz zur Güte von umfragebasierten Vorhersagen

Die Frage nach der Güte von umfragebasierten Vorhersagen wurde in Deutschland bisher nur selten systematisch angegangen. Angesichts der medialen Präsenz von *Wahlabsichtsumfragen* und der dokumentierten Zunahme der Bedeutung von demoskopischen Umfrageergebnissen in der Vorwahlzeit (vgl. Bretschneider 1996, 2000, 2005b) überrascht es, dass bisher nur wenige empirische Befunde zur Validität dieser vorgelegt wurden, ja eine überzeugende systema-

tische Analyse bisher völlig fehlt (vgl. Hofmann-Göttig 2005). In den USA hingegen ist die stetige Überprüfung der Brauchbarkeit von Wahlabsichtsfragen für die Voraussage von Wahlergebnissen aus empirischer Sicht fester Bestandteil der wissenschaftlichen Diskussion. Neben zahlreichen – hier nicht eingehender thematisierten – deskriptiven Auswertungen zur Voraussagegenauigkeit (Buchanan 1986; Converse/Traubgott 1986; Erikson/Sigelman 1995; Erikson/Wlezien 2008; Lewis-Beck 1985; Mitofsky 1998; Mosteller et al. 1949; Pickup/Johnston 2008; Roper 1984, 1986; Traubgott 2001, 2005) liegen auch einige Arbeiten zu Einflussfaktoren auf diese vor. Zentrale Befunde hierzu werden in Abschnitt 3 erörtert.

Mit Bezug auf Deutschland hat Kaase (1973: 146) im Kontext der Analysen der Bundestagswahl 1972 Hinweise zur Güte der Sonntagsfrage vorgelegt. Er vergleicht für die Bundestagswahlen 1961 bis 1972 jeweils das Ergebnis einer Wahlabsichtsfrage (Rohdaten) maximal zwei Wochen vor der Wahl mit dem späteren Wahlergebnis. Hierbei schwanken die absoluten Abweichungen von 0,1 Prozentpunkten (sonstige Parteien bei der Bundestagswahl 1972) bis hin zu 9,7 Prozentpunkten (SPD, Bundestagswahl 1969). Furmaniak et al. (1975) folgern hieraus bereits, dass die ungewichtete Wahlabsicht keinesfalls als Prognoseinstrument brauchbar ist.

Eine weitere Analyse hat Radtke (1977) vorgelegt. Er möchte auf Grundlage veröffentlichter Sonntagsfragen die „Prognosekompetenz der empirischen Wahlforschung“ prüfen. Dazu vergleicht er insgesamt 60 Wahlabsichtsumfragen zu den sieben Bundestagswahlen von 1953 bis 1976 mit dem jeweiligen Wahlergebnis. Dazu greift Radtke (ebd.) zum einen auf verfügbare Rohdaten zurück, betrachtet jedoch auch die von den Befragungsinstituten publizierten Voraussagen, welche den einschlägigen Medien entnommen wurden. Zunächst ist festzuhalten, dass die in den Rohdaten gefundene absolute Abweichung in Prozentpunkten im Schnitt deutlich größer ist als bei den publizierten Wahlabsichten. Im Hinblick auf die Rohdaten ist weiterhin interessant, dass der Anteil an CDU/CSU-Wählern durchgängig unterrepräsentiert ist – Gibowski (1973) führt dies auf Fragereihenfolgeeffekte zurück. In diesem Zusammenhang zentral ist jedoch die Performanz der publizierten Daten: Diesbezüglich berichtet Radtke (1977) eine durchschnittliche Abweichung über alle sieben Wahlen von 1,1 Prozentpunkten für die CDU/CSU und 1,8 Prozentpunkten für die SPD. Diese Befunde wertet Radtke durchaus als Erfolg der Demoskopien, denn es sei den

„Wahl-Prognostikern gelungen, den durchschnittlichen Prognosefehler entsprechend dem zu erwartenden reinen statistischen Irrtumsfehler zu reduzieren“ (Radtko 1977: 677). Gleichwohl kritisiert er die beschränkten Analysemöglichkeiten aufgrund fehlenden Zugangs zu den Rohdaten und der Nicht-Offenlegung der Gewichtungungsverfahren, welche letztlich einer Weiterentwicklung der Wahlforschung im Wege stehen würden. Das gefolgerte Plädoyer für eine größere Transparenz der kommerziellen Wahlforschung verhallte bisher ungehört beziehungsweise ihm wurde zumindest nicht Folge geleistet.

Nachdem die empirische Beschäftigung mit der Güte der Sonntagsfragen lange vernachlässigt wurde, erfahren Wahlprognosen zumindest seit der Bundestagswahl 1998, welche die erste Abwahl einer amtierenden Regierung zur Folge hatte, wieder verstärkte Aufmerksamkeit. Gibowski (2000) greift einige Vorwahlumfragen der Institute Emnid, Forsa, Forschungsgruppe Wahlen, Infratest dimap und des IfD Allensbach heraus und vergleicht diese mit dem amtlichen Zweitstimmenergebnis. Dabei ist festzustellen, dass zwar alle Institute die SPD als stärkste Partei sehen, jedoch kein Institut den Stimmenanteil der CDU/CSU unter 36 Prozent sah (tatsächlich: 35,2 Prozent) und die Stimmenanteile der kleinen Parteien in der Tendenz unterschätzt wurden.

Eine systematischere Analyse auf Grundlage einer breiten Datenbasis hat Antholz (2001: 20ff., 325ff.) in seiner Dissertationsschrift angefertigt. Er hat aus verschiedenen Publikationen 479 publizierte Sonntagsfragenergebnisse zu 142 der insgesamt 14 Bundes- und 159 Landtagswahlen von 1949 bis 2000 zusammengetragen – die Daten decken folglich 82 Prozent der Wahlen in diesem Zeitraum ab. Da einige der von ihm erhobenen Sonntagsfragenergebnisse unvollständig waren, können 454 Wahlabsichtsumfragen ausgewertet werden. Auch wenn es keine gesicherte Erkenntnis darüber gibt, wie viele Wahlabsichtsumfragen veröffentlicht wurden, so stellt die von Antholz zusammen getragene Datenbasis die umfassendste dar, welche bisher vorgelegt wurde. Zudem deckt seine Recherche zur Datenerhebung die wesentlichen überregionalen und regionalen Medien ab, weshalb insgesamt eine hohe Datenqualität unterstellt werden kann (vgl. dazu detaillierter Abschnitt 3.3.1).

Über alle erfassten Wahlprognosen hinweg berechnet Antholz (2001: 98) eine durchschnittliche absolute Abweichung von zwei Prozentpunkten. Dieser erste Befund wird im Folgenden stärker differenziert und der Autor berichtet, dass Ergebnisse von Bundestagswahlen im Durchschnitt etwas besser vorausge-

sagt werden als diejenigen von Landtagswahlen (1,6 versus 2,2 Prozentpunkte). Weiterhin kommt der Autor zu dem Schluss, dass „die durchschnittliche absolute Abweichung der Wahlprognosen ... sich in den vergangenen 50 Jahren kaum [verändert hat]“ (ebd.: 100). Einschränkend muss erwähnt werden, dass sich dieser Befund auf eine Auswertung stützt, welche die jeweilige Abweichung für einzelne Jahrzehnte einander gegenüberstellt.

Weiterhin prüft Antholz (ebd.: 112ff.) die Trefferquote bei kleinen und großen Parteien, indem er hypothetische Fehlerintervalle für den Fall von 1.000 Befragte umfassenden Zufallsstichproben berechnet und von einem Stimmenanteil von 40 Prozent für große und acht Prozent für kleine Parteien ausgeht. Dies ergibt hypothetische Fehlerintervalle von $\pm 1,7$ Prozent für kleine und $\pm 3,1$ Prozent für große Parteien. Dabei zeigt sich, dass im Schnitt für kleine Parteien zu niedrige und für große Parteien zu hohe Stimmenanteile vorhergesagt werden und das insgesamt knapp 31 Prozent aller Wahlprognosen außerhalb der Fehlerintervalle liegen (ebd.: 118).

Schließlich wird die Performanz der einzelnen Institute miteinander verglichen und es zeigt sich, dass das IfD Allensbach mit einer durchschnittlichen absoluten Abweichung von 1,68 Prozentpunkten über alle vorliegenden Prognosen des Instituts am besten abschneidet, dicht gefolgt von Infas (1,70 Prozentpunkte), der Forschungsgruppe Wahlen (1,72 Prozentpunkte), Emnid (1,94 Prozentpunkte), Infratest (1,95 Prozentpunkte), Forsa (2,07 Prozentpunkte) und Infratest dimap (2,20 Prozentpunkte) (vgl. ebd.: 159).

Den jüngsten Beitrag zur Voraussagegüte von Wahlabsichtsumfragen liefert Hoffmann-Göttig (2005). Er untersucht sie anhand von 19 Wahlen im Zeitraum 2001 bis 2005 auf Grundlage von im Internet gesammelten Ergebnissen von Vorwahlumfragen verschiedener Institute, wobei nur Umfragen berücksichtigt wurden, die jeweils nicht länger als zehn Tage vor einer Wahl veröffentlicht wurden. Grundlage der Evaluation bilden insgesamt 47 Umfrageergebnisse, welche er mit den jeweiligen Wahlresultaten vergleicht. Zur Bestimmung der Güte werden die mittleren Abweichungen aufsummiert und mit der erwarteten statistischen Fehlertoleranz verglichen. Ergebnis seiner Analyse ist, dass der Mittelwert des summierten tolerablen Fehlers größer ist als der Mittelwert des tatsächlichen summierten Fehlers der 47 untersuchten Umfragen zu den 19 Wahlen und somit für den betrachteten Zeitraum den Demoskopen kein Versagen nachgesagt werden könne.

Im Rahmen einer Einzelbetrachtung der berücksichtigten Wahlen schenkt Hoffmann-Götig (2005) insbesondere den als Fehlprognose zu wertenden, das heißt außerhalb des Fehlerintervalls liegenden, Voraussagen seine Aufmerksamkeit. Hierbei kommt er zu dem Schluss, dass die demoskopischen Probleme offenbar insbesondere in den neuen Bundesländern liegen, da hier die Fehlerquote deutlich über der der alten Länder liege. Weiterhin legt auch dieser Autor ein Benchmarking der erfassten Institute vor, wobei die zugrunde liegende Fallzahl zu wenig belastbar erscheint, um hieraus Schlüsse zu ziehen (zwischen einer und 15 Umfrageergebnisse pro Institut).

Einen anderen Ansatz zur Bestimmung der Prognosemöglichkeiten auf Umfragebasis verfolgt Ulmer (1987a, 1987b). Ausgehend von einer unsystematischen Dokumentation von Fehlprognosen auf Grundlage von Sonntagsfragen (Ulmer o.J.), zeigt er statistisch, dass allein aufgrund der geringen Stichprobengrößen (in der Regel 1.000 bis 2.000 Befragte) eine hinreichend exakte Prognose der Zweitstimmenanteile der Parteien nicht zu leisten ist. Ausgehend von einem Gedankenexperiment, man hätte 40 Millionen reale Stimmzettel und einen wahren FDP-Zweitstimmenanteil von 4,5 Prozent, zeigt Ulmer, dass – zieht man immer wieder zufällig 1.000 Stimmzettel aus der Grundgesamtheit – 24 Prozent dieser Stichproben vermuten lassen würden, dass die FDP die Fünf-Prozent-Hürde meistert. Um bei realen 4,5 Prozent FDP-Stimmenanteil auf Grundlage einer (perfekten) Stichprobe weniger als ein Prozent falsche „Prognosen“ zu erhalten, wäre ein Stichprobenumfang von 10.000 Befragten notwendig. Fazit von Ulmers Analysen ist, dass jedwede Umfrage, zumal sie nicht perfekt sein kann, nicht geeignet ist, Prognosen für den Zweitstimmenanteil in einer medial verwertbaren Genauigkeit liefern zu können.

Ulmer (1987a, 1987b) verweist hier in der Tat auf treffende Probleme umfragebasierter Voraussagen, die durchaus auch von Wahlforschern geteilt werden (bspw. Rattinger/Ohr 1989). Die publizierten Sonntagsfragen treffen diese Argumente jedoch nicht, da – trotz der Reklamation durch die Institute – durch die Umrechnungsverfahren inferenzstatistische Verfahren ohnehin nicht mehr zulässig sind. Zudem haben bereits die ersten Analysen von Radtke (1977) gezeigt, dass die Performanz ungewichteter Wahlabsichtsumfragen hinter den gewichteten Daten zurückbleibt. Anhand der Daten des Politbarometers wird in Abschnitt 3.4 der Einfluss der politischen Gewichtungungsverfahren dargestellt sowie eine systematische Analyse der Prognosegüte der publizierten Sonntagsfragen

angestrengt, die sowohl hinsichtlich der herangezogenen Datengrundlage als auch in Bezug auf die angewendeten Verfahren über die bisher vorliegenden Arbeiten hinausreicht.

Für *Wahlerwartungsfragen* liegen keinerlei empirische Befunde in Bezug auf ihre Prognosegüte vor, obwohl zumindest das IfD Allensbach in der Vergangenheit die Einschätzung des Wahlgewinners beziehungsweise der stärksten Partei in den politischen Umfragen erfasst hat, wie einzelnen Jahrbüchern des Instituts zu entnehmen ist (Noelle-Neumann 1976: 134; Noelle-Neumann/Köcher 2002b: 791; Noelle/Neumann 1956: 176; 1974: 32).

Obwohl in den USA Wahlerwartungsfragen bereits häufiger eingesetzt wurden, liegen auch hier nur spärliche Befunde zu ihrer Genauigkeit vor. Lewis-Beck und Tien (1999) zeigen anhand von Daten der *National Election Studies* für die Jahre 1956 bis 1996,¹⁶ dass die Befragten die Gewinner von Präsidentschaftswahlen relativ genau vorhersagen können. So liegen die Voraussagen auf der Grundlage der Frage nach der Einschätzung, welcher Kandidat bei der kommenden Präsidentschaftswahl gewählt werden wird, in 71 Prozent der betrachteten Wahlen richtig – damit, so die Autoren, sind diese Voraussagen mindestens ebenso präzise wie diejenigen des Meinungsforschungsinstituts Gallup auf Grundlage von Wahlabsichtsfragen. Als zentraler Vorteil der Wahlerwartungsfragen gegenüber den vergleichend betrachteten Wahlabsichtsfragen wird gesehen, dass die Wahlerwartungsfragen bereits zu einem früheren Zeitpunkt vor dem Wahltermin die Genauigkeit der späteren Wahlabsichtsumfragen aufweisen: Während die Voraussage Gallups auf Umfragen nur wenige Tage vor der jeweiligen Wahl beruhen, liegen die Wahlerwartungseinschätzungen im Schnitt einen Monat vor dem Wahltermin.

Die Präzision der *Wahltagsbefragungen* wurde bisher ebenfalls kaum systematisch untersucht beziehungsweise liegen hierüber keine veröffentlichten Befunde vor. Es ist davon auszugehen, dass die Institute selbst sowie die Nachfrager, insbesondere ARD und ZDF, die Genauigkeit der „18 Uhr-Prognosen“ aufmerksam beobachten, da diese wesentlich für das weitere Engagement der

16 In einer früheren Studie berichten Lewis-Beck und Skalaban (1989) vergleichbare Ergebnisse. Da diese Befunde ebenfalls auf Auswertungen der National Election Surveys beruhen, wird auf diese Studie nicht gesondert eingegangen.

Meinungsforschungsinstitute sind und als zentrales Signal für die Qualität deren Arbeit aufgefasst werden.

Die von Hilmer und Hofrichter (2001: 460ff.) vorliegende Bewertung für den Zeitraum von 1993 bis 2000 belegt deren Güte. Sie zeigen, dass die falsche Voraussage von Wahlsiegern äußerst selten war und auch die Abweichungen der berichteten Voraussagen vom Endergebnis eher gering ausfallen. So lag die durchschnittliche kumulierte Abweichung für 18 Wahlen zwischen 1993 und 1996 bei 5,3 (Infratest dimap) beziehungsweise 4,4 (Forschungsgruppe Wahlen) Prozentpunkten und im Zeitraum 1997 bis 2000 für 16 Wahlen bei 3,9 beziehungsweise 3,8 Prozentpunkten. Hilmer und Hofrichter (2001: 463) gehen in der Folge von einer mittleren Abweichung von weniger als einem Prozentpunkt pro Partei aus, weshalb sie die Bedeutung der *exit polls* insbesondere in der Gewährleistung präziser Prognosen sehen. Gleichwohl erscheint eine Evaluation von Wahltagsbefragungen wenig erkenntnisfördernd, zumal sie aufgrund der geringen *lead* wenig Informationsgehalt für eine Voraussage aufweisen.

2.2 Wahlbörsen

Die Voraussage von Wahlergebnissen auf Grundlage von elektronischen, das heißt internetbasierten experimentellen Märkten ist ein vergleichsweise junges Unterfangen, allerdings mit starken Wachstumstendenzen. Die Teilnehmer an diesen Märkten handeln mit Aktien (genauer: *Futures*), wobei die Höhe der Auszahlung für diese vom Ausgang zukünftiger Ereignisse, etwa der Wahl abhängt (Wolfers/Zitzewitz 2004). Etabliert wurden diese Märkte ursprünglich weniger zu Prognosezwecken als vielmehr dazu „*Informationsbeschaffungsprozesse* sowie deren Bedeutung für Entscheidungsbildung und Markteffizienz“ (Brüggelambert 1999: 24, Hervorhebung im Original) zu analysieren. Deren Entwicklung ist im Rahmen der seit den 1980er Jahren zunehmend Bedeutung erlangenden experimentellen Wirtschaftsforschung zu sehen. Experimentelle Märkte nehmen hierbei neben Laborexperimenten eine zentrale Rolle ein (vgl. z. B. Huber 2002: 4ff.).¹⁷

¹⁷ Einführungen in die experimentelle Wirtschaftsforschung liefern unter anderem Davis und Holt (1993) sowie die Aufsätze in der Edition von Kagel und Roth (1995). Knappe Zusammenfas-

Die Anwendung der Märkte zeigt, dass sie erstaunlich präzise den Ausgang verschiedener Ereignisse vorhersagen können – neben Wahlresultaten auch beispielsweise zukünftige Oscargewinner, den Superbowlsieger oder den Eintritt terroristischer Anschläge (vgl. bspw. Sunstein 2006: 103ff.). Diese prognostischen Erfolge führten zu einer schnellen Etablierung dieser Märkte auch außerhalb der Ökonomik und in verschiedenen Ländern. Besonders rasant war dabei die Etablierung von Wahlbörsen, welche im Folgenden erörtert werden. Dazu wird zunächst deren Funktionsweise dargestellt (Abschnitt 2.2.1), dann ihr prognostischer Wert unter Rückgriff auf die theoretischen Grundlagen dieser Märkte vor wissenschaftstheoretischem Hintergrund bewertet (Abschnitt 2.2.2) sowie empirische Befunde zu ihrer Prognosegüte auch vor dem theoretischen Hintergrund diskutiert (Abschnitt 2.2.3).

2.2.1 Funktionsweise von Wahlbörsen

Die erste Wahlbörse wurde am *Henry B. Tippie College of Business* der *University of Iowa* zur US-Präsidentschaftswahl 1988 zwischen George W. Bush und Michael Dukakis von Forsythe und seinen Kollegen (1992) entwickelt. Dieser erste experimentelle Markt der in der Folge ausgebauten *Iowa Electronic Markets* (IEM) sollte im Rahmen universitärer Forschung und Lehre zur Untersuchung der Preisbildung und Informationsaggregation an Märkten dienen. Grundlegende Idee hierbei ist, dass Marktpreise den besten Prädiktor zukünftiger Ereignisse darstellen und der Marktmechanismus für eine optimale Informationsaggregation auch bei unvollständigen und asymmetrisch verteilten Informationen sorgt (Hayek 1945).

Ergebnis der ersten Wahlbörse von Forsythe et al. (1992) war, dass der Markt den Wahlausgang mit einer Abweichung von nur etwa 0,25 Prozent vorhersagte. Die Autoren schlossen daraus, dass alle verfügbaren Informationen in den Preisen abgebildet wurden und demnach Märkte unabhängig von wirkenden

sungen zentraler Forschungsergebnisse liefern Diekmann (2008) sowie Fehr und Gintis (2007). Eine umfassende Darstellung der Befunde der experimentellen Wirtschaftsforschung ist bei Plott und Smith (2008) zu finden. Eine Übersicht speziell über die deutschen Beiträge zur experimentellen Wirtschaftsforschung sind im Sammelband von Bolle und Lehmann-Waffenschmidt (2002) enthalten.

individuellen Entscheidungsverzerrungen sehr gut als Informationsaggregatoren funktionieren. Dieser Erfolg führte – auch im Rahmen des parallel verlaufenden Erfolgs der gesamten experimentellen Wirtschaftsforschung – zu einer schnellen Verbreitung solcher experimenteller Märkte (vgl. die Übersichten bei Tziralis/Tatsiopoulos 2008; Wolfers/Zitzewitz 2006).

Neben politischen Ereignissen rückten immer mehr andere nicht-marktliche Phänomene in den Fokus. So werden mittlerweile vergleichbare Börsen veranstaltet, um die Ausgänge von Sportereignissen (vgl. bspw. Bochow et al. 2002) oder die Oscargewinner (z. B. Pennock et al. 2001) zu prognostizieren (für eine Auswahl solcher Märkte siehe Abramowicz 2007: 9ff.; Huber 2002: 89ff.; Ray 2006: 2f.; Sunstein 2006: 103ff.; Wolfers/Zitzewitz 2004).

Auch in regionaler Hinsicht verbreitete sich das Instrument schnell. Bereits zur ersten gesamtdeutschen Bundestagswahl 1990 wurde eine Wahlbörse etabliert (Brüggelambert 1999: 42) und auch in zahlreichen anderen Staaten werden inzwischen regelmäßig Wahlkämpfe von *political stock markets* begleitet (Huber 2002: 57ff.). Interessant ist diesbezüglich, dass in Europa Wahlbörsen eine deutlich größere Aufmerksamkeit erfahren als in den USA, was dazu führte, dass etwa die Mehrzahl der deutschen experimentellen Märkte von kommerziellen Anbietern beziehungsweise zumeist überregionalen Zeitungen angeboten werden (so etwa Wahlstreet der Holtzbrinck-Gruppe mit unter anderem der Wochenzeitung *Die Zeit* und dem *Handelsblatt* oder die Strategenbörse der *Financial Times Deutschland*).¹⁸ Hinsichtlich der Teilnehmergruppe sind mittlerweile die meisten Märkte zugänglich für alle Interessierten, während einige frühe Märkte, so auch die an der *University of Iowa*, nur von Studierenden genutzt werden konnten.

Schließlich hat sich auch das Design der Märkte ausdifferenziert. Die ersten Märkte der IEM waren als *vote-share*-Märkte angelegt, während gerade in den USA mittlerweile *winner-take-all*-Märkte sich größerer Beliebtheit erfreuen (vgl. dazu bspw. Erikson/Wlezien 2008). Weiterhin gibt es *seat-share*-Märkte, welche auf die Vorhersage der Sitzanzahl von Parteien in Parlamenten abzielen und vor allem in Großbritannien eingesetzt werden (Huber 2002: 54). Für deutsche Land-

¹⁸ Die Wahlstreet zur Bundestagswahl 2009 wurde nicht mehr von der Holtzbrinck-Gruppe, sondern im Rahmen einer Lehrforschungsprojekts am Institut für Soziologie der Ludwig-Maximilians-Universität München unter der Leitung von Christian Ganser und dem Verfasser durchgeführt.

tags- und Bundestagswahlen haben sich hingegen *vote-share*-Märkte als Prognoseinstrument etabliert. Gefolgert werden kann hieraus, dass das Wahlsystem wesentlich ist für die Herausbildung des dominanten Marktdesigns. Bei der personalisierten Präsidentschaftswahl in den USA liegt das Hauptinteresse am Gewinner der Wahl. Dies wird am besten durch *winner-take-all*-Märkte abgebildet. In Großbritannien sind die Stimmenanteile der einzelnen Parteien aufgrund des Mehrheitswahlrechts nicht interessant, jedoch die Zahl der durch die Partei erreichten Sitze im britischen Unterhaus. In Deutschland hingegen richtet sich vor dem Hintergrund des personalisierten Verhältniswahlrechts mit Fünf-Prozent-Hürde der Blick auf die Stimmenanteile.

Der Unterschied zwischen *vote-share*- und *winner-take-all*-Märkten besteht in der Handhabung des Auszahlungsmechanismus. Bei ersteren orientiert sich die Höhe der Auszahlung am tatsächlichen Stimmenanteil einer Partei oder eines Kandidaten. So hat beispielsweise die SPD bei der Bundestagswahl 2005 34,2 Prozent der Stimmen erhalten. Dies wird übersetzt in eine Auszahlung von 34,2 Cent pro SPD-*Future*. Bei *winner-take-all*-Märkten hingegen erhält man nur für die *Futures* des Siegers eine Auszahlung. So bekamen etwa bei den US-Präsidentschaftswahlen 2008 die Besitzer von Obama-*Futures* für jedes einen Dollar ausbezahlt, während für alle anderen Kandidatenaktien im Portfolio nichts ausbezahlt wurde. Ein zentraler Unterschied zu realen Aktienmärkten wird hier deutlich: Die Auszahlung richtet sich nach dem tatsächlichen Ausgang des zu handelnden Ereignisses, während eine derartige externe Offenlegung des „wahren“ Kurses an Aktienbörsen nicht erfolgt (vgl. Huber 2002: 16ff.).

Im Weiteren werden nur *vote-share*-Märkte betrachtet, da die Stimmenanteile einzelner Parteien im politischen System Deutschlands die wesentlich interessierende Größe darstellt, auf die sich auch die anderen Prognoseansätze beziehen. Allerdings unterscheidet sich der Aufbau der anderen Märkte mit Ausnahme der thematisierten Auszahlungen nicht wesentlich von dem hier beschriebenen.

Kennzeichen experimenteller Märkte ist zunächst die freiwillige Teilnahme an ihnen. Eine Ausnahme hierzu bilden Märkte im Rahmen universitärer Lehrveranstaltungen, bei denen die Kursteilnehmer zum Leistungserwerb an diesen Märkten zur Teilnahme angehalten waren. Die Freiwilligkeit bedingt, dass (bisher) ausschließlich eine spezielle Personengruppe an diesen Märkten handelt: Männer, politisch Interessierte, höher Gebildete, Jüngere sowie Studierende sind

deutlich überrepräsentiert (vgl. etwa Berg et al. 2008; Brüggelambert/Crüger 2002: 170; Huber 2002: 63ff.).

Weiterhin wesentlich ist, dass die Teilnehmer reales Geld investieren, auch wenn die finanziellen Einsätze gering sind (bei den IEM zum Beispiel auf 500 US-Dollar begrenzt) und zumeist alle Teilnehmer nur dieselben Einsätze leisten können. Letzterer Mechanismus wird eingesetzt, um die sonst mögliche Marktmacht einzelner Händler zu unterbinden. Durch den Einsatz realer Geldbeträge verbunden mit der Möglichkeit realer Gewinne und Verluste verspricht man sich überlegte Entscheidungen der Teilnehmer und folglich bessere Approximationen an das spätere Wahlresultat als ihm Rahmen konsequenzenloser Befragungen (siehe hierzu bspw. Berlemann 2000: 12). Allerdings ist die Etablierung von Wahlbörsen mit Realgeldeinsatz insbesondere in Deutschland rechtlich schwierig (Beckmann/Werding 1996: 176; Huber 2002: 52), weshalb sich auch zunehmend experimentelle Börsen etabliert haben, an denen mit fiktivem Geld und ohne Einsatz eigener monetärer Mittel gehandelt wird. Beispiele hierfür sind die Angebote von Wahlieber oder auch die (mittlerweile eingestellte) Strategenbörse der *Financial Times Deutschland*.

Nach der Anmeldung und gegebenenfalls Entrichtung des vorgegebenen Einsatzes an einem experimentellen Markt können die Teilnehmer die *Futures* der jeweiligen Parteien oder Kandidaten frei handeln. Die Märkte sind dabei als Nullsummenspiel konstruiert – ein weiterer Unterschied zu realen Aktienbörsen. Die Bank beziehungsweise der Veranstalter gibt Basisportfolios zu einem Festpreis aus, welche jeweils ein *Future* aller zu handelnden Parteien oder Kandidaten umfassen. Diese Basisportfolios kann man während der gesamten Börsenlaufzeit beim Veranstalter zum Ausgabepreis (zumeist ein Euro) kaufen und verkaufen. Damit erhalten Wahlbörsen den Charakter eines „ständigen Primärmarktes“ (Huber 2002: 23). Diese Konstruktion stellt sicher, dass von jedem Kandidaten oder jeder Partei immer dieselbe Anzahl an *Futures* im Umlauf ist. Auch der Nullsummencharakter ist hierdurch festgelegt, denn bei der Auszahlung addieren sich die Werte jeweils einer Aktie über alle Kandidaten beziehungsweise Parteien zu einem Euro auf. Weiterhin ist damit gegeben, dass der Erwartungswert der Gewinne gleich Null ist – sofern der Veranstalter nicht weitere Gewinne über die Auszahlungen hinaus auslobt. Letzters ist üblich: So werden häufig gerade bei kommerziellen Anbietern Wahlbörsen mit Gewinnspielen verknüpft und an die erfolgreichsten Teilnehmer ausgeschüttet. Auch bei Wahlbörsen ohne Realgeld-

einsatz soll die Aussicht auf einen möglichen Gewinn die Situation an experimentelle Realgeldbörsen annähern (so etwa bei Wahlieber).

Weiterhin bedingt die Konstruktion von Wahlbörsen als ständige Primärmärkte, dass – im Gegensatz zu Aktienbörsen – Arbitragegeschäfte möglich sind. Da diesen im Allgemeinen zugesprochen wird, die Markteffizienz zu erhöhen, wird diese Eigenschaft als besonders nützlich für die Informationsaggregation angesehen (Hayek 1945: 522).

Üblicherweise wird für Wahlbörsen das Vorliegen von Transaktionskosten negiert. Brüggelambert und Crüger (2002: 171) weisen jedoch darauf hin, dass bei den niedrigen zu erwartenden Gewinnen hiervon nicht zwingend ausgegangen werden kann. Für die Optimalität der Informationsaggregation erscheint die Abwesenheit von Transaktionskosten allerdings auch nicht zwingend. Der Veranstalter verdient durch den Handel kein Geld mit den experimentellen Märkten. Die für ihn anfallenden Kosten werden entweder im Rahmen von Forschungsvorhaben durch entsprechend vorhandene Mittel gedeckt, oder, im Fall kommerzieller Anbieter, über Werbung auf den Webseiten der Börse finanziert. Häufig werden entsprechende Angebote in den Internetauftritt von Zeitungen integriert. So soll die Nutzerzahl und der Bekanntheitsgrad der eigenen Seiten und auch der Zeitung gesteigert werden. Zudem verweist Huber (2002: 50) auf die Möglichkeit von Reputationsgewinnen durch Wahlbörsen, die das Wahlergebnis gut approximieren.

Der Handel an Wahlbörsen ist zumeist in Anlehnung an die IEM als *continuous double auction market* konstruiert (ebd.: 23).¹⁹ Damit ist gemeint, dass 24 Stunden am Tag, sieben Tage pro Woche limitierte und unlimitierte Verkaufs- und Kaufordern platziert werden können. Allerdings werden keine Ordern bedient, welche das verfügbare Budget oder die Anzahl der vorhandenen Aktien im Depot überschreiten. Die Preisbildung am Markt erfolgt durch Gegenüberstellung der Kauf- und Verkaufsangebote, wobei der teuerste Käufer beziehungsweise der günstigste Verkäufer den Zuschlag erhält – bei Gleichheit der Angebote wird zuerst das älteste Angebot beziehungsweise Gesuch bearbeitet. Der Kurs spiegelt demnach die Höhe der zuletzt getätigten Transaktion für die entsprechende Aktie wider.

19 Eine Ausnahme stellt beispielsweise die Passauer Wahlbörse dar, welche auf dem Kassamarkt-Mechanismus aufbaut (vgl. Beckmann/Werding 1996).

Interessant erscheint noch die Frage, ob sich Strategien identifizieren lassen, mit denen sich an Wahlbörsen Gewinne realisieren lassen. Huber (2002: 45ff.) unterscheidet zwischen der *Trading*- und der *Payoff*-Strategie.²⁰ Erstere kennzeichnet ein Marktverhalten, bei welchem die Händler möglichst günstig Aktien einkaufen und möglichst teuer wieder verkaufen (*buy-low, sell-high*). Diese Strategie erfordert eine profunde Marktkenntnis und zudem ein reges Handelsvolumen, da hier zumeist Ordern nahe um den Kurswert platziert werden und somit versucht wird, viele kleine Gewinne zu erzielen. Forsythe et al. (1992: 1153ff.) bezeichnen diese Gruppe von Händlern als *marginal trader*. Die zweite Strategie zielt darauf ab, Aktien von – der eigenen Einschätzung nach – unterbewerteten Parteien oder Kandidaten zu kaufen und zu halten, um an der Auszahlung zu verdienen. Grundlage dieser Strategie ist in der Regel die Unterstellung der eigenen Überlegenheit bei der Prognosefähigkeit gegenüber dem Markt. Diese Strategie tritt insbesondere bei weniger aktiven Händlern auf. Bereits Forsythe et al. (ebd.) zeigen, dass diese Gruppe einem erheblichen *judgement bias* unterliegt. In einer Zusammenfassung des Forschungsstandes zeigt Huber (2002: 46f.), dass in Wahlbörsen eine systematische Überschätzung der Stimmenanteile kleinerer Parteien zu beobachten ist, die er auf die Händler zurückführt, welche die *Payoff*-Strategie verfolgen.

2.2.2 Zum prognostischen Wert von Wahlbörsen

Analog zu Vorhersagen auf Basis von Wahlabsichtsfragen liefern experimentelle Märkte Prognosen vor dem Eintritt des Explanandums, womit die erste Anforderung

20 Berlemann (2000: 9f.) differenziert zwischen vier Strategien: der Arbitrage-, der einfachen Erwartungs-, der modifizierten Erwartungs- sowie der Spekulationsstrategie. Diese Unterteilung unterscheidet sich nur in einer stärkeren Ausdifferenzierung der bei Huber (2002) genannten Strategien. So zeichnen sich sowohl Arbitrageure als auch Spekulanten durch ein hohes Handelsvolumen und eine profunde Marktkenntnis aus. Durch die Markt konstruktion sind jedoch sichere Gewinne durch Arbitragegeschäfte möglich, während der Spekulant eher auf mittelfristige Trends setzt, um Gewinne zu realisieren. Berlemann zufolge ist die Unterscheidung dieser Typen wichtig, da Arbitrageure wesentlich zur Markteffizienz beitragen, während Spekulanten im Markt irrationale Bewertungen von Aktien begünstigen. Die einfache Erwartungsstrategie entspricht der *Payoff*-Strategie Hubers, die modifizierte Erwartungsstrategie erweitert diese Strategie um die Einbeziehung möglicherweise falscher Erwartungswerte durch den Händler, welche während der Laufzeit der Börse angepasst werden.

rung erfüllt ist. In der Regel werden die Wahlbörsen Monate vor einer Wahl etabliert und an den Kursen der *Futures* in *vote-share*-Märkten können zu jedem Zeitpunkt direkt die vorhergesagten Stimmenanteile abgelesen werden. Die Märkte werden normalerweise am Tag der Wahl geschlossen und die Auszahlung der Gewinne erfolgt nach Bekanntgabe des amtlichen Endergebnisses.

Im Gegensatz zu Vorhersagen auf Basis von Wahlabsichtsfragen, die in Abschnitt 2.1 behandelt wurden, stützt sich die Vorhersage von Wahlresultaten auf Basis experimenteller Märkte nicht auf eine einfache Aggregationsregel. Experimentelle Märkte basieren auf einem theoretischen Fundament, das den Marktmechanismus als effizienteste Institution zur Aggregation von Informationen auffasst.

Hayeks (1945) Arbeit zur Informationsaggregation durch Märkte zufolge ist zu erwarten, dass der Preis zu einem bestimmten Zeitpunkt immer den besten Indikator für die in einem Markt vorhandenen Informationen darstellt. Der Ausgangspunkt von Hayeks (ebd.) Überlegungen ist, dass in einem Markt, auf welchem rationale Akteure aufeinandertreffen, anders als beispielsweise in Befragungssituationen kein Anreiz zur Informationsverheimlichung gegeben ist, da Personen mit mehr Informationen aufgrund der für sie dadurch zu erwartenden höheren Gewinne diese offenlegen werden. Diese Grundthese wurde im Zeitverlauf weiter entwickelt und auch vielfach empirisch geprüft (z. B. Kou/Sobel 2004; Sunder 1995). Beckmann und Werding (1994, 1998) unterscheiden drei Komponenten der Hypothese.

Die erste Komponente bezeichnen die Autoren als *statische Pareto-Effizienz* von Wettbewerbsmärkten. Gemeint ist damit die von Hayek postulierte Eigenschaft von kompetitiven Märkten, welche zu Allokationen führt, „durch die alle erreichbaren Tauschgewinne ausgeschöpft werden“ (Beckmann/Werding 1998: 757). Dieser Aspekt gilt als Kern der Hayek-Hypothese und die meisten vorliegenden Experimente beziehen sich hierauf (Smith 1982b).

Zweitens umfasst die Hayek-Hypothese das Postulat *rationaler Erwartungen*. Demnach funktioniert dieser Mechanismus selbst beim Vorliegen unvollständiger und asymmetrisch verteilten Informationen, da der Wettbewerb auf den Märkten dafür sorgt, dass sich die Akteure so verhalten, als ob ihnen sämtliche in der Gesellschaft vorliegenden Informationen vorliegen würden. Der Marktmechanismus sorgt demnach für eine effiziente Informationsaggregation. Zudem

ist er anderen Mechanismen überlegen, da der Gesellschaft keine Kosten für die Informationssuche und -aggregation entstehen.

Schließlich identifizieren Beckmann und Werding (1998: 758) die Idee vom *Wettbewerb als Entdeckungsprozess* als dritte Komponente der Hypothese. Hayek postuliert, dass der Wettbewerbsmechanismus den Akteuren einen Anreiz bietet, Informationen zu sammeln und Preise als Signale für das aktuelle gesellschaftliche Wissen aufgefasst werden können.

Wahlbörsen werden jedoch nicht nur in Verbindung mit Hayeks These gebracht, sondern ebenso vor dem Hintergrund der „Weisheit der Vielen“ diskutiert. Surowiecki (2005: 41ff.) selbst führt die IEM als Beleg für die empirische Stichhaltigkeit seiner These an. Auch in der Literatur finden sich in Verbindung mit experimentellen Märkten Verweise hierauf (Erikson/Wleziën 2008; Ray 2006). Hayeks Idee und die postulierte „Weisheit der Vielen“ sind sich ähnlich, zumal im Resultat. Beiden Ansätzen gemein ist, dass unvollständige Information ausreicht, um ein effizientes Marktergebnis etwa in Form einer genauen Wahlprognose zu liefern. Während für Surowiecki (2005: 46) der spezifische Aggregationsmechanismus explizit keine Rolle spielt und er keinen Grund für die Überlegenheit des Marktes gegenüber anderen Methoden zur Erfassung kollektiver Weisheit sieht, ist bei Hayek eine effiziente Informationsaggregation nur in Wettbewerbssituationen zu erwarten. Bei Surowiecki (2005: 47) hingegen liegt „der zentrale Schlüssel [des] Erfolgs (...) nicht so sehr in der Perfektionierung irgendeiner bestimmten Methode, sondern in der Erfüllung der notwendigen Bedingungen – Diversität, Unabhängigkeit, Dezentralisierung (...)“.

Vor allem die Unabhängigkeit der Entscheidungen ist jedoch bei Wahlbörsen fraglich, etwa da anzunehmen ist, dass veröffentlichte Wahlabsichtsumfragen die Entscheidung der Händler beeinflussen. Empirische Studien belegen insbesondere in Deutschland die Relevanz der Sonntagsfragen für das Händlerverhalten (siehe Abschnitt 2.2.3). Auch das Kriterium der Diversität bezüglich der Gruppenzusammensetzung erfüllt bisher kaum eine Wahlbörse, da fast alle bisher durchgeführten elektronischen Märkte nur einen sehr kleinen Teilnehmerkreis umfassen und darüber hinaus keineswegs heterogen zusammengesetzt sind (vgl. Abschnitt 2.2.1). Posner (2008) argumentiert darüber hinaus, dass die Idee der „Weisheit der Vielen“ auf der simplen Tatsache der Reduktion des „sampling errors“ fußt und demnach nicht mit den theoretischen Überlegungen Hayeks in Verbindung gesetzt werden kann.

Vor diesem Hintergrund muss in Bezug auf die Bewertung von Wahlbörsen als Instrument zur wissenschaftlichen Prognose festgehalten werden, dass mit Hayeks These ein profundes Theoriefundament vorliegt. Dieses ist darüber hinaus zumindest bezogen auf die erste Komponente außerhalb von Wahlbörsen empirisch gut bestätigt (vgl. dazu bspw. Beckmann/Werding 1994: 13ff.; Huber 2002: 170ff.; Smith 1982b). Hayeks These beruht – anders als Surowieckis Überlegungen – auf einem präzisen, formalisierbaren Modell und erlaubt auch die empirische Überprüfung der Stichhaltigkeit dieser Postulate (Kou/Sobel 2004; Wolfers/Zitzewitz 2006). Anzumerken ist, dass Hayeks Hypothese beziehungsweise deren weitere Ausarbeitungen weitaus präziser sind als die von Surowiecki genannten Bedingungen und diese insofern vorzuziehen sind. In Bezug auf die Anforderung an eine wissenschaftliche Prognose auf einen Mechanismus zu verweisen, muss konstatiert werden, dass dies im Fall von Wahlbörsen mit dem Verweis auf den Marktmechanismus erfüllt ist.

Hinsichtlich der korrekten Erfassung der Anfangsbedingungen, der dritten in Abschnitt 1.2 abgeleiteten Anforderung, ist die Ausgestaltung des Marktes zu berücksichtigen. Das Design der Wahlbörse muss sicherstellen, dass die von Hayek benannten Komponenten realisiert werden können. Dies erscheint grundsätzlich möglich und diese Aspekte werden im folgenden Abschnitt zur empirisch festgestellten Prognosegüte aufgegriffen.

Zusammenfassend kann festgehalten werden, dass experimentelle Märkte alle drei methodologischen Anforderungen an wissenschaftliche Prognosen erfüllen: Die Prognose kann vor dem Eintritt des Explanandums geäußert werden, der Ansatz verweist auf eine empirisch gut bestätigte und zudem vergleichsweise präzise Gesetzmäßigkeit und durch ein entsprechendes Design der Wahlbörse können auch die Randbedingungen kontrolliert werden.

2.2.3 Vorliegende Evidenz zur Güte von Wahlbörsen

Ausgangspunkt der mittlerweile erheblichen Forschungen auf Grundlage experimenteller Märkte bildet die genaue Vorhersage der Ergebnisse der US-Präsidentenwahl 1988, welche – gemessen am Schlusskurs – den Stimmenanteil von Bush exakt traf und den von Dukakis nur um 0,1 Prozentpunkte verfehlte (Forsythe et al. 1992). Dieser Erfolg konnte bei der US-Präsidentenwahl 1992

wiederholt werden (Forsythe et al. 1994) und beflügelte die Verbreitung experimenteller Märkte. Damit ging eine rapide wachsende Zahl an Forschungsbeiträgen zu diesem Thema einher, aus denen sich die im Folgenden knapp erörterten zentralen Befunde ebenso herausarbeiten lassen wie bestehende Kontroversen und offene Fragen.

Gängige Auffassung in der Literatur ist, dass Wahlbörsen ein valides Prognoseinstrument darstellen und Wahlresultate besser approximieren als Wahlabsichtsfragen (vgl. etwa Berg et al. 2008; Berlemann 1999, 2000; Berlemann/Schmidt 2001; Wolfers/Zitzewitz 2006). Diese Überlegenheit wird mit methodischen Argumenten, aber auch mit empirischen Befunden begründet.

In methodischer Hinsicht wird insbesondere der Vorzug von Wahlbörsen betont, dass keine Zufallsstichprobe für eine akkurate Prognose notwendig sei. Damit ist man auch frei von all den Stichproben- und Befragungsproblemen (Non-Response, Fragereihenfolgeeffekte, Interviewereffekte, soziale Erwünschtheit, hypothetischer Gehalt der Sonntagsfrage etc.), mit denen die Meinungsforscher konfrontiert sind und die letztlich nicht in den Griff zu bekommen seien. Weiterhin sei zu jedem beliebigen Zeitpunkt vor der Wahl – nach Börsenstart – eine Vorhersage möglich und damit könnte auch der Einfluss aktueller Ereignisse schneller verarbeitet und praktisch sofort in neue Prognosen übersetzt werden. Schließlich spreche allein der Kostenaspekt für die Stärke voll elektronischer experimenteller Märkte (vgl. bspw. Berlemann 2000).

Eine systematische Aufarbeitung der empirischen Befunde zu Wahlbörsen zeigt, dass einige der – auf Grundlage der ersten Erfolge der IEM – überschwänglichen Hoffnungen sich nicht oder zumindest nur teilweise einstellen. Für die USA kann man zwar – mit Ausnahme der US-Präsidentenwahl 1996 – eine durchgängig bessere Performanz der Wahlbörsen gegenüber Wahlabsichtsfragen konstatieren (Berg et al. 2003, 2008), jedoch kann dieser Befund insbesondere auf europäische Länder nicht übertragen werden (u. a. Brüggelambert 1997, 1999: 25; Brüggelambert/Crüger 2002; Huber 2002: 67ff.). Differenzen ergeben sich über verschiedene Wahlbörsen hinweg nicht nur in Bezug auf ihre Prognosegüte, sondern auch im Hinblick auf potenzielle Einflussfaktoren darauf.

So sinkt die Genauigkeit mit Zunahme der Anzahl gehandelter Partei- beziehungsweise Kandidatenaktien und es zeigt sich weiterhin konsistent, dass Wahlbörsen in Zweiparteiensystemen bessere Resultate erzielen als in Mehrparteiensystemen (Berg et al. 1996). Insbesondere für Deutschland ist gezeigt worden,

dass kleine Parteien systematisch überschätzt werden und dies maßgeblich zu den zu beobachtenden Abweichungen zwischen den Wahlbörsenprognosen und den tatsächlichen Stimmenanteilen beiträgt (Schaffer/Schneider 2005). Besonders gut schneiden experimentelle Märkte bei der Vorhersage von US-Präsidentenwahlen ab, welche sich noch einmal deutlich akkurater vorhersagen lassen als Kongress-, Senats- oder auch Wahlen auf Ebene der Bundesstaaten (Berg et al. 2008). Weiterhin gibt es Evidenz dafür, dass *vote-share*-Märkte schlechter abschneiden als *winner-take-all*-Märkte. Allerdings muss in diesem Zusammenhang darauf hingewiesen werden, dass entsprechende Vergleiche schwierig sind, da die Konstruktion eines Vergleichsmaßes problematisch ist. Zudem erfreuen sich *winner-take-all*-Märkte vor allem in den USA weitaus größerer Beliebtheit und so wiesen die Stimmenanteilmärkte für Vergleichszwecke häufiger zu wenige Teilnehmer auf. Grundsätzlich scheint sich auch ein höheres Marktvolumen – insbesondere kurz vor einer Wahl – begünstigend auf die Prognosegüte auszuwirken (Berg et al. 1996).

Die institutionelle Ausgestaltung des Marktes besitzt ebenfalls Relevanz. Zwar orientieren sich die meisten Wahlbörsen am Design der IEM, doch implementierten beispielsweise Beckmann und Werding (1996) einen Kassamarkt und sowohl ihre Befunde als auch die Zusammenfassung bei Brüggelambert und Crüger (2002) deuten daraufhin, dass Märkte mit *double-auction*-Mechanismus effizienter sind. Die Erwartung, dass Realgeldmärkte eine bessere Performanz aufweisen als Märkte, auf welchen mit fiktivem Geld gehandelt wird, konnte bisher nicht bestätigt werden (Servan-Schreiber et al. 2004).

Hinsichtlich des Händlerverhaltens besteht Einigkeit darüber, dass auch experimentell Verzerrungen bei den individuellen Entscheidungen zu beobachten sind. So überschätzten die Akteure beispielsweise systematisch den Stimmenanteil der individuell präferierten Partei (für eine detaillierte Analyse siehe Forsythe et al. 1999) wie auch die Eintrittswahrscheinlichkeiten für Ereignisse mit niedriger Wahrscheinlichkeit (vgl. hierzu Snowberg/Wolfers 2007). Dies wirkt sich auch auf das Marktverhalten aus – so berichten Wolfers und Zitzewitz (2004) von einem Zusammenhang zwischen der Parteiidentifikation und dem Marktverhalten. Weiterhin zeigt sich, wie bereits erwähnt, konform zu der Erwartung der systematischen Überschätzung von niedrigen Eintrittswahrscheinlichkeiten die Überschätzung von Stimmenanteilen kleiner Parteien. Dennoch sehen Wolfers und Zitzewitz (ebd.) hierin kein Problem in Bezug auf das effi-

ziente Marktergebnis, solange der Wille zur Profitmaximierung das Händlerverhalten maßgeblich bestimmt und nicht ihre Parteinähe.

Widersprüchliche Befunde liegen zum Einfluss des zugelassenen beziehungsweise handelnden Teilnehmerkreises vor. Huber (2002: 17) berichtet, dass sich im Rahmen österreichischer Wahlbörsen gezeigt hat, dass mit steigender Heterogenität der Händler die Performanz des Marktes steigt. Dies könnte man im Übrigen als Indiz für Surowieckis (2005) These der Weisheit der Vielen auffassen, wäre aber auch Hayek (1945) folgend zu erwarten, denn mit zunehmender Heterogenität der Teilnehmer dürften die im Markt vorhandenen Informationen zunehmen und damit die Markteffizienz steigern. Für die Wahlbörsen der *University of Iowa* zeigt sich hingegen, dass die Markteffizienz insbesondere auf eine bestimmte, sehr homogene Personengruppe zurückzuführen sei, die so genannten *marginal trader*.

Forsythe et al. (1992) untersuchten diese Personengruppe näher und konnten zeigen, dass diese sich durch eine hohe Marktaktivität und durch geringere Fehler beim Handel auszeichnet. *Marginal trader* kaufen und verkaufen Partei- oder Kandidatenaktien immer nahe am Marktpreis und realisieren Arbitragegeschäfte. Dadurch erwirtschaften diese Personen höhere Gewinne als durchschnittliche Händler (vgl. hierzu auch Berg et al. 2000: 6).

Während einige Studien die *marginal-trader*-These stützen (Forsythe et al. 1998, 1999; Huber 2002; Oliven/Rietz 2004), kann Brüggelambert (1997, 1999: 25) für deutsche Wahlbörsen keine markteffizienzsteigernden Effekte von *marginal trader* finden und auch der Befund, dass sie höhere Gewinne realisieren, wird nicht konsistent belegt. Brüggelambert und Crüger (2002) finden vielmehr, dass nicht weniger fehlerbehaftetes Marktverhalten zu höheren Renditen führt, sondern das Ausnutzen asymmetrischer Informationsverteilung. Demnach können politisch besser informierte höhere Gewinne erzielen. Dieser Befund kann jedoch nicht als ein Beleg dafür gewertet werden, dass Experten die besten Prognosen treffen würden. Hierfür gibt es weder für Wahlbörsen noch außerhalb empirische Evidenz (Brüggelambert/Crüger 2002; Lewis-Beck/Rice 1992: 6ff.; Rosenstone 1983: 10ff.).

Die *marginal-trader*-These ist darüber hinaus aus theoretischer Perspektive umstritten. So weist Berlemann (2002) darauf hin, dass eine fundierte theoretische Begründung dieser „market makers“ (Berg et al. 2000: 6) ausbleibt, und auch mit Hayeks These ist dieser Befund schwer in Einklang zu bringen. Auch

Surowiecki (2005: 348ff.) wendet sich gegen diese Interpretation und argumentiert, dass die *marginal trader* eine viel zu geringe Marktmacht hätten, da sie Forsythe et al. (1992) zufolge nur 15 Prozent aller Händler ausmachen und aufgrund des nach oben beschränkten Geldeinsatzes nicht genügend Marktmacht aufbringen, um Markteffizienz herzustellen. Schließlich kann man die arbiträre Definition der *marginal trader* kritisieren, welche zu einer Scheinkorrelation zwischen individueller Marktaktivität und Marginalität führt (Berlemann 2002; Brüggelambert/Crüger 2002).

Weiterhin liegen keine eindeutigen Befunde zur Möglichkeit der Manipulation von Wahlbörsen vor. Ausgangspunkt ist übereinstimmend die Annahme, dass das Profitmotiv ausreichend sei, um Manipulationsversuche schnell im Markt scheitern zu lassen (Wolfers/Zitzewitz 2004). Während sowohl einige empirische Befunde vorliegen, welche diese Annahme bestätigen (Camerer 1998; Hanson/Oprea 2007; Oprea et al. 2007; Rhode/Strumpf 2007; Wolfers/Leigh 2002), haben Aggerwal und Wu (2003) sowohl theoretisch als auch empirisch gezeigt, dass durch Manipulationen durchaus die Markteffizienz beeinflusst werden kann. Hansen et al. (2004) kommen bei ihren Analysen von zwei Wahlbörsen zur Senatswahl in Berlin 1999 ebenfalls zu dem Schluss, dass Wahlbörsen durchaus anfällig für Manipulationsversuche sind.

Wesentlich für die Performanz experimenteller Märkte ist der Informationsstand im Markt. Theoretisch zu erwarten ist, dass mit einer zunehmenden Menge an für das zu prognostizierende Ereignis relevanten Informationen, die Marktperformanz steigt, also die Informationsaggregation bei sinkender Unsicherheit immer effizienter wird. Dementsprechend intensiv wird der Einfluss von politischen Meinungsumfragen auf die Wahlbörsen diskutiert. Kritiker der Wahlbörsen wenden gerne ein, dass insbesondere Wahlabsichtsfragen die Informationen liefern, welche erst das Funktionieren der experimentellen Märkte als Prognoseinstrument ermöglichen würden. Grundsätzlich wird die Ansicht geteilt, dass politische Meinungsumfragen eine wichtige Informationsressource für die Händler darstellen (z. B. Schaffer/Schneider 2005). Inwieweit diese Ergebnisse jedoch die Händler beziehungsweise das Marktgeschehen beeinflussen, ist umstritten. So finden Forsythe et al. (1992) für die IEM und Bohm und Sonnegard (1999) für eine Wahlbörse zum schwedischen EU-Referendum 1994 keine Beeinflussung der Marktpreise durch Meinungsumfragen, während Brüggelambert und Crüger (2002) für deutsche Märkte empirische Hinweise der Beeinflussung der

Marktpreise durch Meinungsumfragen finden. Ein möglicher Erklärungsansatz für diese unterschiedlichen Befunde ist die differierende Praxis von Wahlabsichtsumfragen. Während in den USA häufiger Rohdaten, das heißt der Anteil Unentschlossener angegeben wird, werden in Deutschland ausschließlich politisch gewichtete Befunde berichtet (vgl. Abschnitt 2.1.3), welchen eine höhere prognostische Qualität unterstellt wird.

Dies könnte auch einen Beitrag zur Erklärung der unterschiedlichen Befunde zur Überlegenheit von Wahlbörsen gegenüber Wahlabsichtsfragen leisten, denn für die USA scheint diese unbestritten zu sein, während insbesondere für Deutschland hierzu weniger deutliche Befunde vorliegen (Brüggelambert/Crüger 2002). Trotz der zunehmenden Verbreitung von und dem Interesse an Wahlbörsen (Tziralis/Tatsiopoulos 2008) liegen jedoch insbesondere für Deutschland noch keine befriedigenden systematischen Vergleiche der Prognosegüte von Wahlbörsen und Wahlabsichtsumfragen vor. Die vorliegenden Vergleiche beziehen sich entweder auf Einzelbefunde zu ausgewählten Wahlen oder fassen diese eher kursorisch zusammen. Besonders unzureichend erscheinen bei diesen Vergleichen zum einen die herangezogenen Maßzahlen (siehe hierzu Abschnitt 3.3.2) und zum anderen der Vergleich der letzten veröffentlichten Umfrageergebnisse mit den Schlusskursen der Wahlbörsen.

Dieser Vergleich ist unzulässig, da der *lead* der letzten Umfragen zumeist deutlich größer ist als der der Börsen, an welchen teilweise bis kurz vor Öffnung der Wahllokale gehandelt werden kann (vgl. Schaffer/Schneider 2005). Erikson und Wlezien (2008) haben dies in ihrer Analyse berücksichtigt. Sie vergleichen für die US-Präsidentenwahlen 1988 und 2004 die IEM mit den Wahlabsichtsumfragen und berücksichtigen dabei den *lead*. Ihre Ergebnisse liefern erste Hinweise darauf, dass bei einer solchen Betrachtung Wahlabsichtsfragen keine schlechteren Voraussagen liefern als Wahlbörsen. Dieser Befund stützt auch die Vermutung, warum gerade in Deutschland die empirischen Befunde entsprechend gemischt ausfallen. Im Vorfeld deutscher Wahlen werden anders als in den USA ausschließlich politisch gewichtete Umfragen veröffentlicht. Berg et al. (2003) hingegen kommen auf Basis von Zeitreihenanalysen zu dem Ergebnis, dass auch längerfristige Prognosen auf Grundlage von Wahlbörsen eine hohe Güte aufweisen.

Insgesamt muss für Wahlbörsen festgehalten werden, dass sie einen interessanten wissenschaftlichen Prognoseansatz darstellen. Sie beruhen auf einem

empirisch gut bewährten theoretischen Fundament und erweisen sich auch in den bisher vorliegenden Tests als zumindest konkurrenzfähig mit auf Umfragen basierten Ansätzen. Dennoch bestehen Ungereimtheiten etwa hinsichtlich der Rolle von *marginal trader*, deren Einfluss auf die Markteffizienz noch genauer untersucht werden sollte. Weiterhin wären eine systematische Evaluation deutscher Wahlbörsen und eine entsprechende vergleichende Betrachtung von Wahlabsichtsumfragen und Wahlbörsen im Zeitverlauf über die Arbeiten von Brüggelambert (1999) und Berlemann und Schmidt (2001) hinaus interessant. Zumindest für die Sonntagsfragen wird in dieser Arbeit hierfür eine Datenbasis vorgestellt sowie deren Prognosegüte detailliert analysiert.

2.3 Prognosemodelle

Die letzte Gruppe von Ansätzen zur Prognose von Wahlergebnissen stellen Prognosemodelle dar. Dieser Ansätze versuchen auf Grundlage theoretischer Ansätze der Wahlforschung Wahlergebnisse vorherzusagen (Rosenstone 1983: 37). Damit grenzen sich diese Ansätze durch ihr Theoriefundament zum einen ab von umfragebasierten Voraussagen. Zum anderen unterscheiden sie sich auch von Wahlbörsen, welche auf der theoretischen Idee beruhen, dass Märkte eine effiziente Informationsaggregation erlauben; eine substanzielle Fundierung mit wahltheoretischen Überlegungen unterbleibt hier.

Für die USA liegen zahlreiche Prognosemodelle sowohl für Präsidentschafts- als auch für Kongresswahlen vor. Seit geraumer Zeit werden von führenden Wahlforschern regelrechte Wettkämpfe um das beste Modell vor entsprechenden Wahlen in Fachzeitschriften und auf Kongressen ausgefochten; so etwa regelmäßig in *American Politics Quarterly*, *PS: Political Science & Politics* oder auch im *International Journal of Forecasting*. Schließlich wurde mittlerweile die *Political Forecasting Group* als offizielle Forschungsgruppe der *American Political Science Association* gegründet (für einen Überblick hierzu siehe Campbell/Lewis-Beck 2008). Auch für andere Staaten, etwa Großbritannien oder Frankreich liegen verschiedene Vorschläge bereits vor (Bélangier et al. 2005; Dubois/Fauvelle-Aymar 2004; Jérôme/Jérôme-Speziari 2004; Sanders 1991, 2005; van der Eijk 2005; Whiteley 1979, 2005; Wolfers/Leigh 2002). In Deutschland stellt sich diese Situation anders dar. Hier wurde bisher nur ein entsprechender

Vorschlag erarbeitet und Schaffer und Schneider (2005: 674) diagnostizieren für die deutsche empirische Wahlforschung generell einen zögerlichen Umgang mit Prognosemodellen.

Im Folgenden werden knapp zentrale Charakteristika solcher theoriebasierter Prognosemodelle präsentiert (Abschnitt 2.3.1). Hierbei können und sollen lediglich die verschiedenen Konstruktionsweisen der Modelle sowie deren Entwicklung im Zeitablauf skizziert werden. Eine detaillierte Diskussion der verschiedenen Ansätze kann hier nicht erfolgen (einen Überblick über aktuelle Entwicklungen findet man bei Armstrong et al. 2008). Eingehender wird das für Deutschland vorliegende Modell von Gschwend und Norpoth (2001) behandelt (Abschnitt 2.3.2). Die Diskussion um dessen prognostischen Wert aus wissenschaftstheoretischer Perspektive (Abschnitt 2.3.3) sowie die Erörterung empirischer Befunde zur Prognosegüte der Modelle (Abschnitt 2.3.4) runden die Darstellung relevanter Prognoseansätze ab.

2.3.1 Modellvorschläge für die USA

Anders als die deutsche Wahlforschung bemüht man sich in den USA bereits seit den 1960er Jahren sozialwissenschaftliche Prognosemodelle zu etablieren. Hierbei stehen bis heute die Präsidentschaftswahlen im Mittelpunkt der Bemühungen, doch mittlerweile liegen auch einige Arbeiten zur Vorhersage von Kongresswahlen vor (etwa Kastle et al. 2008; Klarner 2008; Lewis-Beck/Rice 1992: 57ff.). Da sich diese Modelle im Wesentlichen auf dieselben Ideen und Methoden wie die im Folgenden diskutierten Ansätze für die Präsidentenwahlen stützen, wird hierauf im Folgenden nicht weiter eingegangen.

Rosenstone (1983: 37) zufolge sind die Arbeiten von Pool et al. (1964) der Ausgangspunkt von Prognosemodellen, welche auf fundamentalen Erkenntnissen der Wahlforschung beruhen. Die Autorengruppe identifizierte auf Grundlage von Surveydaten über einen Zeitraum von acht Jahren anhand von soziodemographischen und politischen Merkmalen über 100 verschiedene Wählertypen in 48 Staaten der USA, deren Einstellungen in Bezug auf verschiedene politische Streitfragen ebenfalls erfasst wurden. Mit Aggregatdaten zur soziodemographischen Zusammensetzung der einzelnen Bundesstaaten und Annahmen über das Verhältnis zwischen den Einstellungen der 480 Wählergruppen und ihrem Wahl-

verhalten berechneten die Autoren mit Hilfe von Computersimulationen die Wahlausgänge der Präsidentschaftswahlen von 1960 und 1964. Da die Autoren eher eine gute Erklärung als eine Prognose im Blick hatten, passten sie die verwendeten Gewichte in den Computersimulationen mit retrospektiven Wahldaten an, weshalb sie selbst keine Prognose, nicht einmal *out-of-sample*-Prognosen vorlegten, doch ihr Vorgehen stimulierte wesentlich nachfolgende Prognosebemühungen.

Budge und Farlie (1977) legten für die US-Präsidentenwahl 1976 Prognosen vor, welche auf Daten der *National Election Study* von 1974 beruhten. Ihre Annahme war, dass das Wahlverhalten weitgehend stabil bleibt und damit mit dem Votum bei der letzten Wahl, Kandidatenbewertungen und Einstellungen der Kandidaten zu zentralen politischen Sachfragen vorausgesagt werden können.

Auch Sigelman (1979) stützt sich bei seiner Prognose auf Umfrageergebnisse. Er untersucht, inwiefern die mittels *Surveys* ermittelte Popularität von Präsidentschaftskandidaten als Prognoseinstrument eingesetzt werden kann. Da ein starker Zusammenhang zwischen der Popularität und dem späteren Stimmenanteil belegt werden kann, postuliert Sigelman die Anwendungsmöglichkeit von *pre-election-polls* für Prognosezwecke (vgl. auch Brody/Sigelman 1983). Später lieferte Fiorinas (1981) Ansatz des retrospektiven Wählens eine theoretische Begründung für die Bedeutung der Popularität von Kandidaten für die individuelle Wahlentscheidung, auf welche sich folgende Ansätze vornehmlich stützen.

Mit Sigelmans (1979) Arbeit intensivierte sich die Entwicklung von Prognosemodellen durch Politikwissenschaftler und in kurzer Zeit wurden zahlreiche Vorschläge erarbeitet, die heute noch die Grundlage der aktuell diskutierten Prognosemodelle darstellen. Das erste „economy-popularity regression model“ von Lewis-Beck und Rice (1984) mutet mit zwei unabhängigen Variablen recht einfach an und gründet auch auf eher vagen theoretischen Argumenten. Zum einen lehnen sich die Autoren an die Postulate der *economic-voting*-Literatur an (vgl. hierzu Abschnitt 2.1.1). Zum anderen greifen sie auf Sigelmans (1979) Befunde zum engen Zusammenhang zwischen Regierungspopularität und Wahlergebnis zurück.

Die abhängige Variable des Regressionsmodells von Lewis-Beck und Rice (1984) stellt der Stimmenanteil des Kandidaten der Partei des amtierenden Präsidenten bei der kommenden Wahl dar. Vorhergesagt werden soll dieser Stimmenanteil durch die Popularität des Präsidenten und das Wirtschaftswachstum vor

der Wahl. Dazu werden zum einen die aggregierten Ergebnisse von Gallup-Umfragen zur Popularität des Präsidenten sechs Monate vor der Wahl sowie die vierteljährlich erfassten nationalen Wirtschaftswachstumsraten für die USA herangezogen. Damit wurden erstmals ökonomische und politische Variablen in einem Prognosemodell integriert. Die Autoren prägten mit diesem Vorschlag die bis heute in diesem Kontext dominanten so genannten quantitativen Ansätze.

Abramowitz (1988, 1994, 1996, 2004, 2008a, 2008b) erweitert das Modell um die Idee, dass die Länge der bisherigen Regierungszeit eine Rolle spielt. Hintergrund dieses Arguments ist die Überlegung, dass Regieren Stimmen kostet und somit Abnutzungstendenzen zu erwarten sind: „The longer a party has been in power, the more likely the public is to feel that ‚it’s time for a change““ (Abramowitz 1988: 844). Auch Lewis-Beck und Rice (1992: 45ff.) entwickelten ihr Modell weiter und ergänzten es um zentrale Variablen insbesondere des sozialpsychologischen Ansatzes (Campbell et al. 1964) und um Trendfaktoren. So wurden etwa die Parteistärke, gemessen über den Sitzverlust/-gewinn bei den letzten Wahlen, Kandidatenbewertungen oder prospektive Abschätzungen von Wirtschafts- und politischen Entwicklungen zusätzlich eingeführt (vgl. Lewis-Beck/Tien 1996, 2000, 2002, 2004, 2008).

Campbells (1994, 1996, 2008b; Campbell/Wink 1990) Modelle stützen sich zum einen wie die bisher angesprochenen auf ökonomische Indikatoren. Als politische Variablen nutzt Campbell jedoch die Resultate von Umfragen mit Wahlabsichtsfragen. Damit grenzt er sich von Abramowitz und Lewis-Beck ab, welche die Verwendung von Umfrageergebnissen vermeiden und etwa bei den Popularitätsindikatoren sofern möglich auf prozessproduzierte Aggregatdaten zurückgreifen.

Holbrooks (1996, 2007, 2008; Holbrook/DeSart 1999) Modelle verwenden neben der Popularität des Amtsinhabers die *time-for-a-change*-Variable von Abramowitz (1988) zur Vorhersage des Stimmenanteils des Präsidentschaftskandidaten der amtierenden Partei. Allerdings weicht er in Bezug auf die ökonomischen Variablen von der Sichtweise der bisherigen Modelle ab, denn so Holbrook (1996: 509): „After all, when entering the voting booth, perceptions of economic reality rather than the objective economic reality should matter most to voters“. Folgerichtig integriert er eine Variable zur Erfassung dieser subjektiven Einschätzung der Wirtschaftslage in sein Modell, nämlich den aggregierten und

mittels Umfragen erfassten retrospektiven finanziellen Status der Befragten zu einem hinreichend langen Zeitpunkt vor der Wahl.

Der Vorschlag von Norpoth (1996) zieht hingegen wieder das Wirtschaftswachstum heran, führt jedoch die Kandidatenunterstützung ein, um auch das Abschneiden des Oppositionskandidaten prognostizieren zu können. Zudem stützt er sich auf ein autoregressives Modell und wendet damit zeitreihenanalytische Verfahren an. Die bisher angesprochenen Vorschläge stützen sich auf multiple lineare OLS-Regressionen.

Wlezien und Erikson (1996) präsentieren ein Modell, das sich wiederum auf die Popularität des Amtsinhabers stützt, hinsichtlich der ökonomischen Variable jedoch auf die Veränderung des pro-Kopf-Einkommens zurückgreift und somit nur unwesentlich von anderen Vorschlägen abweicht.

Cuzán und Bundrick (1999, 2008) bauen auf den Einfluss von finanzpolitischen Aspekten auf das Wahlergebnis. In Anschluss an die klassischen *economic-voting*-Arbeiten von Fair (1978, 1982, 1988, 1996, 2002) und weiterführende theoretische Argumente zum Zusammenhang von Finanzpolitik und insbesondere der Unterstützung der jeweiligen Amtsinhaber (Cuzán/Heggen 1984; Niskanen 1979; Peltzman 1992), nutzen Cuzán und Bundrick die Resultate dieser Arbeiten, um ein eigenes Prognosemodell zu entwickeln. Demnach sind sechs Variablen zur Vorhersage notwendig: das Wirtschaftswachstum, die Inflationsrate, ein Indikator, der anzeigt, ob der amtierende Präsident noch einmal zur Wahl antritt, die Anzahl an Wahlperioden, die die Regierungspartei bereits im Amt ist, die Staatsausgaben als Anteil am Bruttosozialprodukt, die prozentuale und arithmetische Veränderung der relativen Staatsausgaben in der Regierungsperiode des Amtsinhabers sowie einen Indikator für die verfolgte Finanzpolitik in der letzten Legislaturperiode.

Schließlich legte Hibbs (2008) das so genannte *bread-and-peace*-Modell vor, wonach zur Erklärung und Vorhersage des Ausgangs von US-Präsidentenwahlen zwei Indikatoren ausreichen: der gleitende Durchschnitt des realen pro-Kopf-Einkommens über die Wahlperiode sowie die kumulierten Todesfälle des US-Militärs, welche nicht auf provozierte, feindselige Handlungen der US-Streitkräfte im Rahmen von Auslandseinsätzen zurückzuführen sind.

Neben diesen genannten quantitativen Modellen wurden weitere, vom bisherigen Vorgehen abweichende Vorschläge unterbreitet, welche skizziert werden sollen.

Das von Lichtman (2008a, 2008b) vorgeschlagene Verfahren basiert ebenfalls zentral auf der theoretischen Überlegung, dass die Ergebnisse von US-Präsidentenwahlen maßgeblich aus der Performanz der bisher amtierenden Regierungspartei bestimmt werden können. Wesentlicher Unterschied zu den bisher thematisierten Ansätzen ist, dass Lichtman auf eine weitaus größere Zahl von Indikatoren (so genannte *Keys*) zurückgreift, welche deutlich stärker historisch als ökonomisch motiviert sind. Die insgesamt 13 *Keys* umfassen etwa Informationen zum Abschneiden von Drittkandidaten bei der letzten Wahl, erreichte tiefgreifende Politikwechsel der amtierenden Regierung (als Beispiel werden etwa der *New Deal* genannt) und auch Indikatoren zu außenpolitischen Erfolgen und Fehlschlägen. Schließlich ist dieser Prognosevorschlag nicht regressionsbasiert, sondern fußt auf einer Indexbildung (Lichtman/Keilis-Borok 1981).

Die gänzliche Unabhängigkeit dieses Ansatzes von Umfrageergebnissen bringt zumindest zwei Vorteile mit sich: Zum einen sind Prognosen deutlich früher vor der Wahl möglich. Gerade bei Vorschlägen, welche die vergleichsweise volatile Kandidatenpopularität berücksichtigen, ist der *lead* der Prognose zumeist eher gering. Lichtman (2008a: IXf.) selbst betont in diesem Zusammenhang:

„The Keys indicate incumbent party success or failure long before the polls or any other forecasting models are of any value. Unlike many models developed by political scientists, the Keys include no polling data, but are based on the big picture...“.

Zum anderen kann auf eine deutlich längere Zeitreihe zur Evaluation des Modells zurückgegriffen werden. Problematisch an diesem Vorschlag erscheinen die teilweise willkürlich gesetzten Bewertungen von Ereignissen – bleibt etwa eine klare Bestimmung von zentralen politischen Programmen, die umgesetzt werden konnten, ebenso nebulös wie die Fassung der außenpolitischen Indikatoren.

Im Anschluss an Lichtmans Vorgehen entwickelten Armstrong und Graefe zwei weitere Vorschläge für indexverfahrenbasierte Prognosemodelle. Beim Modell *PollyIssues* werden allgemeine Einschätzungen zur Leistung der zur Wahl stehenden Kandidaten für die Vorhersage genutzt (Graefe/Armstrong 2008). *PollyBio* hingegen stützt sich auf biographische Informationen der Kandidaten mittels derer Vorhersagen über den Ausgang von US-Präsidentenwahlen ermöglicht werden sollen (Armstrong/Graefe 2009).

*PollyVote*²¹ (Cuzán et al. 2005; Graefe et al. 2009) versucht die bisher insgesamt aufgezählten Ansätze, Wahlabsichtsumfragen, Wahlbörsen und (quantitative) Prognosemodelle mit Experteneinschätzungen zu kombinieren und daraus Vorhersagen von Wahlresultaten abzuleiten. Hierzu werden die jeweilig neuesten Vorhersagen für den Stimmenanteil des Präsidentschaftskandidaten der aktuell amtierenden Partei mit jeweils gleicher Gewichtung berücksichtigt. Bei den Wahlabsichten wird der gleitende Durchschnitt der Voraussagen der letzten drei Tage herangezogen, die Expertenmeinungen werden ebenfalls mittels Umfragen ermittelt (Jones et al. 2007). Die Werte der Wahlbörsen speisen sich aus den *Iowa Electronic Markets* (vgl. Abschnitt 2.2.1), wobei *PollyVote* jeweils den gleitenden Durchschnitt der Schlusskurse des *vote-share*-Marktes der letzten sieben Tage heranzieht. Schließlich werden alle für die entsprechende Wahl verfügbaren quantitativen Modelle herangezogen und hieraus ein durchschnittlich zu erwartender Stimmenanteil berechnet.

Zentrale Idee hinter *PollyVote* ist, den allen Ansätzen inhärenten Prognosefehler durch die Sammlung möglichst vieler verschiedener Vorhersagen durch Mittelung zu reduzieren. Damit ist dies zwar kein theoretisches Prognosemodell, sondern vielmehr das pragmatische Ergebnis verschiedener Prognosebemühungen, welches seine Brauchbarkeit in Zukunft erst noch beweisen muss. Zumindest für die Präsidentschaftswahlen 2004 und 2008, für welche mittels *PollyVote* erstmals Prognosen vorgelegt wurden, zeigt sich, dass das Prinzip des *averaging out* viel versprechend zu sein scheint. Bei beiden Wahlen lieferte *PollyVote* gemessen am Stimmenanteil die genauesten Vorhersagen.

2.3.2 Das Prognosemodell für Bundestagswahlen von Gschwend und Norpoth

Anders als für die USA, in denen die Prognosemodelle zunehmende Bedeutung einnehmen und mittlerweile ein eigenes Forschungsgebiet darstellen (Campbell/Lewis-Beck 2008), liegt für Deutschland bisher nur eine Adaption eines quantitativen Prognosemodells vor. Dieses wurde von Gschwend und Norpoth (2001) entwickelt. Dieser Vorschlag orientiert sich an den bisher vorgestellten Modellen

21 Online finden sich die aktuellen Prognosen via *PollyVote*, *PollyIssues* und *PollyBio* sowie zahlreiche Hintergrundinformationen zu diesen Projekten auf <http://www.pollyvote.com>.

sowohl im Hinblick auf die Konstruktionsweise als auch auf die zugrunde liegende theoretische Basis.

Im Vergleich zur US-Präsidentenwahl erscheint die Vorhersage von Ergebnissen einer Bundestagswahl ungleich schwieriger. Während in den USA die Parteienlandschaft von zwei Parteien dominiert wird und deren Präsidentschaftskandidaten die Wahl faktisch unter sich ausmachen, beschränken sich die entsprechenden Modelle auf den Zweikampf des demokratischen mit dem republikanischen Präsidentschaftskandidaten und blenden (zumeist) mögliche weitere Kandidaten gänzlich aus. Das parlamentarische System Deutschlands kennt keine vergleichbare direkt gewählte Exekutive. Das (prognostische) Interesse bei Bundestagswahlen richtet sich daher auf die Stimmenanteile der einzelnen Parteien. Hierbei interessiert neben dem Abschneiden der großen Parteien CDU/CSU und SPD und damit der Festlegung des Kanzlers eben auch das Ergebnis der kleineren Parteien. Gerade vor dem Hintergrund der Fünf-Prozent-Klausel gewinnt dies besondere Bedeutung, da das Einziehen von zusätzlichen Parteien ebenso entscheidende Auswirkungen auf die Sitzverteilung im neuen Parlament hat wie auf mögliche Koalitionsoptionen. Insofern stellt schon aus diesen wahlrechtlichen Gesichtspunkten die Vorhersage von Stimmenanteilen einzelner Parteien eine deutlich größere Herausforderung dar als die Prognose des Siegers einer US-Präsidentenwahl.

Gschwend und Norpoth (2001) reduzieren ihr Prognosevorhaben im Prinzip auf ein Zwei-Parteienproblem, in dem sie als Ziel die Voraussage des gemeinsamen Stimmenanteils der aktuell amtierenden Regierungsparteien anvisieren. Der gemeinsame Stimmenanteil der regierenden Parteien soll mit Hilfe dreier Variablen prognostiziert werden: dem gleitenden Mittel der Stimmenanteile der Regierungsparteien bei den letzten drei Bundestagswahlen, der Zahl der Perioden, welche die amtierende Regierung bereits im Amt ist und dem Anteil der Wähler, der den Kanzler unterstützt. Mit diesen Indikatoren, so die Autoren, werden zentrale theoretisch wie empirisch als einflussreich ermittelte Faktoren erfasst.

Der *erste Parameter* deckt langfristige Parteibindungen ab, womit auf verschiedene, auch konfligierende, theoretische Konzepte Bezug genommen wird. Einer sozialpsychologischen Deutung (Campbell et al. 1964; Miller/Shanks 1996) zufolge sind langfristige Parteibindungen als bereits im Sozialisationsprozess gebildete Präferenzen aufzufassen, welche sich im Lebensverlauf mit zunehmender Wahlerfahrung verfestigen und eine Art psychologische Parteimitgliedschaft

(Converse 1969: 144) begründen. Aus einer Makroperspektive ergibt sich damit das Konzept des *normal vote* (Converse 1966), wonach man durch die Verteilung dieser langfristigen Parteibindungen den Ausgang einer Wahl unter üblichen Bedingungen ermitteln kann.

Soziologische Ansätze sehen langfristige Parteibindungen eher begründet durch den konflikthafter Charakter der Gesellschaft (grundlegend: Lipset/Rokkan 1967). Demnach bilden sich während des Prozesses der Staatsbildung soziale Spannungslinien (*cleavages*) in der Gesellschaft heraus. Diese *cleavages* werden im politischen Prozess durch Parteien repräsentiert, welche als „conglomerates of groups differing on wide ranges of issues, but still united in their greater hostility to their competitors in the other camps“ (ebd.: 6) gesehen werden. Demnach bedingen soziale Prozesse eine Gruppenzugehörigkeit, welche wiederum zu stabilem, gruppenkonformem Wahlverhalten führt.

Schließlich fasst Fiorina (1981: 89) langfristige Parteibindungen als „running tally of retrospective evaluations“ auf, welche in ein ökonomisches Wahlkalkül einfließen. Hiermit wird die Parteiidentifikation stark von der subjektiv bewerteten Leistungsbilanz der Parteien in der Vergangenheit abhängig gemacht (vgl. Pappi/Shikano 2007: 45).

Entsprechend der differierenden theoretischen Konzepte haben sich unterschiedliche Operationalisierungsstrategien herausgebildet: Während im Rahmen des *cleavage*-Ansatzes lediglich die Großgruppenzugehörigkeit des Wählers bekannt sein muss, um auf seine Parteibindung schließen zu können, hat sich insbesondere an die Übertragbarkeit des sozialpsychologischen Konzepts der Parteiidentifikation auf Deutschland eine rege Debatte angeschlossen (Berger 1973; Falter 1984). Ergebnis war ein weitgehender Konsens über die Anwendungsmöglichkeit des Konzepts, ein Standardinstrument zur Messung der Parteiidentifikation in Umfragen²² sowie zahlreiche empirische Befunde, die auf eine hohe Erklärungskraft hinweisen (Gabriel 1997, 2001; Jagodzinski/Kühnel 1990).²³

22 Die weithin benutzte Frage zur Ermittlung der Parteiidentifikation lautet: „Viele Leute in der Bundesrepublik neigen längere Zeit einer bestimmten Partei zu, obwohl sie auch ab und zu eine andere Partei wählen. Wie ist das bei Ihnen? Neigen Sie – ganz allgemein gesprochen – einer bestimmten politischen Partei zu? Wenn ja, welcher? Wie stark? (5-stufige-Skala von sehr schwach bis sehr stark)“.

23 Aktuelle Diskussionen ranken um die langfristige Stabilität der Parteiidentifikation. Hierbei wird zum einen die These des *dealignment*, das Abschmelzen des Anteils der Parteigezogenen

Die Operationalisierung von Gschwend und Norpoth (2001: 482f.) knüpft hierbei am ehesten an Fiorinas Konzept an, da mit dem gleitenden Mittel des Stimmenanteils für die Regierungsparteien über die letzten drei Bundestagswahlen weder die psychologische Komponente der Parteibindung abgebildet wird, noch die Gruppenzugehörigkeit als entscheidend angesehen wird. Vielmehr wird die Idee der retrospektiven Beurteilung von Parteien aufgegriffen und somit ein aggregierter langfristiger Wählerrückhalt erfasst. Diese Vorgehensweise bietet verschiedene Vorteile: Zunächst einmal ist diese Variable exogen – ein Umstand, der nicht ohne Weiteres für das psychologische Konzept der Parteiidentifikation behauptet werden kann. Zudem ist die Variable einfach und zeitlich sehr früh vor einer Wahl zu erfassen, was ebenfalls entscheidende Vorzüge für die Konstruktion eines Prognosemodells mit sich bringt. Problematisch an dieser Konzeption ist insbesondere, dass bei einer engen Auslegung eine Prognose einer Wahl erst nach drei vorliegenden Wahlen möglich erscheint. Dies bedeutet, dass Daten erst für die Bundestagswahlen 1961 und für Gesamtdeutschland ab 2002 zur Verfügung stehen. Dies hätte angesichts der bisher abgehaltenen 16 Bundestagswahlen eine unbotmäßige Reduktion der ohnehin geringen Fallzahl zur Folge.

Gschwend und Norpoth (2001) beziehen deshalb zum einen für die Wahlen von 1953 und 1957 jeweils nur den vorherigen Stimmenanteil als Indikator für die langfristige Parteienstärke heran, verzichten auf die Bundestagswahl 1949 und beziehen in ihr Modell den Stimmenanteil der Regierungsparteien in den alten Bundesländern mit ein, um das Problem für die ersten drei gesamtdeutschen Wahlen 1990, 1994 und 1998 zu umgehen.

Der *zweite Parameter* des Modells, die Anzahl der Amtsperioden der Regierungsparteien, bezieht sich auf die These, dass Regierungsverantwortung Stimmen kostet, und steht damit in Verbindung mit der *economic-voting*-Literatur sowie einer *Rational-Choice*-Perspektive des Wahlverhaltens. Demnach bewerten Wähler die Parteien retrospektiv in Bezug auf ihre Leistungen in der vergangenen Legislaturperiode (Downs 1968: 35ff.; Fiorina 1981: 65ff.). Dabei wurde herausgearbeitet, dass diese Bewertung Wählern für Regierungs-

diskutiert (Arzheimer 2006; Dalton 2005: Kapitel 9) und zum anderen die Stabilität der Parteiidentifikation im Lebensverlauf (Arzheimer/Schoen 2005; Schmitt-Beck et al. 2006). Für die USA würde jüngst das Ende der Ära des *partisan dealignment* verkündet und ein neuer Typ des *partisan voters* identifiziert (Bafumi/Shapiro 2009).

parteien einfacher fällt als für Oppositionsparteien (Kramer 1971). Weiterhin konnte gezeigt werden, dass Regierungen die Verantwortung für schlechte ökonomische Entwicklungen zugeschrieben wird (Nannestad/Paldam 2002; Paldam 1991). Schließlich zeigt Abramowitz (2008b) in seiner *time-for-a-change*-Modellierung für die USA, dass Regieren Stimmen kostet und mit zunehmender Regierungslänge eine Abwahl der amtierenden Partei wahrscheinlicher wird.

Hinsichtlich der empirischen Umsetzung dieses Konzepts ergeben sich nur insofern Schwierigkeiten, als dass in der Geschichte der Bundesrepublik durchaus während Legislaturperioden die Regierungskoalitionen wechselten und sich insofern manchmal die exakte Bestimmung der Zahl der Amtsperioden als schwierig herausstellte (ebenso wie die Festlegung der amtierenden Regierungskoalition). Dies betrifft vor allem die Bundestagswahlen von 1969 und 1983 (Gschwend/Norpoth 2001: 475f.). Gleichwohl wurde eine weitere Amtsperiode auch dann gezählt, wenn der kleinere Koalitionspartner zwischenzeitlich wechselte. So vermerken Gschwend und Norpoth (ebd.) für die Prognose der Bundestagswahl 1969 fünf Amtsperioden der Regierungsparteien; dies trifft für die CDU/CSU zu, welche seit 1949 Regierungsverantwortung inne hatte, der Koalitionspartner jedoch mehrfach wechselte. Auch kann der unterstellte lineare Zusammenhang in Zweifel gezogen werden, könnte man doch vermuten, dass zumindest bei der ersten Wiederwahl eine Regierung im Schnitt ihre Unterstützung ausbauen kann gegenüber der Wahl zuvor.

Als *dritter Parameter* wird eine kurzfristige Komponente berücksichtigt, welche im Wesentlichen den zeitlichen Vorlauf des Prognosemodells bestimmt. Gschwend und Norpoth (ebd.) sehen die herangezogene Kanzlerunterstützung, ermittelt aus Umfragen ohne Berücksichtigung der Unentschiedenen, in Zusammenhang mit der retrospektiven Leistungsbewertung der Regierungsarbeit, wie sie in Fiorinas (1981) Modell des retrospektiven Wählens in Abschwächung des klassischen ökonomischen Ansatzes zum Wählerverhalten von Downs (1968) entwickelt wurde. Gleichwohl könnte man die Einführung dieser Komponente auch mit der Kandidatenpräferenz, wie sie im sozialpsychologischen Ansatz als kurzfristig wirkender Faktor eingeführt wird, begründen.

Auch wenn die empirische Umsetzung dieses Parameters, da er sich auf Umfragedaten stützt, weniger elegant erscheint als die Operationalisierungen der beiden anderen Parameter, ist ein tragfähiges Prognosemodell wohl kaum ohne die Berücksichtigung kurzfristig wirkender Evaluationen von Parteien und/oder

Kandidaten denkbar, zumal zahlreiche empirische Befunde deren Wirksamkeit auch für Deutschland belegt haben (Debus 2008; Gabriel 2001; Jagodzinski/Kühnel 1990; Kellermann/Rattinger 2005; Ohr 2005; Pappi/Shikano 2001). Die Kanzlerpräferenz erscheint aus zweierlei Gründen nahe liegend: Erstens wird damit neben einer Kandidatenpräferenz die kurzfristige retrospektive Evaluation der Regierungspolitik abgedeckt, da der Kanzler als herausragender Repräsentant der Regierung fungiert. Somit ist zu erwarten, dass die Gesamtevaluation der Regierungsleistung sich auf die Popularität des Kanzlers niederschlägt. Zweitens liegen auch für frühere Wahlen Umfrageergebnisse zur Kanzlerpopularität vor. Dies wiederum erscheint gerade angesichts der wenigen zur Verfügung stehenden Datenpunkten essenziell.

Gleichwohl bleibt problematisch, dass sich nicht nur die Erhebungsinstrumente zur Messung der Popularität im Zeitverlauf gewandelt haben, sondern auch die methodischen Differenzen der Umfragen erheblich sind. So beziehen Gschwend und Norpoth (2001: 484) Angaben zur Kanzlerpopularität für die Jahre 1950 bis 1971 aus Umfragen vom IfD Allensbach, seit 1972 auf die Frage der Forschungsgruppe Wahlen zum gewünschten Bundeskanzler. Zudem variieren die Erhebungszeitpunkte, da nicht für alle Wahlen die entsprechenden Angaben zum gewünschten Zeitpunkt, etwa zwei Monate vor der jeweiligen Wahl, erhoben wurden.

Auch wenn das Modell einen ersten wichtigen Schritt in der deutschen Politikwissenschaft hin zur Erstellung von sozialwissenschaftlichen Prognosen darstellt, weist es in inhaltlicher Hinsicht eine entscheidende Schwäche auf: Das zentrale Interesse bei Bundestagswahlen gilt den Stimmenanteilen aller Parteien, welche eine reale Chance besitzen, die Fünf-Prozent-Hürde zu überspringen. Koalitionen spielen in der Bundesrepublik unzweifelhaft eine große Rolle und zumeist stellen sich regierende Koalitionen als „Paket“ wieder zur Wahl und wesentliches Interesse gilt in der Tat der späteren Regierungskoalition. Allerdings werden vor Wahlen nicht immer klare Koalitionsaussagen getroffen, angesichts des etablierten Fünf-Parteien-Systems sind zudem mehrere Koalitionsoptionen denkbar und wie nicht zuletzt am Beispiel der Landtagswahl in Hessen 2008 auch vor dem Hintergrund des Wahlergebnisses unausweichlich und schließlich stellen Große Koalitionen eine Schwierigkeit für das Modell dar. Normalerweise treten die beiden Koalitionspartner nach dieser Konstellation nicht als „Paket“ an und auch Gschwend und Norpoth (2001) betrachten in diesen

Fällen ausschließlich den Stimmenanteil des größeren Partners. Um mögliche Koalitionsoptionen und auch die Chance kleinerer Parteien ins Parlament einzuziehen, berücksichtigen zu können, erscheint es notwendig, die Stimmenanteile aller relevanten Parteien vorherzusagen.

2.3.3 *Zum prognostischen Wert von Prognosemodellen*

Die in diesem Abschnitt diskutierten Prognosemodelle weisen in der Regel einen zeitlichen Vorlauf von zwei bis drei Monaten auf und sind damit in Bezug auf den *lead* den anderen bereits erörterten Ansätzen zumindest ebenbürtig. Damit ist diesen Modellen im Hinblick auf die erste Anforderung an eine wissenschaftliche Prognose zu bescheinigen, dass diese erfüllt wird.

Die vorgestellten quantitativen Prognosemodelle beziehen sich zumeist auf theoretische Ansätze zur Erklärung des Wahlverhaltens und versuchen hieraus zentrale Einflussfaktoren abzuleiten, welche dann für die Prognose eines zukünftigen Wahlergebnisses genutzt werden. Hierbei stehen substantielle theoretische Einsichten im Fokus, welche sich bereits empirisch bewährt haben. Damit scheint die zentrale Anforderung an wissenschaftliche Prognosen, nämlich der Bezug auf empirisch bestätigte Gesetzmäßigkeiten, grundsätzlich sichergestellt. Allerdings müsste im Einzelfall geprüft werden, ob wirklich auf Mechanismen verwiesen wird, oder nicht vorwiegend lediglich empirisch gut bestätigte Korrelationen Basis der Regressionsmodelle sind. Ebenfalls kritisch sind die Modelle von Lichtman (2008a, 2008b) sowie *PollyVote* (Cuzán et al. 2005; Graefe et al. 2009) zu sehen, welche nicht auf Mechanismen verweisen, sondern entweder vor allem historische Faktoren indizieren, oder auf das Prinzip des *averaging out* setzen. Vor diesem Hintergrund ebenfalls problematisch erscheinen Zeitreihenmodelle. Diese versuchen zumeist ohne theoretische Bezüge allein auf Grundlage vorliegender Daten, die beobachteten Werte mit statistischer Finesse in die Zukunft fortzuschreiben.

Ein Problem ergibt sich hinsichtlich der korrekten Erhebung der Anfangsbedingungen. Schwierig erscheint die Erfüllung dieses Kriteriums vor allem in Bezug auf die etwa im Modell von Gschwend und Norpoth enthaltenen Kurzfristfaktoren, welche über Umfragen ermittelt werden. Allerdings sind auch „objektive“ Kriterien wie etwa das Wirtschaftswachstum hinsichtlich ihrer Validität

zu bezweifeln (etwa Lomborg 2001: 68). Selbst über die Ermittlung der Anzahl der Amtsperioden kann man hinsichtlich ihrer korrekten Messung unterschiedliche Positionen vertreten. Gerade in Bezug auf diesen Punkt ist sicher nie ein befriedigendes Ergebnis zu erreichen, jedoch ist allen hier angesprochenen Modellen zu Gute zu halten, dass versucht wird, auf möglichst einfache und objektiv zu messende Faktoren zurückzugreifen.

Die korrekte Spezifikation der Veränderung von Anfangsbedingungen vom Prognose- bis zum Wahlzeitpunkt scheint in diesem Fall unmöglich, da beispielsweise einschneidende Ereignisse die Kanzlerpopularität auch kurzfristig entscheidend beeinflussen können (Bytzek 2007). Entsprechend exogene Ereignisse wie etwa das Elbehochwasser 2002 können mittels solcher Modelle nicht berücksichtigt werden, da die Relevanz solcher Ereignisse nur post hoc ermittelt werden kann.

Insofern sind diese Modelle als konditionale Prognosen aufzufassen, da unterstellt wird, dass relevante Veränderungen der Anfangsbedingungen zwischen dem Zeitpunkt der Prognose und dem Auftreten des Explanandums nicht erfolgen. Eine Ausnahme hierbei bilden Modelle, welche zeitreihenanalytische Verfahren anwenden (z. B. Norpoth 1996) und einen etwaigen Trend bis zum Zeitpunkt der Wahl extrapolieren. Die Trendfortschreibung kontrolliert indes ebenfalls nur die im Modell erfassten Faktoren und unterstellt, dass alle anderen möglicherweise einflussreichen Faktoren konstant bleiben und sich die berücksichtigten gemäß dem postulierten Trend verändern.

Es ist weiterhin festzuhalten, dass diese Modelle einen pragmatischen Weg einschlagen. Der Identitätshypothese zufolge sollte derjenige Ansatz mit der höchsten Erklärungskraft auch die genauesten Prognosen liefern. Gerade die Wahlforschung hat jedoch gezeigt, dass sich die individuelle Wahlentscheidung auf viele Faktoren zurückführen lässt und befriedigende Erklärungen damit einen hohen Komplexitätsgrad aufweisen. Analysen von Gigerenzer (2007: 94ff.) zeigen, dass zwischen dem Komplexitätsgrad des Modells und der Prognosegenauigkeit ein umgekehrt u-förmiger Zusammenhang besteht. Demnach sollten entsprechende Prognosemodelle keinesfalls überkomplex sein (vgl. dazu auch Czerlinski et al. 1999; Goldstein/Gigerenzer 2008; Horst et al. 1941: 68ff.).

Diese Gesichtspunkte zusammenfassend erscheinen Prognosemodelle ein fruchtbares Forschungsfeld nicht nur zur Vorhersage von Wahlergebnissen, sondern auch zur Weiterentwicklung theoretischer Ansätze zum Wählerverhalten.

Bereits im Rahmen der Diskussion zu Wahlbörsen konnte gezeigt werden, dass gerade die empirischen Befunde zu experimentellen Märkten auch die theoretische Diskussion stimuliert hat, etwa in Form von Präzisierungen der Hayek-These durch Kou und Sobel (2004). Insofern wäre es wünschenswert, wenn die deutsche Wahlforschung ihren Prognoseskeptizismus ablegen und eine mit den USA vergleichbare Etablierung von verschiedenen Ansätzen erfolgen würde. Ein Vorschlag hierzu wird in Abschnitt 4 entwickelt.

2.3.4 Vorliegende Evidenz zur Güte von Prognosemodellen

Die meisten der vorliegenden Modelle für die USA weisen mit zwei bis fünf erklärenden Variablen einen überschaubaren Komplexitätsgrad auf und dennoch kann man zusammengefasst diesen Modellen ein durchaus hohes Maß an prognostischer Qualität zuschreiben (für Diskussionen des Abschneidens der US-Modelle vgl. Bartels/Zaller 2001; Campbell 2001, 2008a, 2009; Lewis-Beck 1985). Im Folgenden wird eingehender nur die Prognosegüte des Modells von Gschwend und Norpoth (2001) diskutiert.

Ausgangspunkt der Überprüfung der Brauchbarkeit von Prognosemodellen stellen zunächst *out-of-sample*-Prognosen dar, das heißt ex post wird jeweils ohne die interessierende Wahl mit den Daten, welche bis zu diesem Zeitpunkt vorgelegen hätten, eine „Prognose“ erstellt und diese dann mit dem tatsächlich eingetretenen Ergebnis verglichen.

Das Modell besteht diesen Test mit Bravour (Gschwend/Norpoth 2001, 2005; Norpoth/Gschwend 2003, 2004, 2009). Zunächst ist festzuhalten, dass mit den drei Parametern des Modells für die Bundestagswahlen 1953 bis 1998 fast 94 Prozent der Varianz des gemeinsamen Stimmenanteils der Regierungsparteien in den alten Bundesländern erklärt werden kann. Diese Anpassungsgüte vermag angesichts der insgesamt eher geringen Varianz des gemeinsamen Stimmenanteils (Mittelwert: 50,4 Prozent, Standardabweichung: 5,6 Prozentpunkte) nicht überraschen. Interessant ist jedoch, dass die modellbasierten *out-of-sample*-Prognosen im Schnitt nur um 1,5 Prozentpunkte vom späteren tatsächlichen Wahlergebnis abweichen (Gschwend/Norpoth 2001: 487ff.).

Problematisch gestaltet sich, dass das Modell aufgrund der Messung der langfristigen Parteibindung über das gleitende Mittel der Stimmenanteile bei den

letzten drei Wahlen nur Vorhersagen für die alten Bundesländer zulässt. Da eine Prognose des Stimmenanteils nur für die alten Bundesländer unbefriedigend wäre (Norpoth/Gschwend 2004: 8), werden für die ex-ante-Prognose der Bundestagswahlen 2002 die gesamtdeutschen Stimmenanteile für die jeweils vorhergehenden Wahlen herangezogen. Damit verlassen Norpoth und Gschwend die Strenge ihres Modells und weisen auch auf theoretische Schwierigkeiten hin, da gerade die Prägnanz des Konzepts langfristiger Parteieigungen in den neuen Bundesländern zweifelhaft erscheint (vgl. etwa Rattinger 1994). Vor dem Hintergrund der bisher dünnen Datenbasis an Bundestagswahlen für Gesamtdeutschland und des verhältnismäßig geringen Einflusses der neuen Bundesländer auf das Gesamtwahlergebnis – die neuen Bundesländer stellen weniger als ein Fünftel des deutschen Elektorats – erscheint dieses pragmatische Vorgehen jedoch vertretbar. Gleichwohl wird damit die Prognose unsicherer.

Dennoch trifft die ex-ante-Prognose für Gesamtdeutschland den tatsächlichen gemeinsamen Stimmenanteil von SPD und den Grünen mit 47,1 Prozent exakt, wenngleich auch hier im Zeitverlauf vor der Wahl je nach aktueller Kanzlerpräferenzumfrage schwankende Prognosen festzustellen sind (Norpoth/Gschwend 2004: 9). Auch für 2005, dem zweiten ex-ante-Testfall für das Modell, ist es gelungen, etwa zwei Monate vor der Wahl das gemeinsame Ergebnis von rot-grün mit nur 0,3 Prozentpunkten Abweichung vorherzusagen (Gschwend/Norpoth 2005). Dabei schneidet das Modell besser ab als alle umfragebasierten Voraussagen, wie die Autoren bemerken (ebd.: 68). Allerdings ist hier einzuwenden, dass zum einen die Prognosegüte für 2005 merklich auf eine ad-hoc-Modifikation zurückzuführen ist (Klein 2005b), und zum anderen der Vergleich von aufsummierten, vorhergesagten Stimmenanteilen einzelner Parteien mit einem Modell, das lediglich einen Stimmenanteil vorhersagt, nicht angemessen erscheint.

Der dritte ex-ante-Testfall des Modells, die Bundestagswahl 2009, stellt das Modell insofern vor Probleme, da 2005 bis 2009 eine Große Koalition regierte und die Vorhersage des gemeinsamen Stimmenanteils von CDU/CSU und SPD wenig interessant wäre. Norpoth und Gschwend (2009) präsentieren in der Folge getrennte Schätzungen für die CDU/CSU und die SPD. Weiterhin problematisch ist die Gründung der Partei Die Linke, entstanden aus der Fusion von PDS und WASG, die erstmals als einheitliche Partei bei einer Bundestagswahl antraten. Um diesem Umstand zu begegnen vermindern sie die Kanzlerkandidatenunter-

stützung reaktiv zur Unterstützung von Die Linke gemäß den Ergebnissen von Wahlabsichtsfragen (für eine Kritik an diesem bereits zuvor angewendeten Vorgehen siehe Klein 2005b). Vor dem Hintergrund der Vermutung einer negativen Wirkung der Existenz der neuen linken Partei für den Stimmenanteil der SPD vermuten Norpoth und Gschwend (2009), dass die langfristige Parteibindung mit ihrer Operationalisierung überschätzt wird und ziehen daher lediglich den Stimmenanteil der SPD bei der Bundestagswahl 2005 für die SPD-Prognose heran. Dies ist sicher ebenfalls eine durchaus kritisierbare ad-hoc-Modifikation des Modells.

Das Modell für die Bundestagswahl 2009 liefert somit Vorhersagen für die Stimmenanteile von CDU/CSU sowie SPD – im Juli 2009 ermittelte die Forschungsgruppe Wahlen eine Untertstützung von Angela Merkel von knapp 60 Prozent, während sich 30 Prozent der Befragten Frank-Walter Steinmeier als Bundeskanzler wünschten (Forschungsgruppe Wahlen 2009a). Gegeben diese Konstellation resultiert aus dem Modell die Vorhersage eines Stimmenanteils von 42,8 Prozent für CDU/CSU und 26,8 Prozent für die SPD. Damit sagte das Modell zwar korrekterweise eine Regierungsbildung unter Führung der CDU/CSU voraus, jedoch weichen die Stimmenanteile erheblich von den tatsächlichen ab (CDU/CSU: 9,0 Prozentpunkte; SPD: 3,8 Prozentpunkte).

3 Die Prognosegüte von Wahlabsichtsfragen

Bisher wurden verschiedene Ansätze zur Vorhersage von Wahlergebnissen hinsichtlich der Erfüllung wissenschaftstheoretischer Anforderungen an Prognosen sowie der empirischen Evidenz in Bezug auf ihre Genauigkeit diskutiert. Dabei wurde festgestellt, dass die Prognosebemühungen der deutschen Wahlforschung bisher unterentwickelt sind. Dies trifft sowohl zu auf die systematische Analyse der faktisch als Voraussage zu sehenden Sonntagsfragen als auch auf die Entwicklung (und Bewertung) von wissenschaftlichen Prognosemodellen.

Angesichts der herausgehobenen Bedeutung von Wahlen, dem großen öffentlichen Interesse an Wahlvorhersagen und des Entwicklungsstandes der empirischen Wahlforschung überrascht diese Forschungslücke. In diesem Abschnitt wird diesem Manko insofern begegnet, als eine systematische Analyse der Sonntagsfragen angestrebt wird. Dabei wird ein Test der Vorhersagegüte der Wahlabsichtsumfragen vorgelegt sowie eine methoden- und theoriegeleitete Analyse von Einflussfaktoren auf die Prognosegenauigkeit.

In Abschnitt 3.1 werden zunächst methodischen und theoretischen Überlegungen folgend potenzielle Einflussfaktoren auf die Genauigkeit von auf Wahlabsichtsumfragen basierenden Voraussagen abgeleitet. Die hierzu vorliegenden empirischen Befunde insbesondere aus den USA werden daran anschließend vor diesem Hintergrund in Abschnitt 3.2 skizziert. Methodische Aspekte der empirischen Analyse werden in Abschnitt 3.3 aufgegriffen. Dabei werden zunächst verschiedene Maßzahlen zur Bestimmung der Prognosegüte beziehungsweise des Prognosefehlers vorgestellt und deren Brauchbarkeit für die angestrebte Analyse diskutiert. Weiterhin werden die Datenbasis sowie knapp die verwendeten Auswertungsverfahren erörtert. Schließlich werden die empirischen Befunde präsentiert (Abschnitt 3.4), wobei zwei Schwerpunkte gewählt werden. Zum einen stehen deskriptive Befunde zur Prognosegüte von Sonntagsfragen im Mittelpunkt, zum anderen die hierauf Einfluss nehmenden Faktoren.

3.1 Einflussfaktoren auf die Prognosegüte von Wahlabsichtsfragen

Wahlabsichten werden mittels Bevölkerungsumfragen erfasst. Die individuellen Antworten werden dann aggregiert, um auf inferenzstatistischer Basis Populationsparameter für den zu erwartenden Stimmenanteil für die einzelnen Parteien und/oder Kandidaten zu schätzen. Aus diesem Grund erscheint es sinnvoll, zunächst systematisch zu prüfen, welche methodischen Aspekte die Genauigkeit der Prognose beeinflussen.

Darüber hinaus liegen verschiedene Ansätze zur Erklärung des Wählerverhaltens vor. Sie identifizieren substanzielle Einflussfaktoren auf die individuelle Wahlentscheidung. Diese Erkenntnisse können auf die vorliegende Fragestellung übertragen werden, da anzunehmen ist, dass die Genauigkeit der auf umfragenbasierten Voraussagen beachtlich mit der Sicherheit der anstehenden Wahlentscheidung variiert.

Nahe liegend wäre es weiterhin, Theorien zur Erklärung des Befragtenverhaltens zu berücksichtigen, wie sie beispielsweise Esser (1975, 1986, 1990) oder auch Noelle-Neumann (2001) entwickelt haben. Allerdings liegt weder ein empirisch hinreichend getesteter Ansatz vor, der stringent etwaige Einflussfaktoren auf die Voraussagegenauigkeit ableiten ließe, noch ist es mit der Datenlage möglich, die in den Ansätzen angesprochenen Faktoren zu testen (für einen Überblick über verschiedene Ansätze vgl. Reinecke 1991).

Bezüglich der methodischen Überlegungen sind zunächst *inferenzstatistische Aspekte* anzuführen. Demnach muss man sich vergegenwärtigen, dass die Genauigkeit der Schätzung von Populationsparametern auf Grundlage von Zufallsstichproben von der Stichprobengröße abhängt. Aus dem „Wurzel-N-Gesetz“ ergibt sich, dass mit steigender Stichprobengröße die Standardfehler kleiner werden. Genauer: Zur Halbierung des Standardfehlers – bei einfachen Zufallsstichproben – ist eine Vervierfachung der Stichprobengröße notwendig (Pokropp 1996: 30f.). Dieser Zusammenhang begründet *Hypothese 1: Je größer der Stichprobenumfang ist, desto geringer ist die Abweichung des geschätzten vom tatsächlichen Stimmenanteil.*

Eine weitere inferenzstatistische Überlegung erscheint in diesem Zusammenhang zwingend. Die Schätzung von Populationsparametern ist nur zulässig, sofern Zufallsstichproben verwendet werden (Kaplitza 1982: 137). Das IfD Al-lensbach stützt sich jedoch nach wie vor auf das Verfahren der Quotenauswahl,

eine Form der bewussten Auswahl (Noelle-Neumann/Köcher 2002a). Dieses Stichprobenverfahren ist dadurch gekennzeichnet, dass den Interviewern bestimmte Regeln für die Auswahl der Untersuchungseinheiten vorgegeben werden. Diese Regeln schreiben den Interviewern vor, wie viele Personen mit bestimmten Merkmalen befragt werden sollen. Schmidtchen (1961: 42ff.) weist darauf hin, dass es nicht ein einziges Quotenverfahren gibt, sondern diese sich je nach gewählten Quotierungsmerkmalen stark voneinander unterscheiden können. So quotiert das IfD Allensbach nach den Merkmalen Geschlecht, Alter, Berufstätigkeit und Berufsgruppe, wobei sich die jeweiligen Anteile aus der amtlichen Bevölkerungsstatistik ergeben (Noelle-Neumann/Petersen 2005: 255ff.).

Als Begründung für die Verwendung des Quoten- anstatt eines Zufallsverfahrens werden sowohl praktische als auch inhaltliche Gründe angeführt. Zu den angeführten praktischen Gründen gehören vor allem die vergleichsweise kostengünstige sowie die schnelle Durchführbarkeit (Moser/Kalton 1972: 134). Beide Aspekte haben jedoch seit der Etablierung von Telefonumfragen deutlich an Relevanz eingebüßt. Vermutlich sind mittlerweile telefonische Umfragen schneller und kostengünstiger durchzuführen als persönliche Umfragen auf Basis einer Quotenauswahl. Hierzu liegen jedoch dem Kenntnisstand des Autors zufolge keine systematischen empirischen Befunde vor.

Aus inhaltlicher Perspektive wird argumentiert, dass Quotenverfahren eine hinreichende Annäherung an mehrstufige Zufallsverfahren darstellen und „nachweislich zu einer repräsentativen Stichprobe“ (Schneller 1994: 10) führen. Vor allem bei vorliegenden Erfahrungen mit dem Verfahren soll es zumindest bei spezifischen Fragestellungen sogar Vorteile im Hinblick auf die Schätzung von Populationsparametern aufweisen (Eberlein 2001: 125; Kaplitz 1982: 167f.). Dies gilt Schmidtchen (1961: 93) zufolge insbesondere für Wahlabsichtsfragen, denn nur auf Grundlagen von Quotenstichproben seien realistische Wahlprognosen möglich. Als Grund hierfür wird unter anderem angeführt, dass perfekte Zufallsstichproben, welche dem Diktum der Inferenzstatistik genügen, allein aufgrund von bekannten systematischen Non-Response-Problemen nicht realisierbar erscheinen und durch Quotierungen diese systematischen Verzerrungen geringer ausfallen.

Zur Stützung dieser Einschätzung wurden am IfD Allensbach zeitgleich identische Befragungen zum einen auf Grundlage einer Quotenstichprobe und zum anderen auf Basis einer Zufallsauswahl durchgeführt. Anhand des deskriptiven

Vergleichs einzelner Ergebnisse zeigt Schmidtchen (1961) auf, dass – entgegen häufig geäußerter Vermutungen – keine wesentlichen Unterschiede bei beiden Befragungen gefunden werden konnten. Problematisch an der Studie ist, dass die Analysen unterkomplex sind, um die von Schmidtchen gezogenen Schlüsse unzweifelhaft zu stützen. So werden ausschließlich ausgewählte univariate Statistiken berichtet – Signifikanztests oder multivariate Verfahren vermisst man.

Der Haupteinwand gegen das Quotenverfahren ist theoretischer und weniger praktischer Natur. Die Auswahl der zu befragenden Personen auf der letzten Auswahlstufe erfolgt beim Quotenverfahren willkürlich (Kish 1965: 564). Scheuch (1956) spricht deshalb von einer „Auswahl nach Gutdünken“. Schmidtchen (1961) hingegen sieht diesen Einwand auf Grundlage der empirischen Resultate der IfD-Studie zugunsten des Quotenverfahrens entkräftet. Demnach seien auf der Ebene der Befragtenauswahl im Rahmen des Feldexperiments keine bedeutsamen Unterschiede zwischen den Befragungsarten festzustellen gewesen, vielmehr sei Non-Response beim Quotenverfahren seltener. Diese „Evidenz“ beruht jedoch auf Selbsteinschätzungen der Interviewer und erscheint deshalb wenig belastbar. Zudem ändert dieser praktische Einwand nichts am fehlenden Theoriefundament der Quotenauswahl.

Aus diesem Mangel folgern etwa Wendt (1960) sowie Schnell et al. (1999: 283), dass auf Grundlage von Quotenverfahren inferenzstatistische Schlüsse nicht zulässig sind, zumal keine validen Methoden vorhanden sind, um auf deren Grundlage beispielweise Stichprobenfehler zu berechnen. Aufgrund der fehlenden wahrscheinlichkeitstheoretischen Basis und der subjektiven Komponente bei der Probandenauswahl ist zu vermuten, dass im Durchschnitt geschätzte Populationsparameter auf Basis von Quotenstichproben größere Verzerrungen aufweisen als solche, welche auf Zufallsstichproben beruhen. Aus diesen Gründen wird vermutet, dass die Verwendung von Quotenverfahren mit größeren Prognosefehlern einhergeht. Mit den vorliegenden Daten (vgl. Abschnitt 3.3.1) ist diese Überlegung jedoch nicht stringent zu testen. Das einzige Institut, das nach wie vor Quotenstichproben anwendet, ist das IfD Allensbach. Insofern kann lediglich untersucht werden, ob deren Sonntagsfragen systematisch schlechter abschneiden als die der Konkurrenzinstitute. Durch die Überlagerung des vermuteten Effekts des Stichprobenverfahrens durch die institutsspezifische politische Gewichtung, kann jedoch dieser Effekt nicht isoliert werden.

Eine weitere methodische Überlegung stützt sich auf den *hypothetischen Charakter der Sonntagsfrage*. Labaw (1982: 95f.) formuliert eine Reihe von Annahmen, die zu treffen sind, möchte man mit Befragungen prospektiv Handlungen messen. Insbesondere müsse man unterstellen, dass die Befragten das in der Frage beschriebene Problem beziehungsweise die Situation analysieren können, außerdem in der Lage sind, darüber zu sprechen, und auch bereit sind, dies zu tun. Weiterhin muss angenommen werden, dass die Probanden sich in die hypothetische Lage der skizzierten Situation versetzen können und sich in dieser Situation mögliche Handlungsalternativen samt ihrer Konsequenzen vorstellen können.

In Bezug auf die Sonntagsfrage ist davon auszugehen, dass die Probanden, ausgenommen der vergleichsweise kleinen Gruppe der Erstwähler, sich gut in die Wahlsituation hinein versetzen können und grundsätzlich in der Lage sind darüber zu sprechen. Allerdings hängt der Grad des hypothetischen Charakters der Situation maßgeblich vom zeitlichen Abstand bis zur nächsten Wahl ab. Gerade in Zwischenwahlzeiten oder in zeitlicher Nähe zu anderen Wahlen, aber nicht der interessierenden Bundestagswahl, ist hier von größeren Schwierigkeiten auszugehen.

Auch von der Kenntnis der zur Auswahl stehenden relevanten Handlungsalternativen kann zwar grundsätzlich ausgegangen werden, aber auch diesbezüglich ist mit einem Zeiteffekt zu rechnen. Lange vor einer Wahl müssen nicht alle zur Wahl stehenden Parteien samt der jeweils für den Befragten entscheidungsrelevanten inhaltlichen Positionen oder Kandidaten bekannt sein. Zudem müssen zum Befragungszeitpunkt nicht einmal die antretenden Parteien letztlich feststehen – man denke hier nur an den Zusammenschluss der PDS mit der WASG zur Partei Die Linke im Vorfeld der Bundestagswahl 2005.

Auf Grundlage wahltheoretischer Überlegungen ist ebenfalls von einem Effekt des zeitlichen Abstands der Erhebung vom Wahltermin auszugehen. So können in diesem Zeitraum Ereignisse eintreten, die unter Umständen dazu führen, dass die Handlung der Befragten von der prospektiv berichteten abweicht. Hierbei ist an einschneidende Ereignisse wie etwa die Entführung von Hanns Martin Schleyer 1977, die Wiedervereinigung 1989 oder das Elbehochwasser 2002 zu denken (vgl. Bytzek 2007). Solche Ereignisse können Auswirkungen auf wahlentscheidungsrelevante Themen haben und *issues* in den Vordergrund rücken, welche zum Befragungszeitpunkt eine nur untergeordnete Rolle gespielt

haben. Auch beeinflussen solche einschneidenden Ereignisse die kurzfristige Kandidatenorientierung der Wähler. Gerade in Krisenzeiten können Kandidaten Führungsstärke oder -schwäche zeigen – mit Auswirkungen auf ihre Popularität. Schließlich ist auch von einem Einfluss solcher Ereignisse auf die Leistungsbewertung der Parteien auszugehen, beispielsweise in Bezug auf das gezeigte Krisenmanagement. Dabei sinkt die Wahrscheinlichkeit des Auftretens möglicher einschneidender Ereignisse, welche maßgeblich die Wählerentscheidung beeinflusst, mit zunehmender Nähe zum Wahltermin.

Dem mikrosoziologischen Ansatz der *Columbia School* zufolge beeinflussen insbesondere soziale Kontextfaktoren die individuelle Wahlentscheidung: „A person thinks, politically, as he is, socially. Social characteristics determine political preference“ (Lazarsfeld et al. 1955: 27). Es ist jedoch möglich, dass Personen in Netzwerke mit diffusen oder gar widersprüchlichen Parteipräferenzen eingebettet sind. Lazarsfeld et al. (ebd.: 56ff.) sprechen hier von „cross-pressure“, welchem die Wähler in diesem Fall unterliegen. Für diese Gruppe wird dem Ansatz zufolge davon ausgegangen, dass sie ihre Wahlentscheidung erst im Laufe des Wahlkampfes trifft. Demnach ist zu vermuten, dass der Anteil Unentschiedener bis zum Wahltermin kontinuierlich sinkt und damit auch die Wahlabsichtsfragen bessere Schätzungen der Stimmenanteile ermöglichen.

Empirisch bestätigt diese Vermutung Blais (2004). Er zeigt in einer ländervergleichenden Studie (ohne Berücksichtigung Deutschlands), dass die Häufigkeit eines Meinungswechsels vor der Wahl in der Tat wie angenommen mit Fortschreiten der Wahlkampagne abnimmt. Auch für Deutschland gibt es im Hinblick auf die Veränderung des Ausmaßes der vor der Wahl Unentschiedenen empirische Befunde. So berichten Schmitt-Beck und Faas (2007: 76ff.), dass etwa ein Viertel der Wahlberechtigten sich erst in der heißen Wahlkampfphase für eine Partei entscheidet. Diesen Befunden zufolge sind sogar am Wahltag noch neun Prozent der Wahlberechtigten unentschieden.

Insgesamt liegen demnach gewichtige Argumente vor, um davon auszugehen, dass der zeitliche Abstand einer Wahlabsichtsumfrage zur Wahl eine entscheidende Einflussgröße darstellt. Hierbei ist jedoch anzunehmen, dass dieser Effekt nicht linear verläuft, sondern einen abnehmenden Zuwachs an Genauigkeit aufweist. Insgesamt führen diese Überlegungen zu *Hypothese 2: Je kürzer der zeitliche Abstand zwischen Befragung und Wahltermin, desto besser approximieren Wahlabsichtsumfragen den späteren Stimmenanteil, wobei die Genauig-*

keitszunahme mit sinkendem zeitlichen Abstand zwischen Umfragezeitpunkt und Wahltermin abnimmt.

Aus methodischer Hinsicht erscheint weiterhin die *Qualität der Umfrage* entscheidend. Als zentraler Indikator hierfür wird in der Regel die Ausschöpfungsquote angesehen (Neller 2005; Wüst 1998). Diese wird von den durchführenden Instituten zumeist nicht berichtet, weswegen sie nicht als Qualitätsindikator herangezogen werden kann (vgl. Abschnitt 3.3.1).

Deshalb wird von dem Umstand Gebrauch gemacht, dass die Wahlabsicht normalerweise in regelmäßigen Abständen erhoben und publiziert wird, wobei hinsichtlich des zeitlichen Modus und auch des Veröffentlichungstermins Differenzen zwischen den einzelnen Instituten bestehen. So wird etwa das Politbarometer der Forschungsgruppe Wahlen monatlich – mit in der Regel einer Sommerpause im Juli oder August – erhoben und veröffentlicht, während Forsa wöchentlich Umfragen durchführt. Insbesondere in den Wochen vor einer Wahl nimmt die Zahl an durchgeführten Vorwahlumfragen stark zu, etwa in Form so genannter Blitzumfragen der Forschungsgruppe Wahlen.

Entsprechend variiert auch der Befragungszeitraum der für eine Umfrage angesetzt wird. Da je nach Umfang der Befragung über verschiedene Tage und Tageszeiten die Ausfälle systematisch variieren, kann nun vermutet werden, dass in Umfragen mit einem längeren Erhebungszeitraum diese systematischen Verzerrungen geringer ausfallen als in kürzeren Befragungszeiträumen. Darüber hinaus muss berücksichtigt werden, ob die Umfragen auch am Wochenende oder nur werktags durchgeführt werden. *Hypothese 3a* postuliert entsprechend: *Je länger der Befragungszeitraum, desto besser approximiert die Wahlabsichtsumfrage den tatsächlichen Stimmenanteil.* Und *Hypothese 3b* lautet: *Wenn die Befragung werktags und am Wochenende durchgeführt wird, dann fällt der Voraussagefehler geringer aus als bei Befragungen, welche ausschließlich an Werktagen durchgeführt werden.*

Bisher wurde von der gängigen Praxis der Berichterstattung von Sonntagsfragen in Deutschland insofern abstrahiert, als dass eine einfache Aggregation individueller Angaben unterstellt wurde. Allerdings werden vor der Veröffentlichung politische Gewichtungsfaktoren eingesetzt, insbesondere um den Anteil der zum Zeitpunkt der Befragung Unentschiedenen auf die einzelnen Parteien zu verteilen. Da die genauen Vorgehensweisen hierbei nicht veröffentlicht werden,

können diese nicht im Einzelnen analysiert werden (vgl. zu dieser Praxis Abschnitt 2.1.3).

Die Forschungsgruppe Wahlen stellt ihre Rohdaten für wissenschaftliche Zwecke zur Verfügung. Anhand eines Vergleichs dieser Daten mit den berichteten Projektionen ist eine Abschätzung des Ausmaßes der politischen Gewichtung zumindest für dieses Institut möglich. Hierbei zeigen sich die in Tabelle 3.1 berichteten Abweichungen für die einzelnen Parteien. Auch wenn die mittleren Differenzen eher gering erscheinen, ist die große Varianz auffällig sowie die ausgewiesenen Minima und Maxima (für eine detaillierte Analyse des Ausmaßes der politischen Gewichtung siehe Abschnitt 3.4.1). Zusammengefasst deuten diese Befunde auf einen erheblichen Einfluss der politischen Gewichtung hin.

Neben der politischen Gewichtung, die nach hausinternen Geheimrezepten erfolgt, unterscheiden sich die einzelnen Institute hinsichtlich einiger nicht kontrollierbarer methodischer Aspekte. Zu denken ist hier an spezifische Frageformulierungen, die Anzahl und Reihenfolge der Antwortkategorien, die Platzierung der Sonntagsfrage im Fragebogen, die Ausbildung der Interviewer oder auch an den Umgang mit Non-Response-Problemen (siehe zu den Institutsspezifika auch Abschnitt 2.1.3).

Tabelle 3.1: Verteilungsmerkmale der Abweichungen der Rohdaten von den Projektionen der Forschungsgruppe Wahlen, 1986-2005

	Mittelwert	Standard-abw.	Varianz	Min.	Max.
CDU/CSU	-0,8	4,0	16,3	-13,4	10,9
SPD	1,3	3,7	13,4	-10,1	10,1
FDP	-1,2	1,1	1,2	-4,8	2,6
Bündnis 90/Die Grünen	0,9	1,3	1,6	-2,3	5,9
PDS/Die Linke	-0,5	1,0	1,0	-2,6	5,8

Anmerkungen: Die zugrunde liegende Datenbasis der hier präsentierten Auswertungen wird in Abschnitt 3.3.1 erläutert.

Weiterhin einflussreich ist in diesem Zusammenhang die Komplexität des angewendeten Zufallsstichprobenverfahrens. Lipsmeier (1999) zeigt, dass die Berechnung korrekter Standardfehler mit dem Auswahlverfahren variiert. Zentral

hierbei ist, dass die demoskopischen Institute zumeist geschichtete Zufallsstichproben nach Empfehlungen des Arbeitskreises Deutscher Markt- und Sozialforschungsinstitute e. V. ziehen (vgl. etwa ADM 1979; Häder/Glemser 2006; Hoffmeyer-Zlotnik 1997). Daraus resultieren jedoch gegenüber einfachen Zufallsstichproben – bei Konstanzhaltung der Fallzahlen – unterschiedlich große Standardfehler. Somit wird vermutet, dass auch die Genauigkeit der Stimmenanteilsvorausagen mit dem Komplexitätsgrad des Stichprobenverfahrens variiert. Da die Institute jedoch keine genauen Angaben zu den angewendeten Verfahren machen, ist ein Test dieser These nicht möglich.

Um auf diese unbeachtbaren Institutseigenheiten zu kontrollieren, werden in den folgenden Analysen *house effects* (Crespi 1988: 149ff.; Smith 1978, 1982a) berücksichtigt. Damit ist es zudem möglich, zu überprüfen, ob einzelne Meinungsforschungsinstitute aufgrund ihrer Parteinähe systematische Verzerrungen zugunsten oder zuungunsten bestimmter Parteien in ihren berichteten Voraussagen aufweisen (Noelle-Neumann 1983: 49). Dies wird geprüft, da bereits in Abschnitt 2.1.1 die enge Verbindung zwischen der politischen Meinungsforschung und der Politik beziehungsweise einzelnen Parteien herausgearbeitet wurde. Diese enge Verzahnung führt zu dem immer wieder öffentlich lautstark formulierten Postulat parteispezifisch verzerrter Umfrageergebnisse einzelner Institute. Darüber hinaus liegen Hinweise vor, dass politische Umfrageergebnisse auch gezielt als Wahlkampfinstrument genutzt werden (vgl. Raupp 2007: 122).

Eine empirische Analyse möglicher systematisch verzerrter Wahlabsichtsumfragen nach Institut wurde bisher jedoch noch nicht vorgelegt. Aufgrund des Konkurrenzdrucks zwischen den einzelnen Instituten wird jedoch vermutet, dass zumindest dauerhaft systematisch fehlerhafte Stimmenanteilsvorausagen zugunsten oder zuungunsten von bestimmten Parteien nicht aufrecht erhalten werden können. Darüber hinaus deutet eine Untersuchung des Umgangs der Medien mit Wahlumfragen darauf hin, dass eine entsprechende Selektion nach politischer Orientierung viel eher auf dieser Ebene stattfindet. Demnach berichten Medien über Umfrageergebnisse selektiv nach ihrer politischen Ausrichtung (Wolf 2007: 70ff.).

Insgesamt wird folglich mit *Hypothese 4a* vermutet: *Es bestehen institutspezifische Unterschiede hinsichtlich der Prognosegenauigkeit. Hypothese 4b* postuliert: *Die Institute weisen keine systematische Über- oder Unterschätzung einzelner Parteistimmenanteile auf.*

In Bezug auf die hier aufgestellten Hypothesen ist zudem von einer Überlagerung der *house effects* mit den einzelnen potenziellen Einflussfaktoren auszugehen (Zukin 2004). Insofern wird es insbesondere aufgrund des gewichtigen Einflusses der politischen Gewichtung eventuell nicht möglich sein, entsprechende methodische Effekte zu isolieren.

Im Hinblick auf Überlegungen zu Theorien des Wählerverhaltens ist daran zu denken, dass bei der prospektiven Schätzung des Stimmenanteils einzelner Parteien implizit die *Höhe der Wahlbeteiligung* mit berücksichtigt werden muss. In den Umfragen wird der Proband zumeist zunächst nach der Wahrscheinlichkeit gefragt, mit der er an der kommenden Wahl teilnehmen wird. Wird der Urnengang ausgeschlossen, so wird dem Befragten die Sonntagsfrage erst gar nicht vorgelegt und er wird nicht weiter berücksichtigt (vgl. Abschnitt 2.1.3). Der somit ermittelte potenzielle Nichtwähleranteil unterschätzt üblicherweise den tatsächlichen deutlich. In der Folge berücksichtigt man zur Ermittlung der Stimmenanteile auch Angaben von Personen, die gar nicht an der Wahl teilnehmen werden. Da sich die Gruppe der Nichtwähler nicht zufällig über die Wahlberechtigten verteilt (Armingeon 1994; Blais 2006; Fowler et al. 2008; Kühnel 2001), beeinflusst dies auch die Genauigkeit der Voraussage.

Crespi (1988: 94f.) weist in diesem Kontext auf einen weiteren Grund für die Relevanz der Wahlbeteiligung hin. Er argumentiert, dass im Vorfeld von Wahlen mit niedriger Wahlbeteiligung auch in Umfragen der Anteil potenzieller Nichtwähler steigt. Diese werden jedoch in der Regel nicht nach ihrer Wahlabsicht befragt. Lässt man in diesem Fall die Stichprobengröße unverändert, beruht die Stimmenanteilsschätzung auf insgesamt weniger Probanden. Dies wiederum führt zu breiteren Konfidenzintervallen und somit potenziell zu einer schlechteren Voraussage. Aus diesen Argumenten folgt *Hypothese 5: Je niedriger die Wahlbeteiligung bei einer Wahl, desto größer ist der Voraussagefehler von Wahlabsichtsumfragen.*

Neben der Höhe der Wahlbeteiligung ist die *Stabilität des Wählerverhaltens* zu berücksichtigen. Ausgehend vom sozialpsychologischen Ansatz (Campbell et al. 1954, 1964), welcher als zentralen Bestimmungsfaktor des Wählerverhalten langfristige Parteibindungen postuliert, entwickelte Converse (1966) das Makrokonzept der Normalwahl. Diese liegt vor, wenn alle Wähler gemäß ihrer lang- und mittelfristigen Parteipräferenzen wählen (für eine Analyse für Deutschland vgl. etwa Pappi/Shikano 2002).

In der Regel werden diese mittel- und langfristigen Parteipräferenzen über das Konzept der Parteiidentifikation operationalisiert, welche bereits im Rahmen der Sozialisation gebildet wird und sich im Lebensverlauf mit zunehmender Wahlerfahrung verfestigt (für eine zusammenfassende Darstellung siehe Schoen/Weins 2005). Verfügt nun ein Großteil der Wählerschaft über eine stabile Parteiidentifikation und übt diese einen deutlichen und über die Zeit stabilen Einfluss auf das Wählerverhalten aus, gestaltet sich die Voraussage des Wahlergebnisses einfacher.

Nimmt man diese Konzeption ernst, dann ist zu erwarten, dass in den neuen Bundesländern zum einen weniger Personen überhaupt eine Parteibindung aufweisen und diese zum anderen deutlich schwächer ausgeprägt ist als in den alten Bundesländern. Diese Überlegung wird durch empirische Befunde von Arzheimer und Falter (2005) gestützt: Sie zeigen erstens, dass die Volatilität des aggregierten Wählerverhaltens bei Landtags- und Bundestagswahlen in den neuen Bundesländern von 1990 bis 2003 deutlich höher ist als in den alten Bundesländern. Zweitens wird deutlich, dass in den neuen Bundesländern der Anteil von Wählern mit vorhandener Parteiidentifikation deutlich geringer ist als in den alten Bundesländern. Schließlich zeigen ihre Analysen, dass Kurzfristfaktoren für die Wahlentscheidung der Wähler in den neuen Bundesländern bedeutender sind als für die Wähler in den alten Bundesländern.

Damit ist selbst im Fall einer perfekten Normalwahl im Sinne von Converse (1966) zu erwarten, dass aufgrund der schwächer ausgeprägten mittel- und langfristigen Parteibindungen bei den Wählern in den neuen Bundesländern hier eine Voraussage der Stimmenanteile einzelner Parteien schwieriger ist als in den alten Bundesländern. Demnach lautet *Hypothese 6: Der Voraussagefehler von Wahlabsichtsumfragen ist in den neuen Bundesländern größer als in den alten Bundesländern.*

In der Längsschnittperspektive wird in diesem Zusammenhang das Phänomen des *partisan dealignment* diskutiert. Empirisch wird der postulierte Trend für Deutschland zumindest in Bezug auf einen abnehmenden Anteil an Wählern mit fester Parteibindung als bestätigt angesehen (Dalton 2005: 189; Dalton/Bürklin 2003; Gabriel 1997), wenngleich diese Studien notwendige Zeitreihenanalysen vermissen lassen. Auf individueller Ebene scheinen die Parteibindungen jedoch höchst volatil, so das Ergebnis einer Untersuchung von Schmitt-Beck et al. (2006) auf Grundlage des sozio-oekonomischen Panels. Hinsichtlich der

Veränderung des Einflusses der Parteibindung auf das Wählerverhalten wird von einer stabilen Prägekraft ausgegangen, wie die Analysen von Gabriel (2001), Kellermann (2007) sowie Kellermann und Rattinger (2005) zeigen.

Als ein Mechanismus für das zu beobachtende *partisan dealignment* wird das insgesamt gestiegene Bildungsniveau gesehen. Demnach ist zunehmend mehr Wählern erlaubt, das Wahlsystem zu durchschauen, weshalb in der Folge seltener auf einfache Heuristiken bei der Wahlentscheidung zurückgegriffen wird. Dementsprechend büßt die Parteiidentifikation als *information shortcut* (Popkin 1991: 44ff.) an Bedeutung ein. Schmitt-Beck (1993) zeigt auf Grundlage von Auswertungen der deutschen Wahlstudie von 1990 zwar, dass in den alten Bundesländern nur rund ein Drittel und in den neuen Bundesländern ein Viertel der Wahlberechtigten profunde Kenntnisse des Wahlverfahrens aufweisen, jedoch findet er auch Belege dafür, dass der Kenntnisstand mit dem Bildungsniveau in der angenommenen Richtung variiert.

Als weitere Mechanismen werden die Veränderung der Massenmedien und Individualisierungstendenzen angeführt. Dies führe zu einer Verschiebung der relativen Bedeutung von Einflussfaktoren auf die Wahlentscheidungen der Wähler zuungunsten der Parteibindungen und zugunsten des Einflusses von Wahlkämpfen, Meinungsumfragen und auch Kandidaten (vgl. für entsprechende Argumente bspw. Neu 2008; Ohr 2005). Die These der Veränderung der Massenmedien beziehungsweise spezieller der veränderten Berichterstattung in der Vorwahlzeit mit Tendenzen zum *horse-race*-Journalismus werden weithin geteilt (Brettschneider 1996; Hohlfeld 2006). Empirisch umstritten sind indes die Thesen sich zunehmend individualisierender Wahlentscheidungen und einer stetigen Bedeutungszunahme des Einflusses von Kandidatenpräferenzen und Wahlkämpfen auf das Wählerverhalten.

So sehen Schnell und Kohler (1995) die Individualisierungsthese anhand ihrer Analysen zu Einflussfaktoren auf die Parteipräferenz bestätigt. Müller (1997) hingegen führt diese Befunde im Wesentlichen auf methodische Artefakte zurück und auch Jagodzinski und Quandt (1997) finden in ihrer Replik auf die Arbeit von Schnell und Kohler kaum empirische Belege dafür. Pappi und Shikano (2001) können anhand der Politbarometer-Daten eine Personalisierung der Wahlentscheidung, gemessen am Kandidateneinfluss auf die individuelle Wahlentscheidung, seit den 1980er Jahren nicht feststellen. Kellermann und Rattinger (2005) analysieren den relativen Einfluss verschiedener Faktoren auf die Wahl-

entscheidung für die CDU/CSU oder die SPD. Sie folgern aus ihren Analysen, dass für „die beiden großen Volksparteien CDU/CSU und SPD (...) vieles für einen schwankenden Einfluss der drei Bestimmungsfaktoren [spricht], der sich aus der jeweils spezifischen Ausgangslage einer jeden Bundestagswahl ergibt“ (ebd.: 209). Ein Trend in Richtung einer Relevanzverschiebung ist ihnen zufolge nicht festzustellen, wobei die Studie nur drei Wahlen umfasst und damit einen zu kurzen Zeitraum, um in Bezug auf die zugrunde liegende These Schlüsse zu ziehen.

Die empirische Erforschung der Ausgestaltung und Wirkung von Wahlkämpfen erfolgt zumeist für einzelnen Wahlen (Schmitt-Beck 2001; Schmitt-Beck/Faas 2007; Schoen 2004a) oder Ereignisse wie die populär gewordenen TV-Duelle (Klein 2005a; Klein/Rosar 2007). In einer Analyse der Wahlkämpfe von 1980 bis 1998 untersucht Schoen (2004b) zwar nicht direkt den Einfluss von Wahlkämpfen auf das Wahlverhalten, er findet jedoch Belege dafür, dass Bundestagswahlkämpfe die Kandidatenorientierungen der Wähler beeinflussen – mit wahlspezifischen Variationen. Entsprechend kommt Schmitt (2001) zu dem Schluss, dass etwaige Trendhypothesen bisher empirisch keineswegs abschließend beantwortet sind.

Auch aus methodischen Überlegungen heraus sind Veränderungen der Vorhersagegenauigkeit im Zeitablauf zu vermuten. Einerseits ist davon auszugehen, dass mit zunehmender Verbreitung von Telefonumfragen und verbesserten technischen Möglichkeiten durch etwa die Einführung von CATI-Systemen und *Random-Digit-Dialing*-Verfahren (Häder/Glemser 2006: 149) die Präzision der Messung der Wahlabsicht zunimmt. Mittlerweile stützen sich mit Ausnahme des IfD Allensbach alle großen Meinungsforschungsinstitute maßgeblich auf Telefonumfragen. Da diese eine qualitativ hochwertige Stichprobenziehung erlauben und umfassende Kontrollmöglichkeiten der gesamten Feldphase ermöglichen, entwickelten sie sich zum bevorzugten Befragungsinstrument (Steeh 2008: 223).

In jüngerer Zeit sehen sich Telefonumfragen jedoch verschiedenen Herausforderungen ausgesetzt. So wird häufiger postuliert, dass aufgrund einer Übersättigung potenzieller Probanden die Ausschöpfungsquoten bei Telefonumfragen sinken. Die vorliegenden empirischen Befunde können diese Vermutung jedoch nicht bestätigen (Stögbauer 2000; Tortora 2004). Unbestritten ist jedoch, dass Telefonumfragen durch die Abnahme von Festnetzanschlüssen und den Bedeutungszuwachs von Mobilfunktelefonen insbesondere bei Spezialpopulationen vor neue Herausforderungen gestellt sind (Häder/Glemser 2006: 163ff.; Steeh 2008:

232ff.), welche zumindest vermuten lassen, dass die Zunahme der Präzision der Messung im Zeitablauf nicht linear erfolgte. Vielmehr wird angenommen, dass die Präzision mit abnehmender Rate zunimmt und möglicherweise aus den genannten Gründen in jüngster Zeit sogar von einer sinkenden Präzision auszugehen ist.

Aufgrund der verschiedenen theoretischen wie methodischen Argumente und widersprüchlichen empirischen Befunde, wird geprüft, ob sich in Bezug auf die Genauigkeit der Wahlvoraussagen auf Basis von Wahlabsichtsumfragen Trends erkennen lassen. Die skizzierten Aspekte führen zum einen zu *Hypothese 7a: Im Zeitablauf wird der Voraussagefehler von Wahlabsichtsumfragen größer*. Zum anderen wird *Hypothese 7b* geprüft: *Der Voraussagefehler von Wahlabsichtsumfragen weist im Zeitablauf einen umgekehrt u-förmigen Verlauf auf*.

Die Diskussion potenzieller Einflussfaktoren auf die Vorhersagegenauigkeit von Wahlabsichtsumfragen abschließend sind in Tabelle 3.2 die abgeleiteten und im Folgenden empirisch zu prüfenden Hypothesen zusammengefasst.

Tabelle 3.2: Potenzielle Einflussfaktoren auf die Vorhersagegenauigkeit von Wahlabsichtsumfragen

Hypothese	Einflussfaktor	Effekt auf die Vorhersagegenauigkeit
1	Zunahme der Stichprobengröße	steigt
2	Zeitlicher Abstand zum Wahltermin sinkt	steigt mit abnehmender Grenzrate
3a	Zunahme der Länge des Befragungszeitraums	steigt
3b	Befragung an Werktagen und am Wochenende	steigt
4a	house effects	liegen vor
4b	Institutsspezifischer Parteibias	kein Effekt
5	Höhere Wahlbeteiligung	steigt
6	Vorhersagen für neue Bundesländer	sinkt
7a	Zeitverlauf	sinkt
7b	Zeitverlauf	umgekehrt u-förmig

Anmerkung: Eigene Zusammenstellung.

3.2 Forschungsstand

Für Deutschland hat bisher nur Antholz (2001: 147ff.) mögliche Korrelate der Genauigkeit von Wahlabsichtsumfragen thematisiert. Hierzu ist jedoch zu bemerken, dass die Identifikation von Einflussgrößen weder zentrales Anliegen seiner Arbeit ist, noch diese systematisch aus methodischen und/oder theoretischen Überlegungen hergeleitet werden. Vielmehr werden entsprechende Faktoren ad hoc eingeführt und empirische Resultate hierzu beschrieben. Hinsichtlich der angewendeten Prüfverfahren ist zudem zu konstatieren, dass diese unterkomplex sind. Antholz stützt sich überwiegend auf deskriptive und bivariate Befunde und lässt angemessene multivariate Analysen vermissen.

Aus einer Gegenüberstellung der Entwicklung der absoluten mittleren Fehler und der durchschnittlichen Wahlbeteiligung über alle Wahlen hinweg folgert Antholz (ebd.: 107), dass mit einer sinkenden Wahlbeteiligung die „Wahlprognosevalidität“ abnimmt.

Ausführlich werden *house effects* thematisiert (ebd.: 151ff.). Ziel dabei ist ein Ranking zu erstellen. Dieses stützt sich nicht nur auf die Performanz der Voraussagen, sondern auch auf Ergebnisse einer Politikerbefragung zu ihrer Einschätzung der einzelnen Institute. Den Auswertungen zu Folge zeigen sich durchaus institutsspezifische Unterschiede hinsichtlich der durchschnittlichen absoluten Abweichung der Voraussagen vom tatsächlichen Wahlergebnis, welche allerdings nicht besonders groß ausfallen. Das IfD Allensbach weist demnach mit 1,7 Prozentpunkten die niedrigste mittlere absolute Abweichung auf, während Infratest dimap mit 2,2 Prozentpunkten die größte zeigt. Unklar bleibt hier allerdings sowohl, ob die berichteten Unterschiede signifikant sind, als auch die mögliche Beeinflussung dieser Befunde durch andere Faktoren.

Schließlich untersucht Antholz (ebd.: 108ff.) den Einfluss der Art der Wahl und kann zeigen, dass zu Landtagswahlen systematisch schlechtere Voraussagen abgegeben werden als zu Bundestagswahlen. Dies analysiert er mit Hilfe einer einfaktoriellen Varianzanalyse, welche einen signifikanten Unterschied ausweist. Allerdings sind nur 3,6 Prozent der Varianz der mittleren absoluten Abweichungen auf die Art der Wahl zurückzuführen.

Für die USA liegen ausgehend von der Diskussion um die Gründe der „Fehlprognose“ der Demoskopien zur US-Präsidentenwahl von 1948 (vgl.

Abschnitt 2.1.1) einige Befunde zu Einflussfaktoren auf die Genauigkeit von Voraussagen auf Grundlage von Wahlabsichtsumfragen vor.

Das zur Ursachenforschung des schlechten Abschneidens der Meinungsforscher im Rahmen der US-Präsidentschaftswahl 1948 eingesetzte Komitee des *Social Science Research Councils* (SSRC) stellt die erste Analyse der Voraussagequalität von Wahlabsichtsumfragen dar. Als Gründe für die Abweichung der Vorwahlumfragen vom späteren Wahlergebnis wurden verschiedene Aspekte diskutiert. Ein zentraler und bis heute nicht geklärter Streitpunkt stellt der Einfluss des verwendeten Stichprobenverfahrens dar. Der Bericht enthält hierzu keine präzise Aussage. Einerseits werden methodischen Aspekten, wie dem Stichprobenverfahren und dem Befragungsinstrument, eine zentrale Bedeutung zugesprochen, andererseits wird als Erkenntnis des Berichts angegeben, dass die „over-all operation of making election predictions from pre-election polls is a complex one“ (Moeller et al. 1949: 201). Wenig verwunderlich, dass diese Analyse Ausgangspunkt von Kontroversen sowohl unter Meinungsforschern als auch Wissenschaftlern war.

Bradburn und Sudman (1988: 29) sehen demnach inadäquate Quotenstichproben ebenso als Hauptursache ungenauer *pre-election polls* an wie Scheuch (1956: 53) und verweisen hierbei auf die von Mosteller et al. (1949: 95) berichtete systematische Verzerrung der Quotenstichproben. So zeigte sich etwa, dass die Gruppe der niedrig Gebildeten deutlich unterrepräsentiert war (Katz 1949). Ba-cheleder, Crossley und Gallup entgegneten, dass der Einfluss des Stichprobenverfahrens nicht entscheidend gewesen sein kann, da hinsichtlich der Abweichungen der Prognosen vom Wahlergebnis dem SSRC-Bericht zufolge Voraussagen auf Basis von Quotenstichproben nicht systematisch schlechter abschneiden (Stuit et al. 1949).

Zumindest ist für die Wahl von 1948 unumstritten, dass sich ein beträchtlicher Teil der Wählerschaft zu einem sehr späten Zeitpunkt entschieden hat oder sogar in den letzten Tagen vor der Wahl unentschieden hat. Mosteller et al. (1949: 251ff.) sprechen in diesem Zusammenhang vom Phänomen des „last minute swings“. Dies stellte ein erhebliches Problem dar, da die letzten Umfragen mindestens zwei Wochen vor der Wahl durchgeführt wurden und damit dieser Umschwung nicht erfasst werden konnte (siehe hierzu auch Moeller et al. 1949; Noelle-Neumann/Petersen 2005: 296). Crossleys Einschätzung, dass vor diesem

Hintergrund die Meinungsforscher selbst im Rahmen der US-Präsidentschaftswahl 1948 sehr gute Prognosen lieferten, wird jedoch nicht geteilt (Stuit et al. 1949).

Weiterhin zeigt der Bericht, dass es nicht gelungen ist, das Wahlverhalten der zum Zeitpunkt der Befragung Unentschiedenen vorherzusagen (Mosteller et al. 1949: 263ff.). Dies steht in engem Zusammenhang mit der Wahlbeteiligung, welcher ebenso ein bedeutender Einfluss in der bereits erörterten Art und Weise zugeschrieben wurde (Stuit et al. 1949). Insgesamt liefert der Bericht zwar Hinweise auf potenzielle Einflussfaktoren und auch einige empirische Befunde, jedoch stützen sich die Analysen ausschließlich auf die Umstände der Wahl von 1948, was eine systematische Untersuchung von Einflussfaktoren nicht ermöglicht.

Crespi (1988) hat diese Lücke für die USA geschlossen. Seine Untersuchung verfolgt die Ziele: „... to (1) assess the extent of poll accuracy (and inaccuracy); (2) to identify the correlates of accuracy; (3) to investigate the capabilities and limitations of current state-of-the-art polling; and (4) to assess the reasons for existing limitations“ (ebd.: 11). Um diesen Ansprüchen gerecht zu werden, hat er 430 Umfrageergebnisse für US-Wahlen zusammengetragen, hierbei allerdings entgegen der bisherigen Studien versucht, möglichst alle letzten Wahlabsichtsumfragen vor einer Wahl, die zwischen 1979 und Mitte der 1980er Jahre veröffentlicht wurden, zu erfassen. Ergänzend dazu hat er eine Umfrage unter den Meinungsforschungsinstituten durchgeführt, um weitere Informationen wie die Stichprobengröße, den Erhebungszeitraum und das Erhebungsverfahren zu ergänzen (vgl. ebd.: 19f.).

In einer multivariaten Betrachtung zeigt Crespi (ebd.: 163ff.), dass der zeitliche Abstand einer Umfrage bis zum Wahltermin in Bezug auf die getesteten Faktoren am einflussreichsten auf die Güte der Prognose ist. Weiterhin als relevant haben sich die Höhe der Wahlbeteiligung, die Knappheit des Wahlausgangs und der Anteil Unentschlossener in der Umfrage gezeigt – jeweils in der erwarteten Richtung. Ein weiterer wesentlicher Befund von Crespis Analyse ist, dass kein Effekt der Stichprobengröße auf die Güte der Voraussage gefunden werden konnte. Gerade dieses überraschende Ergebnis hat weitere Forschungen stimuliert.

So untersucht Lau (1994) methodische Einflussfaktoren auf die Voraussagegüte von Wahlabsichtsumfragen im Vorfeld der US-Präsidentschaftswahl 1992. Dazu analysiert er 56 Umfragen, welche im letzten Monat vor dem Wahltermin veröffentlicht wurden. Als Resultat dieser Studie ist zunächst festzuhalten, dass sich die Länge der Feldphase als wesentlich herausgestellt hat. Demnach verbes-

sert sich die Genauigkeit der Voraussage mit zunehmender Anzahl an Tagen an denen Interviews geführt werden und auch Umfragen, welche nicht ausschließlich an Wochentagen durchgeführt wurden, schnitten signifikant besser ab. Interessant ist ebenfalls, dass Umfragen basierend auf so genannten *tracking polls* signifikant geringere Voraussagefehler aufweisen als „normale“ Surveys.

Tracking polls sind dadurch gekennzeichnet, dass täglich Interviews auf Grundlage kleiner Zufallsstichproben geführt werden und die berichteten Ergebnisse auf Zusammenfassung beispielsweise einer gesamten Woche beruhen (für ein Beispiel vgl. Merkle et al. 2008). Als Vorteil erwartet man sich hier, dass durch das fortlaufende Monitoring etwaige kurzfristige Stimmungswchsel erfasst und durch die gleitende Durchschnittsbildung Fehler besser ausgeglichen werden können als bei etwa nur jeden Monat durchgeführten Umfragen, welche eine kurze Feldphase mit vergleichsweise hohen Fallzahlen aufweisen.

Analog zu Crespi (1988) konnte Lau (1994) keinen Einfluss der Stichprobengröße belegen. Im Rahmen des kurzen Zeitfensters von einem Monat konnte auch kein signifikanter Einfluss des zeitlichen Abstands der Umfrage bis zum Wahltermin gefunden werden.

An Laus (1994) Studie schließen DeSart und Holbrook (2003) an. Sie erweiterten jedoch die Datenbasis in zweierlei Hinsicht: Es werden die US-Präsidentenwahlen 1992, 1996 und 2000 betrachtet und es werden auch Umfragen berücksichtigt, die früher als ein Monat vor dem Wahltermin veröffentlicht wurden. Ihre Analysen zeigen einen deutlichen Effekt des zeitlichen Abstands – und zwar in der in Abschnitt 3.1 postulierten Art und Weise eines kurvilinearen Effekts. Eine Ausnahme hierzu bildet die Wahl von 2000, für welche kein signifikanter Einfluss des zeitlichen Abstands zur Wahl gefunden wurde. Entgegen der Befunde von Lau (1994) können DeSart und Holbrook (2003) keinen Unterschied hinsichtlich des Voraussagefehlers bei Umfragen, welche nur wochentags und solchen, die auch das Wochenende als Befragungstag umfassen, feststellen. Weitere Differenzen zu vorherigen Befunden ergeben sich im Hinblick auf den Einfluss der Stichprobengröße: Hier finden die Autoren einen signifikanten Effekt in der erwarteten Richtung, wonach die Voraussagegüte mit zunehmender Stichprobengröße steigt.

Insgesamt ist aus diesen Studien für die USA der Schluss zu ziehen, dass für die hier abgeleiteten Einflussfaktoren auf die Voraussagegenauigkeit bereits empirische Hinweise auf deren Relevanz vorliegen. Gleichwohl sind diese – etwa

in Bezug auf den Stichprobenumfang – nicht konsistent. Pickup und Johnston (2008) verweisen zudem darauf, dass in diesen Studien teilweise erhebliche methodische Defizite festzustellen sind. Dies betrifft einerseits die Datenbasis, welche verbreitert werden sollte. Andererseits müssten verstärkt Verfahren der Mehrebenen- und Zeitreihenanalyse eingesetzt werden. Dieser Aspekt betrifft vor allem die Analyse systematischer Verzerrung aufgrund von *house effects*. Für Deutschland gilt dies, wie gezeigt wurde, umso mehr, weshalb die folgenden empirischen Analysen sich auf eine bisher beispiellos breite Datenbasis stützen sowie angemessene statistische Auswertungsverfahren zur Anwendung kommen.

3.3 Methodik

3.3.1 Daten

Die folgenden empirischen Analysen verfolgen drei Zielsetzungen: Erstens wird anhand der Politbarometer-Studien der Forschungsgruppe Wahlen das Ausmaß der politischen Gewichtung im Rahmen der Erstellung von Stimmenanteils Voraussagen auf Grundlage von Wahlabsichtsfragen untersucht. Zweitens wird die Güte der umfragebasierten Voraussagen im Zeitablauf deskriptiv dargestellt. Dies soll Klarheit über die tatsächliche Stärke oder Schwäche demoskopischer Wahlprognosen liefern. Zugleich wird hierbei die Vergleichsgröße ermittelt, an denen sich zukünftige wissenschaftliche Prognosemodelle, wie das in Abschnitt 4 entwickelte, messen müssen. Drittens werden die zuvor in Abschnitt 3.1 abgeleiteten Hypothesen empirisch getestet, um Einflussfaktoren auf die Genauigkeit von Sonntagsfragen zu identifizieren.

Zentrale Datenbasis der folgenden Analyse stellen 3.610 publizierte Sonntagsfragenergebnisse zu den Bundestagswahlen von 1949 bis 2009 dar. Hierbei sind allerdings sowohl Umfrageergebnisse enthalten, welche sich nach 1990 auf Gesamtdeutschland oder ausschließlich auf die alten oder die neuen Bundesländer beziehen. 540 Sonntagsfragen (15,0 Prozent) konnten für die Wahlen vor der Wiedervereinigung gesammelt werden. Die verbleibenden Voraussagen beziehen sich zumeist auf das gesamte Bundesgebiet (1.947 bzw. 63,4 Prozent). Der Datensatz umfasst somit für das Bundesgebiet 2.487 Wahlvoraussagen.

Diese Daten wurden aus unterschiedlichen Quellen zusammengetragen, da der Versuch, die veröffentlichten Umfrageergebnisse direkt über die jeweiligen Umfrageinstitute zu beziehen, weitgehend gescheitert ist. Einzig die Forschungsgruppe Wahlen stellte eine Zeitreihe ihrer Projektionen zur Verfügung. Auch andere, ähnlich gelagerte Versuche entsprechende Daten direkt über die Meinungsforschungsinstitute zu beziehen, sind gescheitert (Antholz 2001: 26ff.; Ehrenpfordt/Maaßen 2007: 6f.).

Es war somit leider nicht möglich, gesichert eine vollständige Liste aller, oder zumindest die von den großen Demoskopieinstituten veröffentlichten Sonntagsfragenergebnisse zu erhalten. Antholz (2001: 26ff.) hat für die Jahre 1949 bis 2000 aufwendige Medienrecherchen durchgeführt, um einen Datensatz mit möglichst allen der letzten zwei vor eine Wahl publizierten Umfrageergebnisse für Bundes- und Landtagswahlen zu erstellen. Auf diese Datenquelle konnte im Rahmen dieser Arbeit zurückgegriffen werden.

Insgesamt umfasst seine „Wahlprognosetabelle“ (ebd.: 38) 454 auswertbare Ergebnisse von Sonntagsfragen für 82 Prozent aller Land- und Bundestagswahlen von 1949 bis 2000. Allerdings zeigt sich, dass die Daten von Antholz durchaus Probleme aufweisen. So enthielt der Datensatz Duplikate und zum anderen musste festgestellt werden, dass insbesondere für die 1990er Jahre ein erheblicher Teil publizierter Sonntagsfragen nicht erfasst wurde. In dieser Arbeit soll auch der Einfluss des zeitlichen Abstands zur Wahl überprüft werden. Auf Grundlage der Datenbasis von Antholz ist dies nur eingeschränkt möglich, da hier nur die jeweils zwei letzten Sonntagsfragenergebnisse erfasst wurden.

Dementsprechend wurde zunächst eine Bereinigung vorgenommen, um etwa Duplikate und andere offenkundig problematische Fälle auszuschließen. Unter letztere Gruppe fallen beispielsweise Voraussagen des Stimmenanteils von nur einer Partei oder solche, die nach Angaben von Antholz nicht auf Umfragen beruhen, sondern von Experten oder Journalisten getätigt wurden. Um die Lücken im Datensatz zu beseitigen und auch länger vor einem Wahltermin publizierte Sonntagsfragen zu ergänzen, wurde auf verschiedene Quellen zurückgegriffen.

Für die jüngeren Wahlen – etwa seit 1994, wobei dies nach Institut variiert – konnten zahlreiche Sonntagsfragenergebnisse von der von Martin Fehndrich betriebenen Webseite www.wahlrecht.de heruntergeladen werden. Selbst die angeschriebenen Demoskopieinstitute verwiesen zum Teil auf diese Datenquelle. Die Angaben auf www.wahlrecht.de wurden stichprobenartig überprüft, indem

entsprechende Zeitungsrecherchen unternommen wurden, um die erfassten mit den dort veröffentlichten Resultaten zu vergleichen. Dabei wurden die zu überprüfenden Resultate per Zufall ausgewählt und über alle Institute gestreut – die entsprechenden Recherchen ergaben in keinem Fall Abweichungen zu den von der Homepage wahlrecht.de entnommenen Werten.

Die verbliebenen Lücken im Datensatz, insbesondere in Bezug auf die 1980er und frühen 1990er Jahre, wurden durch eigene Zeitungsrecherchen geschlossen. Hierbei wurde gezielt auf die Kooperationspartner der jeweiligen Institute zurückgegriffen und so wurden etwa aus *Die Woche* insgesamt weitere 189, zuvor nicht im Datensatz enthaltene Voraussagen von Forsa entnommen.

Schließlich konnte auf zwei weitere Datenquellen zurückgegriffen werden. Zum einen stellte Gebhard Kirchgässner eine von ihm zusammengetragene Zeitreihe mit Sonntagsfragen von 1950 bis 1985 des IfD Allensbach über den Mailverteiler des Arbeitskreises Wahl- und Einstellungsforschung der Deutschen Vereinigung für Politische Wissenschaft (DVPW) zur Verfügung. Zum anderen konnte auf einen Datensatz von Christopher J. Anderson mit entsprechenden Informationen aus den Jahren 1960 bis 1989 zurückgegriffen werden.²⁹

Insgesamt hatte die Datensammlung das Ziel einen möglichst vollständigen Bestand an veröffentlichten Sonntagsfragen für den Zeitraum von 1949 bis 2009 zu erheben. Aufgrund fehlender Informationen ist es nicht möglich, abzuschätzen, inwieweit dies gelungen ist. Durch das Heranziehen unterschiedlicher Quellen konnten zahlreiche Redundanzen entdeckt werden (etwa in Bezug auf die Datenquellen von Anderson, Antholz und Kirchgässner). Zudem konnte für den jüngeren Zeitraum auf die äußerst umfassende und tagesaktuelle Sammlung von Umfrageergebnissen bei wahlrecht.de zurückgegriffen werden. Aus diesen Gründen wird vermutet, dass insgesamt ein hoher Abdeckungsgrad erreicht werden konnte.

Zusammen mit den jeweiligen vorausgesagten Stimmenanteilen für die einzelnen Parteien wurden, sofern möglich, die interessierenden Einflussfaktoren erhoben. Dies betrifft zunächst das Institut, welches die Umfrage durchgeführt hat. Im Datensatz sind Umfragen des IfD Allensbach (23,0 Prozent der Umfragen,

29 Der Datensatz kann über die Homepage von Christopher J. Anderson bezogen werden: http://dvn.iq.harvard.edu/dvn/dv/anderson/faces/study/StudyPage.xhtml?studyId=661&studyListingIndex=1_34cfe269babe4d2db75b90a3529c (zuletzt abgerufen am 7. Januar 2010).

welche sich nach 1990 nicht ausschließlich auf die neuen oder alten Bundesländer beziehen), von Infratest dimap (17,2 Prozent), TNS Emnid (18,9 Prozent), Forsa (25,7 Prozent), der Forschungsgruppe Wahlen (11,3 Prozent), GMS Dr. Jung GmbH (2,1 Prozent), Infas (0,4 Prozent) und sonstigen, nicht näher differenzierten Instituten (1,4 Prozent). Insbesondere die geringe Anzahl von Infas-Voraussagen verwundert und deutet auf eine möglicherweise unzureichende Erfassung entsprechender Umfragen hin. Dies ist allerdings vor dem Hintergrund zu sehen, dass Infas heute keine politischen Meinungsumfragen mehr durchführt und bereits Mitte der 1990er Jahre durch die Übergabe der ARD-Fernsehberichterstattung an Infratest dimap an Bedeutung für Wahlvoraussagen verlor. Entsprechend ist die jüngste im Datensatz enthaltene Infas-Sonntagsfrage zur Bundestagswahl von 1998 und seitdem hat die Zahl der veröffentlichten Umfragen generell sprunghaft zugenommen, während in den 1960er und 1970er Jahren häufig nur wenige Sonntagsfragen vor einer Wahl publiziert wurden. Jedoch ist anzunehmen, dass insgesamt der Abdeckungsgrad mit dem jeweiligen Zeitraum variiert. Für ältere Jahrgänge gestaltet sich die Recherche deutlich schwieriger (Antholz 2001: 35ff.), etwa weil die Prominenz von Meinungsumfragen in der journalistischen Berichterstattung im Zeitablauf zugenommen hat (vgl. z. B. Bretschneider 1996). Einzig für das IfD Allensbach besteht dieses Problem weniger, da die Umfragen des Instituts in den Allensbacher Jahrbüchern dokumentiert sind.

Erhoben wurden auch jeweils das Veröffentlichungsdatum der Voraussage sowie die Anzahl der in der Voraussage berücksichtigten Parteien. Herangezogen wurde hier jeweils der Termin an welchem die Voraussage in einer Zeitung oder im Fernsehen erstmals veröffentlicht wurde. Durch die Erfassung der nächst gelegenen Wahltermine konnten mit diesen Angaben (zumeist) tagesgenau der zeitliche Abstand der Voraussage zur Wahl berechnet werden.

Soweit verfügbar wurden methodische Angaben zur Umfrage gesammelt. Dies betrifft die Zahl der realisierten Interviews – Angaben von Bruttostichproben oder Ausschöpfungsquoten werden nicht gemacht – und dem Erhebungszeitraum. Beide Angaben fehlen jedoch häufig. In Bezug auf die Stichprobengröße konnten nur für 42,4 Prozent der erfassten Voraussagen Informationen zur Stichprobe erhoben werden. Aufgenommen wurden entsprechende Angaben nur bei expliziter Nennung; einige Institute bieten weiterführende Informationen auf ihrer Homepage an (so etwa Infratest dimap), der man allgemeine Hinweise entnehmen kann, wie etwa zu Fehlerintervallen der Sonntagsfragenergebnisse. Vage

Informationen dieser Art, welche nicht in unmittelbarem Zusammenhang mit der Publikation der jeweiligen Umfrage stehen, wurden nicht berücksichtigt. Deren Aussagekraft wäre aufgrund fehlender Varianz zudem gering, da beispielsweise die meisten Institute angeben, in etwa 1.000 Personen zu befragen. Die Folge dieser unzulänglichen Dokumentation methodischer Aspekte der Umfragen ist, dass zwar die formulierten Hypothesen zu Einflussfaktoren auf die Voraussagegenauigkeit getestet werden können, dies jedoch mit einem selektiven Sample aus allen erfassten Sonntagsfragen geschieht, von welchen angenommen wird, dass es zumindest eine gute Approximation der tatsächlichen Sonntagsfragenlage für den Zeitraum von 1949 bis 2009 darstellt.

Schließlich wurde die Wahlbeteiligung erfasst. Diese Daten wurden der Homepage des Bundeswahlleiters entnommen und zusammen mit den Daten der Bundestagswahlen sowie den tatsächlichen Zweitstimmenanteilen der einzelnen Parteien dem Datensatz zugespielt.

Für den Vergleich mit den Rohdaten zur Darstellung des Ausmaßes der politischen Gewichtung werden die Politbarometer-Daten herangezogen, welche über das GESIS – Leibniz-Institut für Sozialwissenschaften für wissenschaftliche Zwecke bezogen werden können.³⁰ In allen Wellen wurden jeweils die Stimmenanteile für CDU/CSU, SPD, FDP, Bündnis 90/Die Grünen sowie PDS/Die Linke anhand der immer gestellten Sonntagsfrage ermittelt. Bei der Berechnung der Stimmenanteile wurde das Designgewicht berücksichtigt, jedoch wurden die zum jeweiligen Befragungszeitpunkt Unentschiedenen und potenziellen Nichtwähler bei der Berechnung ausgeschlossen. Befragte, die angegeben haben eine andere Partei zu wählen als die in den folgenden Abbildungen enthaltenen, wurden in einer Kategorie „Sonstige“ zusammengefasst, welche jedoch nicht dargestellt wird. Diese Kategorie umfasst je nach Fragebogendesign unterschiedlich viele und auch im Zeitablauf wechselnde Parteien.

30 Die Daten für die Studie Politbarometer wurden erhoben von der Forschungsgruppe Wahlen. Sie wurden von der GESIS für die Analyse aufbereitet und dokumentiert. Weder die Forschungsgruppe Wahlen noch das GESIS tragen irgendeine Verantwortung für die Analyse oder Interpretation der Daten in dieser Arbeit. Verwendet wurden die Studien mit folgenden ZANummern: 1987, 1920, 2102, 2114, 2275, 2287, 2378, 2390, 2391, 2546, 2559, 2765, 2777, 2894, 3045, 3160, 3261, 3262, 3425, 3426, 3554, 3555, 3849, 3850, 4001, 4002, 4221, 4222, 4256, 4257.

Zur Ermittlung der gesamtdeutschen Stimmenanteile musste für die Jahre 1990 bis 1995 sowie 2000 bis 2003 ein zusätzlicher Berechnungsschritt eingeführt werden. Die Forschungsgruppe Wahlen hat in diesen Jahren getrennte Erhebungen in den alten und neuen Bundesländern durchgeführt mit einem sich auch teilweise unterscheidenden Fragenkatalog. Ein Zusammenführen der Datensätze war somit nicht möglich. In diesen Fällen wurden die Stimmenanteile der einzelnen Parteien getrennt für die neuen und alten Bundesländer berechnet und anschließend noch nach Bevölkerungsanteilen gewichtet zusammengeführt. Die jeweiligen Bevölkerungsanteile wurden entsprechenden Bevölkerungsstatistiken des Statistischen Bundesamtes für das jeweilige Jahr zum Stichtag 31.12. entnommen.

Schließlich ist zu erwähnen, dass das Politbarometer in der Regel nur elfmal im Jahr durchgeführt wird – üblicherweise finden entweder im Juli oder August keine Erhebungen statt. Zur Vervollständigung der Reihe wurden die jeweils fehlenden Datenpunkte durch den Mittelwert des vorherigen und des nachfolgenden Monats ersetzt. Da entsprechend auch die Zeitreihe der Projektionen Lücken aufweist, wurde diese nach demselben Verfahren für die Analysen in Abschnitt 3.4.1 imputiert.

3.3.2 *Maßzahlen zur Bestimmung des Prognosefehlers*

Im Hinblick auf die Evaluation der Voraussagegüte von Sonntagsfragen ist zunächst zu klären, wie die Abweichung der Vorhersage vom tatsächlichen Wahlergebnis gemessen werden kann. Ziel dabei ist es, ein Maß zu konstruieren, das möglichst objektiv und valide ist, jedoch zugleich eine möglichst intuitive Interpretation erlaubt und zudem den Vergleich verschiedener Prognosearten miteinander erlaubt (vgl. Barrot 2007: 417).

Eine Problematik bei der Erfassung des Fehlers ergibt sich im vorliegenden Fall daraus, dass immer mehrere Parteien zur Wahl stehen und die Voraussagen entsprechend eine – zumindest über die Zeit – variierende Anzahl von einzelnen zu prognostizierenden Stimmenanteilen umfassen. Dennoch soll eine Maßzahl gefunden werden, welche diese Informationen sinnvoll zusammenfasst. Hierzu liegen verschiedene Vorschläge vor, von denen im Folgenden die zentralen Konzepte knapp dargestellt und bezüglich ihrer Brauchbarkeit für die hier angestrebten

Analysen bewertet werden (für einen umfassenderen Überblick und eingehende Diskussionen verschiedener Maßzahlen vgl. etwa Armstrong/Collopy 1992; Barrot 2007; Flores/Wichern 2005; Hyndman/Koehler 2006).

Nach der „Fehlprognose“ von 1948 im Vorfeld der US-Präsidentenwahl wurde ein Komitee zur Erforschung der Gründe für dieses „Versagen“ demoskopischer Instrumente und zur Entwicklung möglicher Maßnahmen zur Sicherstellung möglichst zutreffender Voraussagen. Im Rahmen dieser Evaluation schlugen Mosteller et al. (1949: 55ff.) verschiedene Maßzahlen zur Bewertung der Prognosegüte von Wahlabsichtsumfragen vor, welche bis heute zentral für deren Analyse sind. Die acht vorgeschlagenen Maßzahlen beziehen sich unmittelbar auf Präsidentenwahlen und einige sind nur anwendbar, wenn sich die Prognose ausschließlich auf zwei Parteien oder Kandidaten bezieht. An dieser Stelle werden nur diejenigen Vorschläge aufgegriffen, welche prinzipiell auf Mehrparteiensysteme anwendbar sind.

3.3.2.1 Der mittlere absolute Fehler

Bekanntestes und das bis heute im Rahmen der empirischen Wahlforschung zumeist verwendete Maß ist die mittlere absolute Abweichung (*mean absolute error* – MAE). Hierzu werden für alle Parteien oder Kandidaten die absoluten Differenzen zwischen vorausgesagtem und tatsächlichem Stimmenanteil in Prozentpunkten betrachtet und ihr Mittelwert gebildet:

$$MAE = \frac{1}{k} \sum_{i=1}^k |\hat{x}_i - x_i|$$

Dabei und auch im Folgenden wird mit x_i der tatsächliche Stimmenanteil einer Partei i bezeichnet und \hat{x}_i repräsentiert den geschätzten Stimmenanteil einer Partei i . Weiterhin steht k für die Anzahl berücksichtigter Parteien.

Wesentliche Vorteile dieser Maßzahl sind ihre einfache Berechnung sowie die intuitive Interpretation als durchschnittliche Abweichung der Prognose vom tatsächlichen Ergebnis in Prozentpunkten. Gegenüber der einfachen mittleren Abweichung weist dieses Maß den Vorteil auf, dass negative und positive Abweichungen sich nicht gegenseitig aufheben und damit der Wert Null eindeutig als perfekte Prognose gewertet werden kann.

Allerdings weisen bereits Mosteller et al. (1949: 56) auf eine zentrale Schwäche des Maßes hin: Es ist abhängig von der Zahl der inkludierten Parteien. Beispielsweise wird bei erheblichen Abweichungen bezüglich der Stimmenanteile zweier großer Parteien bei gleichzeitiger Berücksichtigung vieler kleiner Parteien die mittlere absolute Abweichung gering, obwohl die vorrangig interessierenden Stimmenanteile der großen Parteien unzureichend vorausgesagt werden. Hieraus ergibt sich auch das Problem, dass der MAE nur bei konstanter Parteienzahl zu Vergleichszwecken angewendet werden darf. Da sich das Parteiensystem Deutschlands im Zeitverlauf geändert hat und neue Parteien (Bündnis 90/Die Grünen sowie die PDS/Die Linke) relevant geworden sind, ist die Anwendung der mittleren absoluten Abweichung nur eingeschränkt möglich – zumindest nicht über den gesamten Zeitrahmen und alle erfassten Voraussagen hinweg.

Diese Sensibilität hinsichtlich der berücksichtigten Parteienzahl erfordert mithin eine Entscheidung des Forschers darüber, wie viele Parteien berücksichtigt werden sollen (vgl. hierzu Mitofsky 1998: 235). Im Rahmen dieser Analyse ist diese Entscheidung relativ leicht, denn es werden immer diejenigen Parteien berücksichtigt, für die Stimmenanteile vorausgesagt werden. Verglichen werden können mittels des MAE dann nur Voraussagen, welche auf der gleichen Anzahl berücksichtigter Parteien beruhen. Trotz der erwähnten Veränderungen des deutschen Parteiensystems wiegt dieses Problem nicht so schwer, denn die Anzahl vorausgesagter Parteien variiert nur zwischen mindestens zwei und höchstens sechs.³¹

Weiterhin wird an dieser Stelle das Ziel verfolgt, mögliche systematische Verzerrungen zugunsten oder zuungunsten einzelner Parteien zu untersuchen. Aufgrund der Betrachtung absoluter Abweichungen kann dieser potenzielle *bias* mit dieser Maßzahl nicht analysiert werden. Da die mittlere absolute Abweichung als das dominante Maß zur Bestimmung der Güte von Wahlprognosen angesehen wird und auch im Rahmen der spärlich vorliegenden empirischen Untersuchungen in Deutschland verwendet wurde, wird es auch in dieser Arbeit

31 Es liegen acht Voraussagen (0,3 Prozent) für zwei Parteien vor. Hierbei werden jeweils Stimmenanteile von CDU/CSU und SPD betrachtet. Diese Voraussagen beziehen sich auf Wahlen von 1976 und früher. Missachtet man spezifische Umfrageergebnisse für die alten und neuen Bundesländer liegen 355 (14,3 Prozent) Voraussagen für sechs Parteien vor. Dies sind Umfragen zu den Bundestagswahlen 1990, 1994, 1998 und 2002, welche zusätzlich Voraussagen der Stimmenanteile für die Republikaner umfassen.

herangezogen. Allerdings wird es ausschließlich im Rahmen der deskriptiven Auswertungen zu Vergleichszwecken mit anderen Befunden und aufgrund der intuitiven Interpretation verwendet. Zudem wird hierbei die zugrunde liegende Parteienanzahl zu berücksichtigen sein.

Als Erweiterungen dieser Maßzahl haben insbesondere die mittlere quadratische Abweichung (*mean square error* – MSE) sowie die Wurzel der mittleren quadratischen Abweichung (*root mean square error* – RSME) an Bedeutung gewonnen (Hyndman/Koehler 2006).

$$MSE = \frac{1}{k} \sum_{i=1}^k (\hat{x}_i - x_i)^2$$

$$RMSE = \sqrt{MSE}$$

Diese Maßzahlen tragen der Überlegung Rechnung, dass größere Abweichungen stärker berücksichtigt werden sollten als kleinere. Das grundsätzliche Defizit der Abhängigkeit von der Anzahl der berücksichtigten Parteien wird hier jedoch nicht gelöst. Zudem reagieren beide Maße im Gegensatz zur mittleren absoluten Abweichung sensibel auf Ausreißer. Aus diesen Gründen werden sie im Weiteren nicht herangezogen.

3.3.2.2 Der mittlere absolute prozentuale Fehler

Als skalenunabhängige Maßzahlen sind insbesondere solche gebräuchlich, die auf Prozentuierung beruhen (Hyndman/Koehler 2006: 683). Das am weitesten verbreitete Maß innerhalb dieser Gruppe ist der so genannte mittlere absolute prozentuale Fehler (*mean absolute percentage error* – MAPE). Er gibt die durchschnittliche absolute prozentuale Abweichung des prognostizierten vom tatsächlichen Wahlergebnis an (Barrot 2007: 419):

$$MAPE = \frac{1}{k} \sum_{i=1}^k \frac{|\hat{x}_i - x_i|}{x_i} \cdot 100$$

Der Vorteil dieses Maßes ist, dass es dimensionslos und unabhängig von der Größenordnung der betrachteten Werte ist. Weiterhin ist die intuitive Interpretation

hervorzuheben. Die Maßzahl weist mit ihrer klaren Interpretierbarkeit des Wertes Null als perfekte Prognose durch den Ausschluss der Möglichkeit der gegenseitigen Aufhebung von positiven und negativen Abweichungen eine weitere Parallelität zum MAE auf. Gleichwohl ergeben sich auch Nachteile: Die Berücksichtigung der Abweichung in Relation zur Höhe des Stimmenanteils führt dazu, dass Abweichungen bei Parteien mit kleinen Stimmenanteilen überbewertet werden. Dieses Problem ist folglich analog zu den beschriebenen Schwierigkeiten der mittleren absoluten Abweichung: Es besteht eine Abhängigkeit von der berücksichtigten Anzahl der Parteien. Im Fall des MAPE wird jedoch mit steigender Parteienzahl der Fehler größer und nicht kleiner wie beim mittleren absoluten Fehler.

Gerade im deutschen Parteiensystem kann diese „Bestrafung“ von Fehleinschätzungen des Stimmenanteils kleiner Parteien erwünscht sein, weshalb im Rahmen der deskriptiven Analyse auf den relativen absoluten Fehler zurückgegriffen wird. Durch die Fünf-Prozent-Hürde ist gerade für die kleinen Parteien eine möglichst exakte Prognose besonders relevant, denn das Überspringen der Fünf-Prozent-Marke entscheidet, ob sie im künftigen Parlament vertreten sein werden oder nicht. Wichtig ist an dieser Stelle zu betonen, dass in dieser Arbeit ausschließlich auf die Zweitstimme Bezug genommen wird. Neben dem Erreichen eines Stimmenanteils von fünf Prozent kann eine Partei auch bei Gewinn von drei oder mehr Direktmandaten in den Bundestag einziehen (Grundmandatsklausel). Sie erhält dann dennoch eine ihrem bundesweiten Stimmenanteil entsprechende Sitzanzahl im Parlament. Dieser Fall ist bereits eingetreten. So gewannen vier Kandidaten der PDS bei der Bundestagswahl 1994 Direktmandate, die Partei erreichte bundesweit jedoch nur 4,4 Prozent der abgegebenen Stimmen. Die Partei erhielt in der Folge 30 Sitze im 13. Deutschen Bundestag.

3.3.2.3 Der Vergleichsquotient v

Eine alternative Betrachtung der Prognosegüte ist möglich, wenn man die Verteilungen der Vorhersagen mit den Verteilungen der tatsächlichen Stimmenanteile miteinander vergleicht. Hierfür kann man den Herfindahl-Index h als Ausgangs-

punkt nehmen (Herfindahl 1950).³² Zur Berechnung des Indexes werden die quadrierten Anteile addiert. Für die tatsächlichen Stimmenanteile x_i ergibt sich für n Parteien damit:

$$h = \sum_{i=1}^n x_i^2, \quad \text{wobei } \frac{1}{n} \leq h \leq 1$$

Die Untergrenze des Herfindahl-Indexes wird bei der Gleichverteilung unter allen n Parteien erreicht. Der Höchstwert ergibt sich bei $n = 1$, also bei Vorliegen eines Monopols. Analog hierzu ist der Herfindahl-Index h^* für die vorhergesagten Stimmenanteile zu berechnen:

$$h^* = \sum_{i=1}^n x_i^{*2}, \quad \text{wobei } \frac{1}{n} \leq h^* \leq 1$$

Bezieht man diese beiden Herfindahl-Indizes nun aufeinander zum Vergleich der beiden Verteilungen, so kann man Braun und Keuschnigg (2008) folgend einen logarithmierten Vergleichsquotienten bilden:

$$v := \ln\left(\frac{h}{h^*}\right), \quad \text{wobei } -\ln(n) \leq v < \ln(n)$$

Das Logarithmieren der Maßzahl bedingt Symmetrie, so dass der Wert Null eine Gleichverteilung von h und h^* angibt. Je weiter der Wert v von Null entfernt ist, desto größere Unterschiede bestehen zwischen den Verteilungen. Geprüft wird im Weiteren jeweils, wie hoch der Anteil der Vorhersagen ist, die innerhalb von \pm einer beziehungsweise zwei Standardabweichungen liegen.

Für die deskriptive Betrachtung der Prognosegüte wird auch auf dieses alternative Maß zurückgegriffen, da es einer anderen Logik als die bisher diskutierten Maßzahlen folgt und sich auf die Erfassung der Konzentration der Stimmenanteile auf einzelne Parteien bezieht.

Allerdings weist auch dieses Maß für die hier verfolgten Ziele Defizite auf. So ist der Herfindahl-Index ebenfalls von der Fallzahl abhängig, weshalb ähnlich wie bei den bisher erörterten Maßzahlen ein Vergleich über die Zeit aufgrund

32 Rinne (2008) gibt einen Überblick über alternative Konzentrationsmaße.

variierender Parteienanzahl nur eingeschränkt möglich ist. Darüber hinaus ist es mit dieser Maßzahl nicht möglich, die parteispezifische Prognosegüte zu analysieren. Lediglich wahl- beziehungsweise institutsspezifische Analysen über alle Parteien sind auf Grundlage der Maßzahl v möglich. Als Ungenauigkeit kommt hinzu, dass die kleinen Parteien zur Kategorie „Sonstige“ zusammengefasst und als eine Partei betrachtet werden müssen, da zu diesen Gruppierungen keine Vorhersagen vorliegen.

3.3.2.4 Das A-Maß

Wesentlicher Mangel der bisherigen Maße ist, dass mögliche systematische Unter- oder Überschätzungen von Stimmenanteilen bestimmter Parteien nicht erfasst werden können und zudem die Abhängigkeit von der Zahl der berücksichtigten Parteien bisher nicht gelöst ist. Martin et al. (2005) haben ein Maß entwickelt, welches – zumindest bei Berücksichtigung von nur zwei Parteien – die Größe des Fehlers sowie dessen Richtung misst und dabei unabhängig ist von variierenden Faktoren (etwa der Anteil sonstiger Parteien).

Die vorgeschlagene Maßzahl A beruht auf relativen Chancen (*odds ratios*). Zunächst wird der Fall mit zwei Parteien a und b betrachtet:

$$A_a = \ln \left[\frac{\hat{x}_a / \hat{x}_b}{x_a / x_b} \right]$$

Im Nenner des Terms innerhalb der Klammer steht das Verhältnis der vorausgesagten Stimmenanteile für Partei a und b . Ergibt sich hier ein Wert größer eins, sieht die Prognose einen Vorsprung für Partei a , bei Werten kleiner eins entsprechend einen Vorsprung für Partei b . Analog verhält es sich in der dargestellten Relation der tatsächlichen Stimmenanteile von Partei a und b , welche im Zähler des Terms innerhalb der Klammer steht. Durch die Bildung des Verhältnisses der *odds* für die Voraussage und das Wahlergebnis erhält man den Prognosefehler ohne die Information über die Richtung der Abweichung zu verlieren. Die Logarithmierung der *odds ratio* erfolgt, um ein symmetrisches Maß zu erhalten.

Nimmt A den Wert Null an, liegt eine perfekte Prognose vor, da beide Quotienten dieselbe relative Verteilung von Wähleranteilen zwischen Partei a und b aufweisen und die *odds ratio* damit gleich eins wird. Positive Werte können

dahingehend interpretiert werden, dass die Voraussage das Abschneiden von Partei a überschätzt, bei negativen Werten wurde der Stimmenanteil von Partei b überschätzt. Je größer die Differenz von A zu Null, desto größer ist die jeweilige Verzerrung und damit auch der Prognosefehler.

Durch Entlogarithmierung erhält man zudem eine inhaltlich bedeutsame Interpretation der Abweichung. Repräsentiert a die CDU/CSU und b die SPD und ist $A = 0,21$, bedeutet dies, dass die Voraussage den Stimmenanteil von CDU/CSU gegenüber ihrem tatsächlichen Ergebnis um den Faktor $e^{0,21} = 1,23$ überschätzt.

Wesentlicher Vorteil dieses Maßes ist, dass durch die Standardisierung Vergleiche über verschiedene Wahlen hinweg möglich sind, zumal die betrachteten relativen Anteile nicht auf Veränderungen der Gruppengröße reagieren. Entsprechend sehen auch Martin et al. (2005: 353) den Anwendungsfall für dieses Maß in Meta-Analysen, welche Einflussfaktoren auf Prognosefehler analysieren.

Zentrales Problem der Maßzahl A in der präsentierten Variante ist, dass es bei Wahlen mit bedeutenden Stimmenanteilen für weitere Parteien sehr klein werden kann, wie folgendes Beispiel zeigt. Am 5. September 1998 veröffentlichte Infratest dimap das Ergebnis einer Sonntagsfrage und ermittelte für die CDU/CSU einen Stimmenanteil von 37,0 Prozent und für die SPD 43,0 Prozent. Das Ergebnis der Wahl am 27. September 1998 ergab für die CDU/CSU einen Stimmenanteil von 35,2 Prozent, für die SPD 40,9 Prozent. Alle anderen Parteien wären nach der Sonntagsfrage zusammen auf 20,0 Prozent gekommen, tatsächlich erreichten sie 23,9 Prozent. Die jeweiligen Abweichungen für CDU/CSU und SPD betragen 1,8 beziehungsweise 2,1 Prozentpunkte. Berechnet man auf Grundlage dieser Daten nun das A -Maß, ergibt sich ein Wert nahe Null:

$$A_{CDU/CSU} = \ln \left[\frac{37,0 / 43,0}{35,2 / 40,9} \right] = -0,0002$$

Insofern ist dieses Maß nur sinnvoll anwendbar, wenn sonstige Parteien nur einen geringen Anteil an allen Stimmen aufweisen, oder man das Maß für die Anwendung auf mehrere Parteien erweitert. Letzteres ist im Kontext der Zielsetzung dieser Arbeit notwendig.

Martin et al. (2005) erweitern illustrativ ihr entwickeltes Maß für den Drei-Parteien-Fall. Da für den hier interessierenden Fall eine Erweiterung auf fünf Parteien zentral ist, zumal dies der Regelfall ist (64,1 Prozent der auf das gesamte Bundesgebiet bezogenen Umfragen), wird eine Weiterentwicklung des Maßes

vorgeschlagen. Die Ausweitung beruht weiterhin auf der Idee, dass das Abschneiden einer Partei relativ zum Abschneiden anderer Parteien gesehen werden soll. Folgt man Martin et al. (ebd.) ergäbe sich für den Fünf-Parteien-Fall folgendes A -Maß:

$$A_{a5} = \ln \left[\frac{\hat{x}_a / (\hat{x}_b + \hat{x}_c + \hat{x}_d + \hat{x}_e)}{x_a / (x_b + x_c + x_d + x_e)} \right]$$

Ein entscheidender Vorteil hiervon ist, dass das Maß bei jeweils korrekter Berücksichtigung der Parteienanzahl im Kontext der Berechnung über verschiedene Wahlen hinweg miteinander verglichen werden kann. Dies ist der Fall, denn es gilt:

$$Odds\ ratio_{Umfra\ge} = \left[\frac{\hat{x}_a / (\hat{x}_a + \hat{x}_b)}{\hat{x}_b / (\hat{x}_a + \hat{x}_b)} \right] = \left[\frac{\hat{x}_a / (\hat{x}_a + \hat{x}_b + \hat{x}_c + \hat{x}_d + \hat{x}_e)}{\hat{x}_b / (\hat{x}_a + \hat{x}_b + \hat{x}_c + \hat{x}_d + \hat{x}_e)} \right] = \frac{\hat{x}_a}{\hat{x}_b}$$

Das erweiterte A -Maß misst allerdings nur die Genauigkeit der Voraussage und die Richtung der Verzerrung in Bezug auf die jeweilige Referenzpartei a . Die Abweichungen der geschätzten Stimmenanteile für die anderen Parteien bleiben unberücksichtigt. Insofern eignet sich dieses Maß zur Überprüfung möglicher systematischer Verzerrungen einzelner Institute zugunsten oder zuungunsten bestimmter Parteien, nicht jedoch als generelles Maß für die Erfassung der Prognosegüte im Mehrparteienfall.

Hierzu wird im Folgenden für jede betrachtete Partei ein A -Maß berechnet und die nach Parteienanzahl gewichtete Summe als Maß für die generelle Prognosegüte herangezogen:

$$A_{adj} = \frac{1}{k} \sum_{i=1}^k |A_i|$$

Wesentliche Nachteile dieses adjustierten A -Maßes ist, dass der Parteibias nicht mehr erkennbar ist und dass eine intuitive Interpretation nicht mehr gegeben ist. Der Vorteil besteht darin, dass es auch bei variierender Parteienanzahl eingesetzt werden kann und entgegen dem einfachen A -Maß alle Abweichungen berücksichtigt werden. Dabei werden die Fehler der Voraussagen relativ betrachtet. Dies vermeidet die angesprochenen Probleme des MAE beziehungsweise des MAPE in dieser Hinsicht. Deshalb wird das adjustierte A -Maß für die Analyse

potenzieller Einflussfaktoren auf die Voraussagegüte als abhängige Variable herangezogen, nicht jedoch im deskriptiven Teil aufgegriffen. Hier kommt das von Martin et al. (2005) vorgeschlagene A-Maß zur Identifizierung von systematischen Über- oder Unterschätzungen bestimmter Parteistimmenanteile zum Einsatz.

3.3.3 Auswertungsverfahren

Im Rahmen der empirischen Analysen kommen univariate, bivariate sowie vor allem multivariate statistische Verfahren zum Einsatz. Letztere umfassen verschiedene regressionsanalytische Techniken, welche jeweils die zugrunde liegende Datenstruktur berücksichtigen. Die im Rahmen der Analyse des Ausmaßes der politischen Gewichtung (Abschnitt 3.4.1) herangezogenen Politbarometer-Daten weisen Zeitreihencharakter auf, weshalb hier knapp wesentliche Konzepte und Analyseschritte der univariaten Zeitreihenanalyse skizziert werden (Abschnitt 3.3.3.1). In Abschnitt 3.4.3 werden die Einflussfaktoren auf die Prognosegüte von Wahlabsichtsfragen untersucht. Hierbei wird vermutet, dass sich sowohl die Vorhersagen für die einzelnen Bundestagswahlen sowie die Vorhersagen der jeweiligen Institute einander ähnlicher sind als zwischen den Bundestagswahlen und den Instituten. Damit liegen hierarchisch gegliederte Daten vor, weshalb die Anwendung von Mehrebenenverfahren notwendig ist. Auch auf diese wird hier knapp eingegangen (Abschnitt 3.3.3.2).

Die Darstellung der univariaten Zeitreihenanalyse wird allgemein gehalten, während sich die Erörterung der Mehrebenenanalyse unmittelbar auf die Problemstellung der Arbeit bezieht. Diese Differenz rührt daher, dass unterschiedliche Zeitreihen vorliegen, welche jeweils angemessene Modelle zu ihrer Beschreibung bedingen. Im Rahmen der Analyse der Einflussfaktoren auf die Vorhersagegüte von Wahlabsichtsfragen liegt immer dieselbe Datenstruktur zugrunde, weshalb diese Modelle spezifischer behandelt werden können.

3.3.3.1 Zeitreihenanalyse

Von einer Zeitreihe wird gesprochen, wenn Beobachtungen vorliegen, welche einer zeitlich vorgegebenen definitiven Ordnung folgen (Kirchgässner/Wolters 2006: 1). Beispiele für Zeitreihen sind etwa die Entwicklung der Arbeitslosen-

quoten, die Schlusskurse des Deutschen Aktienindex oder auch die monatliche Zahl der Verkehrstoten. Charakteristisch für solche Daten ist, dass häufig die Beobachtungen voneinander abhängen, also beispielsweise die Zahl der Arbeitslosen in einem Monat wesentlich von der Zahl der Arbeitslosen im Monat zuvor abhängt.

Solche Regelmäßigkeiten in den Daten sollen mit den Verfahren der Zeitreihenanalyse aufgedeckt werden und dienen häufig Prognosezwecken. Im Rahmen der auch im Folgenden zur Anwendung kommenden univariaten Zeitreihenanalyse geht es darum, die interessierende Variable zu ihren Ausprägungen in der Vergangenheit in Verbindung zu setzen (Verbeek 2008: 269). Hierbei ist man insbesondere daran interessiert, etwaige Trends, saisonale und konjunkturelle Schwankungen zu identifizieren und von zufälligen Effekten zu trennen (Schlittgen/Streitberg 2001: 9). Es wird demnach versucht, einen komplexen Prozess aus mehreren einfachen Prozessen aufzubauen. Zur Prognose von Ereignissen werden die aufgedeckten Trends und/oder Saisoneffekte um die Zufallsschwankungen bereinigt in die Zukunft fortgeschrieben. Die Klasse der Modelle zur Analyse solcher Zeitreihenprozesse werden ARIMA-Modelle genannt (*autoregressive integrated moving average models*).

Im Rahmen der Zeitreihenanalyse wird angenommen, dass die Beobachtungen zu einem bestimmten Zeitpunkt t als die Realisation einer Zufallsvariablen Y_t aufgefasst werden, welche durch einen stochastischen Prozess beschrieben werden können. Dabei ist $t \in T$ und T eine geordnete und – in der Regel diskrete Indexmenge. Im vorliegenden Fall ist Y_t der Stimmenanteil einer spezifischen Partei in einem bestimmten Monat. Ziel ist es nun, diesen stochastischen Prozess mit einem möglichst sparsamen Modell zu beschreiben, das die zeitlichen Abhängigkeiten umfasst, um die dynamischen Komponenten der Zeitreihe offenzulegen.

Eine Möglichkeit solch einer zeitlichen Abhängigkeit ist, dass der Stimmenanteil zum Zeitpunkt t vom Stimmenanteil zum Zeitpunkt $t-1$ abhängt:

$$Y_t = \delta + \theta Y_{t-1} + \varepsilon_t$$

Für den Term ε_t wird dabei angenommen, dass dieser keine serielle Autokorrelation aufweist, homoskedastisch ist, einen Mittelwert von Null und eine konstante Varianz hat (Verbeek 2008: 270). Dieser Prozess wird als autoregressiver Prozess erster Ordnung beziehungsweise AR(1)-Prozess bezeichnet. Inhaltlich

besagt dies, dass der aktuelle Stimmenanteil einer Partei Y_t sich ergibt aus einer Konstanten δ plus θ mal den Stimmenanteil des Vormonats (Y_{t-1}) plus dem Residualterm ε_t . Entscheidend in Bezug auf ε_t ist, dass dieser die Zufallsschwankungen abbildet und demnach als zufälliger, nicht vorhersagbarer Prozess aufzufassen ist. Im Rahmen der Zeitreihenanalyse wird dies als *white-noise*-Prozess bezeichnet.

Durch autoregressive Prozesse höherer Ordnung können auch andere zeitliche Abhängigkeiten modelliert werden. Sind etwa konjunkturelle oder saisonale Schwankungen zu beobachten, kann dies durch die entsprechende Berücksichtigung von *lag*-Termen höherer Ordnung einbezogen werden. So ist beispielsweise bekannt, dass die monatliche Arbeitslosenrate saisonal schwankt und in den Wintermonaten regelmäßig höher ist als in den Sommermonaten. Betrachtet man nun die beispielsweise Arbeitslosenrate im Januar eines Jahres, so ist zu vermuten, dass diese nicht nur von der Arbeitslosenrate des Vormonats abhängt, sondern auch von der des Monats Januar vor einem Jahr. Durch eine Erweiterung der obigen Gleichung durch entsprechende *lag*-Variablen ist dies modellierbar:

$$Y_t = \delta + \theta_1 Y_{t-1} + \theta_2 Y_{t-12} + \varepsilon_t$$

Der resultierende Erwartungswert $E\{Y_t\}$ für einen AR(1)-Prozess kann folgendermaßen berechnet werden:

$$E\{Y_t\} = \delta + \theta E\{Y_{t-1}\}$$

Unterstellt man nun, dass der Erwartungswert unabhängig von der Zeit ist, ergibt sich für den Mittelwert μ :

$$\mu \equiv E\{Y_t\} = \frac{\delta}{1-\theta}$$

Im Rahmen zeitreihenanalytischer Betrachtungen wird zumeist eine vereinfachte Schreibweise verwendet, die Veerbeck (2008: 270f.) folgend auch im Weiterem Anwendung findet. Demnach wird $y_t \equiv Y_t - \mu$ definiert. Dadurch lässt sich der AR(1)-Prozess auch so ausdrücken:

$$y_t = \theta y_{t-1} + \varepsilon_t$$

Verallgemeinert auf AR(p)-Prozesse ergibt sich:

$$y_t = \theta_1 y_{t-1} + \theta_2 y_{t-2} + \dots + \theta_p y_{t-p} + \varepsilon_t, \quad \text{mit } p = 1, 2, 3, \dots$$

Ein anderer einfacher Prozess einer möglichen Zeitabhängigkeit stellt der so genannte *moving-average*-Prozess erster Ordnung dar (MA(1)-Prozess). Demnach sind die Werte von Y_t als Prozess der *white-noise*-Komponente definiert:

$$Y_t = \mu + \varepsilon_t + \alpha \varepsilon_{t-1}$$

Dies bedeutet, dass – um bei unserem Beispiel zu bleiben – der Stimmenanteil einer Partei in einem gegebenen Monat vom Mittelwert μ des Stimmenanteils dieser Partei, der Zufallskomponente ε_t und einer gewichteten Zufallskomponente des Vormonats ε_{t-1} abhängt. Mit dieser Modellierung können Trendprozesse erfasst werden. Unter der Maßgabe, dass $y_t \equiv Y_t - \mu$, lässt sich ein MA(1)-Prozess alternativ folgendermaßen formulieren:

$$y_t = \varepsilon_t + \alpha \varepsilon_{t-1}$$

Entsprechend dem obigen Vorgehen ist auch im Fall von MA-Prozessen eine Erweiterung auf q Ordnungen möglich:

$$y_t = \varepsilon_t + \alpha_1 \varepsilon_{t-1} + \alpha_2 \varepsilon_{t-2} + \dots + \alpha_q \varepsilon_{t-q}, \quad \text{mit } q = 1, 2, 3, \dots$$

Kombiniert man AR-Prozesse p -ter Ordnung mit MA-Prozessen q -ter Ordnung erhält man das allgemeinere ARMA(p, q)-Modell (Verbeek 2008: 275ff.):

$$y_t = \theta_1 y_{t-1} + \dots + \theta_p y_{t-p} + \varepsilon_t + \alpha_1 \varepsilon_{t-1} + \dots + \alpha_q \varepsilon_{t-q}$$

AR- und MA-Prozesse weisen jeweils unterschiedliche Autokovarianzen auf, welche die gemeinsame Verteilung aller Werte von Y_t charakterisieren. Angesprochen sind dabei die Kovarianzen zwischen Y_t und den entsprechenden *lags* Y_{t-k} mit $k = 1, 2, 3, \dots$ ³³ Unterstellt wird dabei jeweils, dass die Varianzen sowie Autokovarianzen unabhängig sind von der Zeit t – diese Annahme wird als Statio-

33 Die Indizes k , p und q kennzeichnen jeweils die Anzahl an *lags*. Wird sowohl auf AR- als auch auf MA-Prozesse rekuriert, wird der Index k verwendet. p bezieht sich ausschließlich auf AR-Prozesse, q auf MA-Prozesse.

narität bezeichnet. Für AR(k)-Prozesse lässt sich die Autokovarianz berechnen durch:³⁴

$$\text{cov}\{Y_t, Y_{t-k}\} = \theta^k \frac{\sigma^2}{1-\theta^2}, \quad \text{für } k = 1, 2, 3, \dots$$

Für MA(k)-Prozesse ergibt sich:

$$\text{cov}\{Y_t, Y_{t-k}\} = 0, \quad \text{für } k = 2, 3, 4, \dots$$

Die Annahme der Stationarität ist im Rahmen der Zeitreihenanalyse zentral, weshalb hierauf etwas genauer eingegangen wird. Es wurde bereits angedeutet, dass mit Stationarität die Unabhängigkeit der Autokovarianzen vom Zeitparameter t angesprochen ist. Im Fall der strikten Stationarität wird unterstellt, dass die Verteilung von Y_t dieselbe ist wie jede andere Verteilung von Y_t . Schwache Stationarität hingegen unterstellt lediglich, dass die Mittelwerte, Varianzen und Kovarianzen der Zeitreihe Y_t unabhängig von t sind. Formal ausgedrückt muss ein schwach stationärer Prozess diese Bedingungen erfüllen (Verbeek 2008: 273):

- (i) $E\{Y_t\} = \mu$
- (ii) $\text{var}\{Y_t\} = E\{(Y_t - \mu)^2\} = \gamma_0 < \infty$
- (iii) $\text{cov}\{Y_t, Y_{t-k}\} = E\{(Y_t - \mu)(Y_{t-k} - \mu)\} = \gamma_k, \quad k = 1, 2, 3, \dots$

Da die Autokovarianzen γ_k nicht einheitsfrei sind, werden diese üblicherweise standardisiert. Dies geschieht durch die Definition der Autokorrelationen ρ_k :

$$\rho_k = \frac{\text{cov}\{Y_t, Y_{t-k}\}}{\text{var}\{Y_t\}} = \frac{\gamma_k}{\gamma_0}, \quad \text{wobei } -1 \leq \rho_k \leq 1$$

Betrachtet man die Autokorrelationen als Funktion der Anzahl an *lags* k , dann ergibt sich die Autokorrelationsfunktion beziehungsweise das Korrelogramm der Zeitreihe Y_t . Diese Funktion beschreibt den Prozess der zeitlichen Entwicklung von Y_t , womit der Autokorrelationsfunktion eine zentrale Rolle bei der Entdeckung der der Zeitreihe zugrunde liegenden zeitlichen Abhängigkeitsstruktur

34 Für eine Herleitung der Autokovarianzen vgl. Verbeek (2008: 271ff.) oder ausführlicher Kirchgässner und Wolters (2006: 25ff.).

zukommt. So ergibt sich für MA(q)-Modelle, dass die Autokorrelationsfunktion nach q lags den Wert Null annimmt. Damit lässt sich anhand – beispielsweise graphischer Darstellungen – der Autokorrelationsfunktionen das Vorliegen von Trendkomponenten bei der gegebenen Zeitreihe feststellen und die korrekte Spezifikation des MA-Modells festlegen.

Zur Identifikation der lag-Struktur in Bezug auf autoregressive Komponenten wird üblicherweise die partielle Autokorrelationsfunktion herangezogen. Diese misst die partielle Korrelation zwischen Y_t und Y_{t-p} bei Kontrolle auf die intervenierenden Werte $Y_{t-1}, \dots, Y_{t-p+1}$. Das angemessene AR(p)-Modell weist partielle Autokorrelationen nahe Null nach dem p -ten lag auf (Verbeek 2008: 298).

Die Bedingungen der schwachen Stationarität können bei empirisch zu beobachtenden Zeitreihen verletzt werden, beispielsweise dann, wenn die *moving-average*-Koeffizienten in einem MA(1)-Modell über die Zeit variieren:

$$y_t = \varepsilon_t + g(t) \varepsilon_{t-1}$$

In diesem Fall ergibt sich, dass die Kovarianzen nicht mehr – wie gefordert – von t unabhängig sind. Auch im Fall eines AR(1)-Prozesses ($y_t = \theta y_{t-1} + \varepsilon_t$) kann die Annahme der Stationarität verletzt werden. Dies ist etwa dann der Fall, wenn $\theta = 1$, da keine Lösung der Varianz des Prozesses, welche konsistent mit den Bedingungen an stationäre Prozesse wäre, vorliegt (Verbeek 2008: 280ff.). Prozesse mit $\theta = 1$ werden als autoregressive Prozesse erster Ordnung mit *unit root* oder auch als *random walk* bezeichnet.

Da die Erfüllung der Bedingungen schwacher Stationarität zur korrekten Schätzung von ARMA-Modellen notwendig ist, müssen gegebenenfalls nicht-stationäre Zeitreihen stationär gemacht werden. Dies kann in der Regel durch einmalige Differenzierung³⁵ der Zeitreihe gewährleistet werden. Liegt beispielsweise dieser nicht-stationäre ARMA(2, 1)-Prozess vor (Verbeek 2008: 281),

$$y_t = 1,2y_{t-1} - 0,2y_{t-2} + \varepsilon_t - 0,5\varepsilon_{t-1},$$

35 Eine ausführliche Einführung in die mathematischen Grundlagen der Differenzenrechnung bietet Goldberg (1968).

dann ergibt einmaliges Differenzieren die stationäre Zeitreihe:

$$\Delta y_t = 0,2 \Delta y_{t-1} + \varepsilon_t - 0,5 \varepsilon_{t-1}$$

Es ist nun möglich, die ARMA(p, q)-Modelle zu ARIMA(p, d, q)-Modellen zu erweitern, indem die d-te Ordnung der Differenzierung zur Herstellung von Stationarität der Zeitreihe zusätzlich berücksichtigt wird. Das Vorliegen eines *unit roots* kann mit Hilfe des von Dickey und Fuller (1979) entwickelten Testverfahrens untersucht werden.

Dieses Verfahren testet die Nullhypothese, dass die vorliegende Zeitreihe einen *unit root* ($\theta = 1$) aufweist. Das grundsätzliche Vorgehen wird anhand eines einfachen AR(1)-Prozesses erläutert (siehe zur Verallgemeinerung bspw. Verbeek 2008: 284ff.). Ein AR(1)-Prozess ist – wie oben eingeführt – definiert durch:

$$Y_t = \delta + \theta Y_{t-1} + \varepsilon_t$$

In einem stationären AR(1)-Prozess muss die Konstante $\delta = (1 - \theta) \mu$ sein, wobei μ den Mittelwert der Zeitreihe repräsentiert (siehe oben). Gleichfalls impliziert die Nullhypothese des Vorliegens eines *unit roots*, dass $\delta = 0$. Der Dickey-Fuller-Testwert (DF) ergibt sich nach Schätzung eines OLS-Modells durch:

$$DF = \frac{\hat{\theta} - 1}{se(\hat{\theta})}$$

Dabei steht im Nenner der Standardfehler aus der OLS-Regression. Im Prinzip entspricht dies der t-Statistik im OLS-Modell, jedoch zeigen Dickey und Fuller (1979), dass die t-Werte unter der Nullhypothese eines *unit roots* nicht asymptotisch t-verteilt sind, sondern eine im Vergleich hierzu nach links verschobene Verteilung aufweisen. In der Folge würde bei Unterstellung einer t-Verteilung ein *unit root* fälschlicherweise zu oft zurückgewiesen werden. Dementsprechend ist der Vergleich mit anderen kritischen Werten notwendig, welche Dickey und Fuller (1979: 430) mittels Monte-Carlo-Simulationen ermitteln.

Mit den vorgestellten Konzepten ist es möglich, die vorliegenden Zeitreihen angemessen zu beschreiben. Mit Hilfe der Autokorrelationsfunktion können *moving-average*-Prozesse, mit Hilfe partieller Autokorrelationsfunktionen autoregressive Komponenten von Zeitreihen aufgedeckt werden. Die Dickey-Fuller-Tests erlauben schließlich den Test auf Vorliegen von Stationarität (auch diffe-

renzierter Zeitreihen) und liefern in ihrer erweiterten Form zudem Signifikanztests für die Relevanz der vermuteten AR(p)- sowie MA(q)-Prozesse.

Auf Grundlage dieser Informationen können die ARIMA(p, d, q)-Modelle geschätzt werden. Die Schätzung erfolgt im verwendeten Statistikprogramm Stata 10.1 mittels des *Maximum-Likelihood*-Verfahrens (StataCorp 2007a: 42ff.). Weiterhin notwendig erscheint die Bestimmung des angemessenen Modells. Da häufig keine theoretischen Vorgaben vorliegen, welche die korrekte Spezifikation des Zeitreihenmodells vorgeben, wird in der Regel das statistisch angemessenste Modell gesucht. Hierzu werden zunächst die Ergebnisse der Analyse der Autokorrelationsfunktionen, der partiellen Autokorrelationsfunktionen sowie der Dickey-Fuller-Tests herangezogen. Nach Schätzung der ARIMA-Modelle können unterschiedliche Spezifikationen für dieselbe Zeitreihe anhand des Bayesianischen Informationskriteriums (BIC) oder des Akaike Informationskriteriums (AIC) miteinander verglichen werden. Dabei gilt grundsätzlich, dass diese Maße Kriterien zur Bewertung des *trade offs* zwischen einer sparsamen Modellierung und der besten Anpassung umfassen. Ziel ist demnach anhand dieser Maßzahlen zur Beschreibung von Zeitreihen dasjenige Modell auszuwählen, für den dieser *trade off* – gemessen am jeweiligen Kriterium – am besten ausfällt. AIC und BIC lassen sich folgendermaßen berechnen:

$$AIC = \log \hat{\sigma}^2 + 2 \frac{p+q+1}{T}$$

$$BIC = \log \hat{\sigma}^2 + \frac{p+q+1}{T} \log T$$

Dabei werden wie ersichtlich unterschiedliche Kriterien für den *trade off* zwischen Angemessenheit (*Loglikelihood*-Wert) und Sparsamkeit (Zahl der freien Parameter) angelegt. Beiden Kriterien gemeinsam ist, dass Modelle mit kleineren AIC- beziehungsweise BIC-Werten Modellen mit größeren Werten bevorzugt werden sollten. In dieser Arbeit wird der Empfehlung Verbeeks (2008: 300) folgend das Bayesianische Informationskriterium zur Modellselektion herangezogen.

3.3.3.2 Mehrebenenanalyse

Grundsätzlich wäre bei der multivariaten Analyse der Einflussfaktoren auf die Prognosegüte die Verwendung des linearen Regressionsmodells zur Schätzung der Parameter möglich, da die abhängige Variable metrisches Skalenniveau aufweist. Jedoch liegen hierarchisch gegliederte Daten vor. Dies betrifft zwei Aspekte. Zum einen wurde vermutet, dass aufgrund der politischen Gewichtungen, welche institutsspezifisch variieren und Geheimnis der jeweiligen Unternehmen sind, *house effects* vorliegen (*Hypothese 4a*). Zum anderen, so wurde argumentiert, kann man Veränderungen der Prognosegüte über die Zeit vermuten (*Hypothesen 7a und 7b*). Insofern kann man davon ausgehen, dass die beobachteten Abweichungen nicht nur systematisch nach Institut variieren, sondern auch wahlspezifisch.

Treffen diese Vermutungen zu – und in Abschnitt 3.4.3 wird gezeigt, dass dies der Fall ist – wird bei der Anwendung von OLS-Regressionen die Modellvoraussetzung der Homoskedastizität verletzt, da sich Elemente innerhalb einer Gruppe ähnlicher sind als Elemente verschiedener Gruppen. Statistisch gesehen führt eine Nichtbeachtung dieser Besonderheiten zu einer Unterschätzung der Standardfehler (vgl. Snijders/Bosker 1999: 7). Darüber hinaus besteht bei der Vernachlässigung der hierarchischen Datenstruktur die Gefahr eines „fallacy of the wrong level“ (Hox 2002: 3), da Zusammenhänge, die auf einer Ebene gelten, fälschlicherweise auf eine andere Ebene übertragen werden.

Um daraus resultierende falsche Ergebnisse zu vermeiden, müssen Methoden der Mehrebenenanalyse angewendet werden. Unter diesem Begriff werden eine Reihe verschiedener Modelle zusammengefasst, welche die Gruppierung der Daten berücksichtigen. Hervorzuheben ist für die Klasse dieser Modelle, dass sie auch eine adäquate Berücksichtigung von Kontextmerkmalen als erklärende Variablen ermöglichen (vgl. Hinz 2005). Im vorliegenden Fall würde es sich hierbei um Eigenschaften der Meinungsforschungsinstitute beziehungsweise der einzelnen Wahlen handeln. Mit der als einflussreich erachteten Höhe der Wahlbeteiligung (*Hypothese 5*) liegt ein solches Kontextmerkmal auf Wahlebene vor.

Die statistische Modellierung der Mehrebenenanalyse trägt der Tatsache Rechnung, dass sich die Beobachtungen innerhalb einer Gruppe ähnlicher sind als diejenigen zwischen den Untersuchungseinheiten. Zunächst wird die Ebene der Meinungsforschungsinstitute betrachtet. Dabei repräsentiert der Index i die

Wahlabsichtsfragen und j die Meinungsforschungsinstitute. Am einfachsten ist die Berücksichtigung größerer Ähnlichkeiten der Beobachtungen innerhalb einzelner Meinungsforschungsinstitute durch die Einführung von Dummy-Variablen für die einzelnen Gruppen j in ein OLS-Modell (Baltagi 1999: 309):

$$Y_{ij} = \beta_0 + \beta_1 \mathbf{X}_{ij} + \sum_{j=1}^N u_j D_j + \varepsilon_{ij}, \quad \text{mit } i = 1, 2, 3, \dots, n_j \text{ und } j = 1, 2, 3, \dots, N$$

Hierbei ist D_j eine Dummy-Variable für das j -te Meinungsforschungsinstitut. Der Fehlerterm ε_{ij} entspricht dem üblichen Fehlerterm in der OLS-Regression mit einem Mittelwert von Null und einer konstanten Varianz. \mathbf{X}_{ij} repräsentiert einen Vektor mit den erklärenden Variablen auf Ebene der Wahlabsichtsfragen. Die Berücksichtigung von Dummy-Variablen für die einzelnen Untersuchungseinheiten reduziert die Zahl der Freiheitsgrade. Dies führt insbesondere bei einer hohen Anzahl von Gruppen zu Problemen, weshalb in diesem Fall auf *fixed-effects*-Modelle zurückgegriffen wird, welche sich durch eine Aufspaltung des Fehlerterms in der Regressionsgleichung auszeichnen, um wiederum die größere Ähnlichkeit der Beobachtungen innerhalb der Gruppen abbilden zu können (Snijders/Bosker 1999: 39ff.):

$$Y_{ij} = \beta_0 + \beta_1 \mathbf{X}_{ij} + u_j + \varepsilon_{ij}$$

Der hier eingeführte wahlunabhängige Term u_j bildet die konstanten Unterschiede der Meinungsforschungsinstitute im Hinblick auf deren durchschnittliche Prognosegüte ab. Im Ergebnis sind die beiden Modellierungen identisch. Die *fixed-effects*-Modellierung erlaubt es allerdings nicht Kontextmerkmale zu berücksichtigen, welche konstant sind innerhalb der Gruppen j (Snijders/Bosker 1999: 43).

Für die Meinungsforschungsinstitute werden im Rahmen dieser Analyse Dummy-Variablen eingefügt. Diese Vorgehensweise wird gewählt, da erstens nur relativ wenige Institute unterschieden werden und somit das Problem geringer Freiheitsgrade hier nicht gegeben ist. In diesem Zusammenhang ist zweitens auf die Simulationsergebnisse von Maas und Hox (2005) hinzuweisen, die gezeigt haben, dass für eine korrekte Schätzung der Varianz der *Level-2-intercepts* mindestens 50 Gruppen erforderlich sind. Da dies hier nicht der Fall ist, erscheint auch aus statistischer Perspektive die Dummy-Modellierung angemessener. Drittens bietet die Einführung von Dummys die Möglichkeit der direkten Beo-

bachtung der Institutsunterschiede – sowohl hinsichtlich des Ausmaßes als auch der Richtung. Schließlich werden auf Ebene der Meinungsforschungsinstitute keine erklärenden Variablen betrachtet. Grund hierfür ist, dass diese Arbeit nicht zum Ziel hat, die Varianz der Prognosegüte zwischen den Instituten mit Eigenschaften der Institute zu erklären. Auch deshalb erscheint die Dummy-Modellierung die sparsamste zur Berücksichtigung der hierarchischen Datenstruktur.

In Bezug auf die Wahlen, im Folgenden mit dem Index t bezeichnet, stellt sich dies anders dar. Die theoretischen Überlegungen legen den Einfluss der Höhe der Wahlbeteiligung auf die Prognosegüte nahe. Insofern muss zur korrekten Berücksichtigung dieser – für einzelne Wahlen konstanten – Kontextvariable eine angemessene Modellierungsstrategie angewendet werden. Da *fixed-effects*-Modelle nur die Berücksichtigung variierender Kontextmerkmale über die einzelnen Bundestagswahlen ermöglichen, werden *random-intercept*-Modelle spezifiziert (Snijders/Bosker 1999: 45ff.). Diese lassen zu, dass der *intercept* der jeweiligen Modelle über die Gruppen hinweg zufällig variiert. Damit werden sowohl Zusammenhänge innerhalb als auch zwischen den Gruppen berücksichtigt. Das geschätzte Modell entspricht hierbei in der Form dem *fixed-effects*-Modell, jedoch wird im Fall des *random-intercept*-Modells angenommen, dass u_j eine normalverteilte latente Variable ist, wobei nur dessen Varianz im Rahmen des Schätzverfahrens geschätzt wird.

Kombiniert man nun das *random-intercept*-Modell für die Bundestagswahlen auf Ebene 2 mit dem OLS-Modell inklusive Dummies für die Meinungsforschungsinstitute, ergibt sich folgende Regressionsgleichung, die im Rahmen der empirischen Analyse geschätzt wird:

$$Y_{ijt} = \beta_0 + \beta_1 \mathbf{X}_{ijt} + \sum_{j=1}^N u_j D_j + \beta_2 z_t + u_t + \varepsilon_{ijt}$$

Die in Abschnitt 2.3.4 vorgestellte Analyse umfasst 15 Bundestagswahlen. Den Ergebnissen von Maas und Hox (2005) können somit aufgrund der geringen Gruppenanzahl die diesbezüglich berichteten Standardfehler unterschätzt werden. Gleichwohl zeigt die Studie von Maas und Hox (ebd.), dass die Schätzung der Standardfehler der festen Regressionskoeffizienten nur in geringem Maß von der Gruppenanzahl abhängt. Da die Effekte der Variablen im Mittelpunkt der Analysen stehen, erscheint dieses Problem weniger zentral für die nachfolgend

berichteten Befunde. Zudem zeigt sich bei Berechnung klassischer linearer Regressionsmodelle das angesprochene Problem unterschätzter Standardfehler der Koeffizienten. Insofern scheint das gewählte Vorgehen mit Berücksichtigung der Mehrebenenstruktur im Hinblick auf das Forschungsinteresse angemessener zu sein. Gleichwohl müssen die Varianzkomponenten der *random intercepts* mit Vorsicht interpretiert werden.

Diesen Überlegungen folgend werden in Abschnitt 3.4.3 Modelle geschätzt, welche auf Ebene 1 die Sonntagsfragenergebnisse und auf Ebene 2 die Bundestagswahlen umfassen. Die Institute werden als Dummy-Variablen einbezogen und so als weitere Ebene betrachtet.³⁶

3.4 Empirische Resultate

Die im Folgenden berichteten empirischen Resultate beziehen sich auf drei Aspekte. Erstens wird anhand der Politbarometer-Daten der Einfluss der politischen Gewichtung auf Sonntagsfragen exemplarisch analysiert (Abschnitt 3.4.1). In diesem Kontext soll das Ausmaß der politischen Gewichtung im Zeitablauf dargestellt werden, um dessen Relevanz abschätzen zu können. Betont werden muss, dass damit die Zielsetzung bereits umschrieben ist. Ziel ist also nicht, die zur Anwendung kommenden Gewichtungsfaktoren *ex post* zu ermitteln und mittels multivariater Verfahren mit Zusatzinformationen der Umfragen, etwa der Rückerinnerungsfrage, die Rohdaten möglichst exakt in die berichteten Projektionen zu überführen.

Im zweiten Schritt wird die Prognosegüte der berichteten Sonntagsfragen beschrieben (Abschnitt 3.4.2). Ziel hierbei ist es, eine systematische empirische Beschreibung der Qualität der Wahlabsichtsumfragen unter Einbezug der diskutierten Gütemaße zu erstellen, welche als Vergleichsmaßstab zur empirischen Bewertung anderer Prognoseversuche herangezogen werden kann.

Drittens wird die Relevanz der abgeleiteten Einflussfaktoren auf die Prognosegüte der Sonntagsfragen empirisch geprüft (Abschnitt 3.4.3).

36 Sämtliche Berechnungen wurden mit Stata 10.1 durchgeführt. Eine Darstellung der Mehrebenenanalyse mit Stata liefern Rabe-Hesketh und Skrondal (2008).

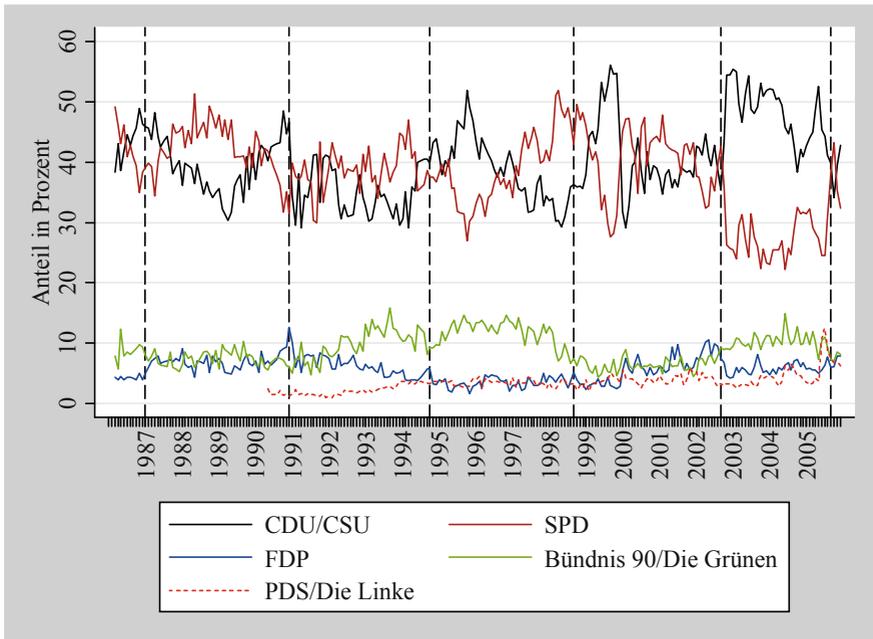
3.4.1 Die politische Gewichtung der Wahlabsichtsfrage

In einem ersten Analyseschritt werden zur Abschätzung des Einflusses der politischen Gewichtung bei der Erstellung von Voraussagen, welche auf Sonntagsfragen basieren, anhand der Politbarometer-Studien monatlich die gemessenen Stimmenanteile gemäß der Rohdaten mit den veröffentlichten Projektionen der Forschungsgruppe Wahlen miteinander verglichen. In Bezug auf die Auswertung der Politbarometer-Daten ist vorzuschicken, dass es sich hierbei um Zeitreihendaten handelt, welche spezielle Verfahren zu ihrer Beschreibung und Analyse erfordern (siehe hierzu Abschnitt 3.3.4.1).

In Abbildung 3.1 sind die Stimmenanteile von CDU/CSU, SPD, FDP, Bündnis 90/Die Grünen sowie der PDS/Die Linke dargestellt, wie sie sich aus den Rohdaten gemäß der Wahlabsichtsfrage für den Zeitraum vom März 1986 bis Dezember 2005 ergeben. Der Startzeitpunkt ergibt sich daraus, dass die Forschungsgruppe Wahlen erstmals im März 1986 eine politisch gewichtete Projektion veröffentlicht hat (vgl. Wüst 2003). Die Bundestagswahl 2009 kann hier nicht betrachtet werden, da die Rohdaten des Politbarometer bisher nur bis Ende 2007 als *scientific use file* bezogen werden können. Die berichteten Stimmenanteile beziehen sich bis einschließlich März 1990 auf die alten Bundesländer und seit April 1990 auf das gesamte Bundesgebiet. Weiterhin sind in der Abbildung die jeweiligen Termine der Bundestagswahlen (vertikale gestrichelte Linien) eingezeichnet.

Abbildung 3.1 zeigt bei allen Parteien deutliche Schwankungen im Zeitverlauf und insbesondere erkennbare monatliche Ausschläge nach oben und unten. Klare Trends, saisonale oder konjunkturelle Effekte lassen sich graphisch nicht ausmachen, vielmehr scheinen die ermittelten Stimmenanteile um die jeweiligen Mittelwerte unregelmäßig zu schwanken. Einzig bei der PDS/Die Linke ist ein Aufwärtstrend im Zeitverlauf relativ deutlich. Betrachtet man die Zeitreihe der SPD genauer, kann man einen leichten Abwärtstrend graphisch ablesen – beide Vermutungen werden durch Dickey-Fuller-Tests mit Berücksichtigung einer Trendkomponente bestätigt (nicht dargestellt). Die anderen Parteien weisen bei entsprechenden Tests wie nach der graphischen Analyse vermutet keinen signifikanten Trend auf.

Abbildung 3.1: Ergebnisse der Sonntagsfragen 1986-2005, Politbarometer-Rohdaten



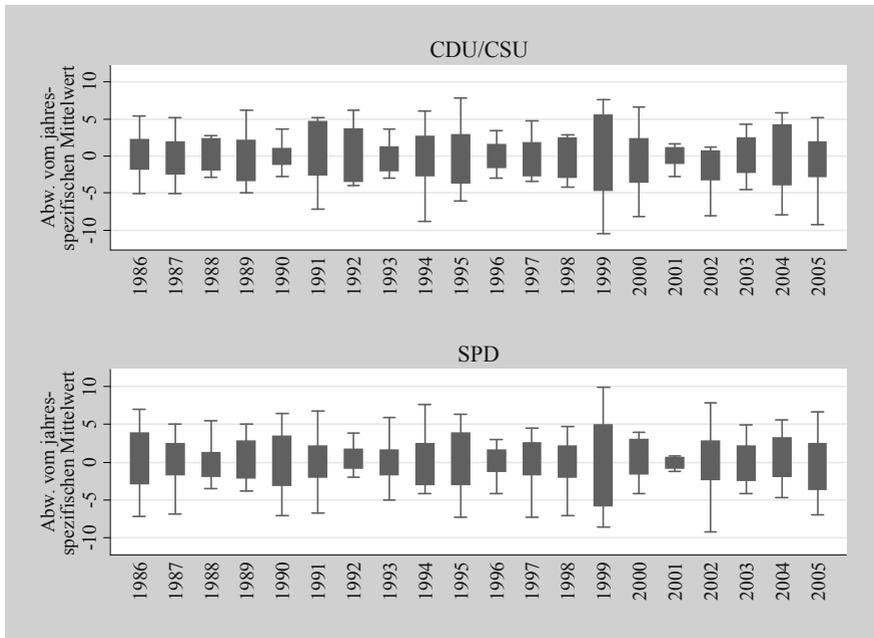
Anmerkungen: Die Darstellung beruht auf eigenen Auswertungen der Politbarometer-Daten. Bei der Berechnung der Stimmenanteile der einzelnen Parteien wurde das Designgewicht verwendet. Zum Zeitpunkt der Befragung Unentschiedene und potenzielle Nichtwähler wurden nicht berücksichtigt. Bis einschließlich März 1990 beziehen sich die dargestellten Stimmenanteile auf die alten Bundesländer, ab April 1990 auf das gesamte Bundesgebiet. Die vertikal gestrichelten Linien sind die Zeitpunkte der Bundestagswahlen im betrachteten Zeitraum.

Häufiger wird vermutet, dass die politische Stimmung, gemessen anhand der Sonntagsfrage, in jüngerer Zeit größere Ausschläge aufweist (vgl. bspw. Antholz 2001: 102). Abbildung 3.1 liefert hierfür zumindest für den durch das Politbarometer abgedeckten Zeitraum keine Hinweise. Zwar zeigen sich die deutlichsten Ausschläge bei den großen Parteien nach der Bundestagswahl 2002, aber bereits nach der Wahl von 1987 zeigt sich eine Volatilität, welche kaum geringer war als etwa nach der Wahl von 1998. Betrachtet man die kleinen Parteien, dann zeigen sich auch hier deutliche Schwankungen. So variiert der Stimmenanteil der Grünen von fünf bis über 15 Prozent und auch bei der FDP sind deutliche Schwan-

kungen und auch temporäre Ausschläge im Zeitablauf zu erkennen. Ein Trend hinsichtlich der Volatilität ist hier allenfalls für die PDS/Die Linke zu konstatieren; deren Anhängerschaft scheint im Zeitverlauf insgesamt größer zu werden, allerdings scheinen auch die Schwankungen zuzunehmen.

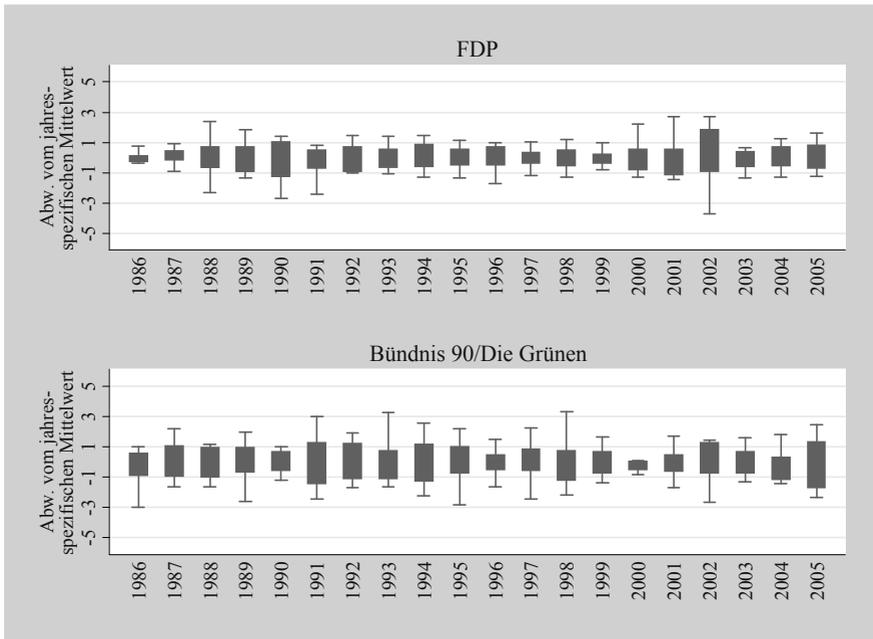
Eine genauere Analyse der Volatilität im Zeitablauf ist den Abbildungen 3.2 bis 3.4 zu entnehmen. Dargestellt sind hier die Schwankungen um die für jede Partei jahresspezifischen Mittelwerte ihrer Stimmenanteile.

Abbildung 3.2: Volatilität der Stimmung bei der CDU/CSU und der SPD nach Jahren, 1986-2005, Politbarometer-Rohdaten



Anmerkungen: Eigene Berechnungen und Darstellung auf Grundlage der Politbarometer-Daten. Abgebildet sind Box-Plots, wobei die grauen Kästen jeweils das 25- bzw-75-Prozent-Quartil repräsentieren. Die Whiskers geben jeweils das Quartil plus den 1,5-fachen Interquartilsabstand an. Etwaige Ausreißer wurden bei der Berechnung berücksichtigt, jedoch nicht dargestellt. Die Nulllinie repräsentiert den jahresspezifischen Mittelwert der in den Rohdaten ermittelten Stimmenanteile nach Partei.

Abbildung 3.3: Volatilität der Stimmung bei der FDP und Bündnis 90/Die Grünen nach Jahren, 1986-2005, Politbarometer-Rohdaten

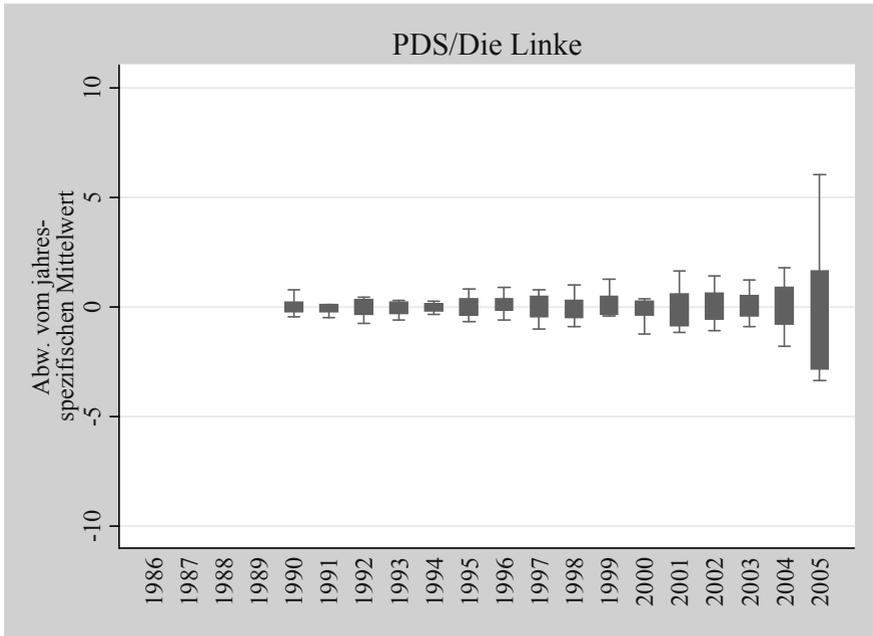


Anmerkungen: Eigene Berechnungen und Darstellung auf Grundlage der Politbarometer-Daten. Abgebildet sind Box-Plots, wobei die grauen Kästen jeweils das 25- bzw. 75-Prozent-Quartil repräsentieren. Die Whiskers geben jeweils das Quartil plus den 1,5-fachen Interquartilsabstand an. Etwaige Ausreißer wurden bei der Berechnung berücksichtigt, jedoch nicht dargestellt. Die Nulllinie repräsentiert den jahresspezifischen Mittelwert der in den Rohdaten ermittelten Stimmenanteile nach Partei.

Dabei bestätigt sich das bereits skizzierte Bild: Außer für die PDS/Die Linke sind anhand der Rohdaten des Politbarometers keine bedeutenden Volatilitätszunahmen zu erkennen. Selbst hier ist der zu beobachtende Trend vorsichtig zu konstatieren, denn zum einen zeigt sich die zunehmende Volatilität erst in den jüngsten Jahren und damit stützt sich diese Interpretation auf wenige Datenpunkte. Zum anderen fusionierten die in den alten Bundesländern insbesondere von ehemaligen SPD-Mitgliedern getragene WASG und die vorwiegend in den neuen Bundesländern verwurzelte PDS. Insofern erscheint es fraglich, ob eine Fortschreibung der ermittelten Stimmenanteile für die PDS für die neue Partei Die

Linke zulässig ist. Dies soll hier nicht weiter diskutiert werden, da im Hinblick auf die im Rahmen dieser Arbeit gestellten Zielsetzung keine alternative Behandlung möglich erscheint. Dies ist nicht zuletzt auf die Praxis der Meinungsforschungsinstitute zurückzuführen, deren Daten den hier vorgestellten Analysen zugrunde liegen, und die eine entsprechende Fortschreibung der Stimmenanteile vollziehen.

Abbildung 3.4: Volatilität der Stimmung bei der PDS/Die Linke nach Jahren, 1990-2005, Politbarometer-Rohdaten



Anmerkungen: Eigene Berechnungen und Darstellung auf Grundlage der Politbarometer-Daten. Abgebildet sind Box-Plots, wobei die grauen Kästen jeweils das 25- bzw-75-Prozent-Quartil repräsentieren. Die Whiskers geben jeweils das Quartil plus den 1,5-fachen Interquartilsabstand an. Etwaige Ausreißer wurden bei der Berechnung berücksichtigt, jedoch nicht dargestellt. Die Nulllinie repräsentiert den jahresspezifischen Mittelwert der in den Rohdaten ermittelten Stimmenanteile nach Partei.

Diese rein graphische Deskription wird um eine zeitreihenanalytische Betrachtung erweitert, um der vorliegenden Datenstruktur gerecht zu werden und etwaige Fehlschlüsse aufgrund der Anwendung nicht angemessener statistischer Verfahren zu vermeiden. Das hierbei relevante Vorgehen wurde eingangs bereits skizziert.

Anhand partieller Autokorrelationsplots wurde jeweils die vorliegende *lag*-Struktur eingehender analysiert (nicht dargestellt). Es zeigt sich für CDU/CSU sowie die PDS/Die Linke, dass nur ein *lag* berücksichtigt werden muss. Bei SPD und FDP sind zwei *lags* relevant und bei Bündnis 90/Die Grünen drei. Die Autokorrelationen weisen darüber hinaus für alle Zeitreihen außer für die Partei Bündnis 90/Die Grünen auf vorliegende Stationarität hin (nicht dargestellt). Dies wurde mit Dickey-Fuller-Tests überprüft, wobei jeweils eine Konstante ungleich Null, die zuvor ermittelte Anzahl *lags* sowie gegebenenfalls eine Trendkomponente berücksichtigt wurden (nicht dargestellt).

Als Ergebnis dieser Analysen lässt sich festhalten, dass außer bei der Partei PDS/Die Linke kein signifikanter Trendeffekt festgestellt werden konnte. Die Vermutung eines leichten Abwärtstrends bei der SPD aus Abbildung 3.1 kann demnach nicht bestätigt werden. Weiterhin zeigt sich, dass die Zeitreihe für Bündnis 90/Die Grünen einmal differenziert werden muss, um Stationarität aufzuweisen.

In Tabelle 3.3 sind die Ergebnisse geschätzter ARIMA-Modelle zur Beschreibung der einzelnen Zeitreihen aus Abbildung 3.1 zusammengefasst. Dargestellt sind jeweils die gemäß des Bayesianischen Informationskriteriums (BIC) angemessensten Modelle (hierzu z. B. Verbeek 2008: 299f.). Dabei variiert je nach Partei die Anzahl der berücksichtigten *lags*. Für die PDS/Die Linke ist weiterhin ein schwacher negativer *moving-average*-Term festzustellen, der auf einen schwachen – in diesem Modell jedoch nicht signifikanten – Aufwärtstrend hinweist.

Tabelle 3.3: Ergebnisse der ARIMA-Modelle für die Stimmenanteile von CDU/CSU, SPD, FDP, Bündnis 90/Die Grünen und PDS/Die Linke, 1986-2005, Politbarometer-Rohdaten

	CDU/CSU	SPD	FDP	Bündnis 90/ Die Grünen (erstes Diffe- renzial)	PDS/Die Linke
AR(1)	0,842*** (0,035)	0,689*** (0,056)	0,600*** (0,063)	-0,558*** (0,064)	0,857*** (0,051)
AR(2)		0,191*** (0,054)	0,234*** (0,054)	-0,377*** (0,074)	
AR(3)				-0,220*** (0,069)	
MA(1)					-0,114 (0,063)
Konstante	40,197*** (1,377)	38,428*** (1,711)	5,609*** (0,483)	-0,000 (0,047)	3,505 (0,544)
Wald Chi ²	580,75***	595,20***	522,85***	77,85***	1.044,91***
BIC	1.266,775	1.283,898	760,509	877,361	506,000
N	238	238	238	237	188

Anmerkungen: Dargestellt sind die unstandardisierten Koeffizienten sowie die Standardfehler (in Klammern) von ARIMA-Modellen mit der jeweils in der Tabelle angegebenen Anzahl an autoregressiven Termen (AR) und gegebenenfalls einem *moving-average*-Term (MA); die Zeitreihe für Bündnis 90/Die Grünen wurde einmal differenziert, um Stationarität herzustellen; *** $p \leq 0,001$, ** $p \leq 0,01$, * $p \leq 0,05$.

Betrachtet man noch einmal die Rohdaten-Zeitreihen der beiden großen Parteien, CDU/CSU und SPD in Abbildung 3.1, so ist zu vermuten, dass der Abstand beider Parteien kurz vor Wahlterminen zumeist weniger stark ausgeprägt ist als in den Zwischenwahlzeiten. Zugleich scheint die jeweilige Oppositionspartei in den Zwischenwahlzeiten von ihrer Oppositionsrolle profitieren zu können: So liegt nach den Bundestagswahlen von 1987, 1990 und 1994 die SPD vor der regierenden CDU/CSU. Allerdings „überholt“ die SPD nach der Wahl von 1994 die CDU/CSU erst im Jahr 1997. Gleichfalls liegt nach den Bundestagswahlen 1998 und 2002 die CDU/CSU vor der SPD.

Analysiert man die absolute Differenz der Stimmenanteile von CDU/CSU und SPD regressionsanalytisch, dann kann dieser Befund jedoch nicht bestätigt werden. Die Zeitreihe der absoluten Differenz der Stimmenanteile weist einen *lag* sowie einen schwach positiven Trend auf. Das entsprechend spezifizierte ARIMA-Modell ist Tabelle 3.4 zu entnehmen, welches zeigt, dass der zeitliche Abstand zur Wahl keinen signifikanten Zusammenhang mit der absoluten Stimmenanteilsdifferenz der beiden großen Parteien aufweist.

Tabelle 3.4: Zusammenhang zwischen dem zeitlichen Abstand zur nächsten Bundestagswahl und der absoluten Differenz des Stimmenanteils zwischen CDU/CSU und SPD, 1986-2005, Politbarometer-Rohdaten

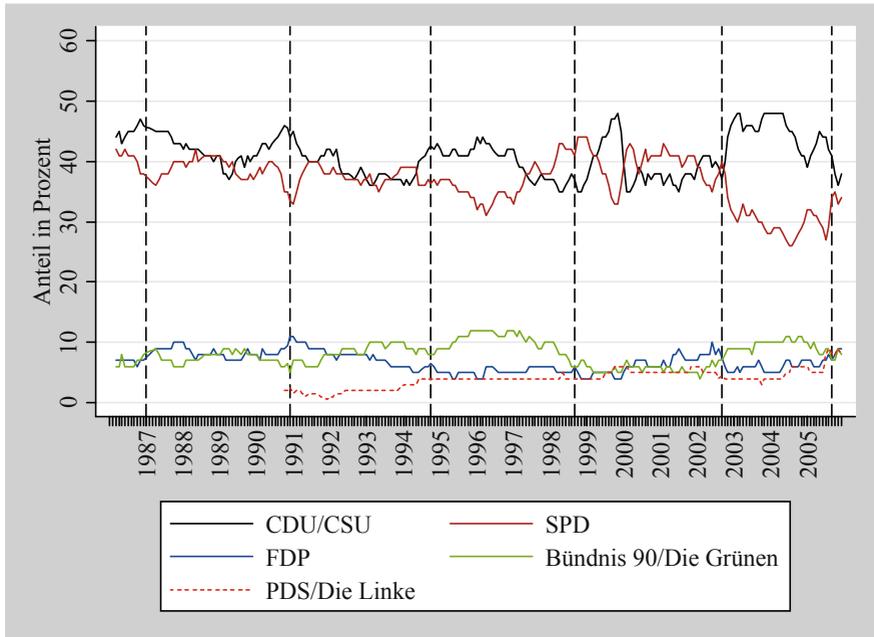
	Absolute Stimmenanteilsdifferenz zwischen CDU/CSU und SPD
Zeitlicher Abstand zur Wahl (in Monaten)	0,360 (0,276)
Quadrierter zeitlicher Abstand zur Wahl	-0,007 (0,006)
AR(1)	0,835*** (0,048)
MA(1)	-0,211** (0,081)
Konstante	6,932*** (1,130)
Wald Chi ²	412,38***
BIC	1.459,256
N	238

Anmerkungen: Dargestellt sind die unstandardisierten Koeffizienten sowie die Standardfehler (in Klammern) von ARIMA-Modellen mit einem autoregressiven Term (AR) und einem moving-average-Term (MA); *** $p \leq 0,001$, ** $p \leq 0,01$, * $p \leq 0,05$.

Nun werden die Daten der von der Forschungsgruppe Wahlen veröffentlichten Projektionen von März 1986 bis Dezember 2005 betrachtet. Diese sind Abbildung 3.5 zu entnehmen. Wiederum wurden fehlende Werte für einzelne Monate

in der Zeitreihe durch den jeweiligen Mittelwert des vorangegangenen und des nachfolgenden Monats ersetzt und die Daten beziehen sich bis einschließlich März 1990 auf die alten Bundesländer und seitdem auf Gesamtdeutschland.

Abbildung 3.5: Ergebnisse der Sonntagsfragen 1986-2005, Politbarometer-Projektionen



Anmerkungen: Eigene Darstellung auf Grundlage der veröffentlichten Politbarometer-Projektionen. Bis einschließlich März 1990 beziehen sich die dargestellten Stimmenanteile auf die alten Bundesländer, ab April 1990 auf das gesamte Bundesgebiet. Die vertikal gestrichelten Linien sind die Zeitpunkte der Bundestagswahlen im betrachteten Zeitraum.

Auf den ersten Blick erkennbar ist, dass für alle Parteien die monatlichen Schwankungen im Gegensatz zu den Ausschlägen in Abbildung 3.1 deutlich geringer ausfallen. Insgesamt ist eine deutliche Glättung der Zeitreihen festzustellen. Dies verdeutlicht Tabelle 3.5. Erkennbar ist insbesondere für die beiden großen Parteien bei jeweils sehr ähnlichen Mittelwerten eine fast nur halb so große Standardabweichung bei den Projektionen im Vergleich mit den Rohdaten. Weniger

deutlich fällt dies bei den kleineren Parteien aus – bei der PDS/Die Linke sind die Standardabweichungen sogar identisch. Interessant hier sind die doch teilweise deutlichen Mittelwertunterschiede der Zeitreihen. Während die FDP in den Projektionen im Schnitt einen um 1,2 Prozentpunkte höheren Stimmenanteil aufweist als in den Rohdaten, ist der Mittelwert bei Bündnis 90/Die Grünen in den Rohdaten um 0,9 Prozentpunkte höher als in den Projektionen.

Tabelle 3.5: Vergleich der Politbarometer-Rohdaten mit Projektionen für alle Parteien, 1986-2005

		Mittelwert	Standardabw.	N
CDU/CSU	Rohdaten	40,2	6,3	238
	Projektionen	41,0	3,4	238
SPD	Rohdaten	38,3	6,6	238
	Projektionen	37,0	3,9	238
FDP	Rohdaten	5,6	1,9	238
	Projektionen	6,8	1,6	238
Bündnis 90/Die Grünen	Rohdaten	9,0	2,6	238
	Projektionen	8,1	2,0	238
PDS/Die Linke	Rohdaten	3,5	1,5	188
	Projektionen	4,0	1,5	183

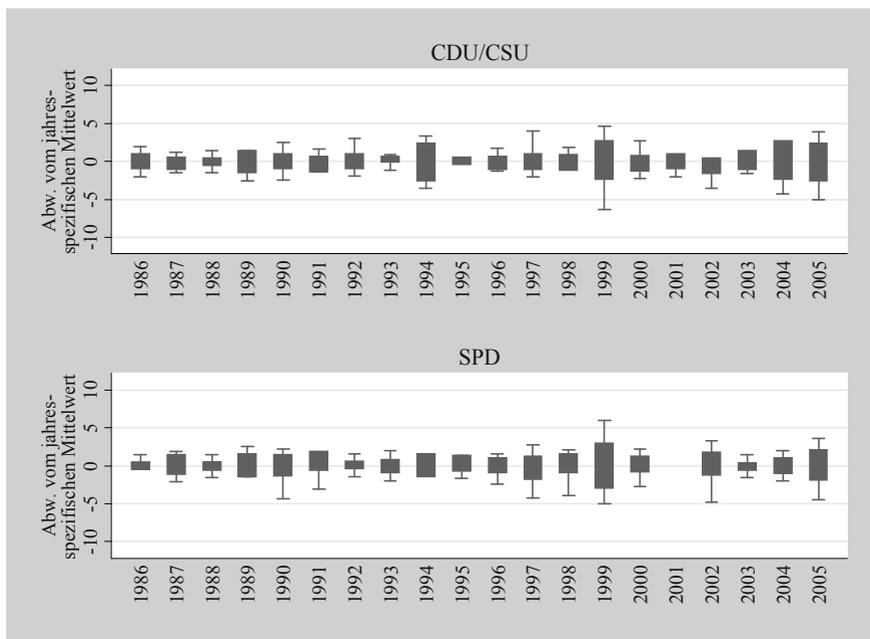
Anmerkungen: Eigene Berechnungen auf Grundlage der Politbarometer-Rohdaten sowie der veröffentlichten Politbarometer-Projektionen.

Ein denkbarer Grund für diese Abweichung ist, dass im Rahmen der politischen Gewichtung für die Projektionen der Überlegung Rechnung getragen wird, dass die FDP besonders häufig von strategischen Wählern profitiert, die ihre Stimmen splitten und beispielsweise trotz CDU/CSU-Präferenz ihre Zweitstimme am Wahltag der FDP geben.

Der Vergleich von Rohdaten und Projektionen lässt auch vermuten, dass die Volatilität hier im Zeitverlauf zumindest etwa für die SPD durchaus zugenommen hat. Um diese Vermutung zu untersuchen, wurden analog zur vorherigen Analyse für jede Partei jahresspezifische Mittelwerte der Stimmenanteile berechnet und die Differenz zu den jeweils berichteten Stimmenanteilen pro Monat

gebildet, um die Entwicklung der Streuung im Zeitablauf betrachten zu können. Die entsprechenden Box-Plots sind in den Abbildungen 3.6 bis 3.8 dargestellt.

Abbildung 3.6: Volatilität der Stimmung bei der CDU/CSU und der SPD nach Jahren, 1986-2005, Politbarometer-Projektionen

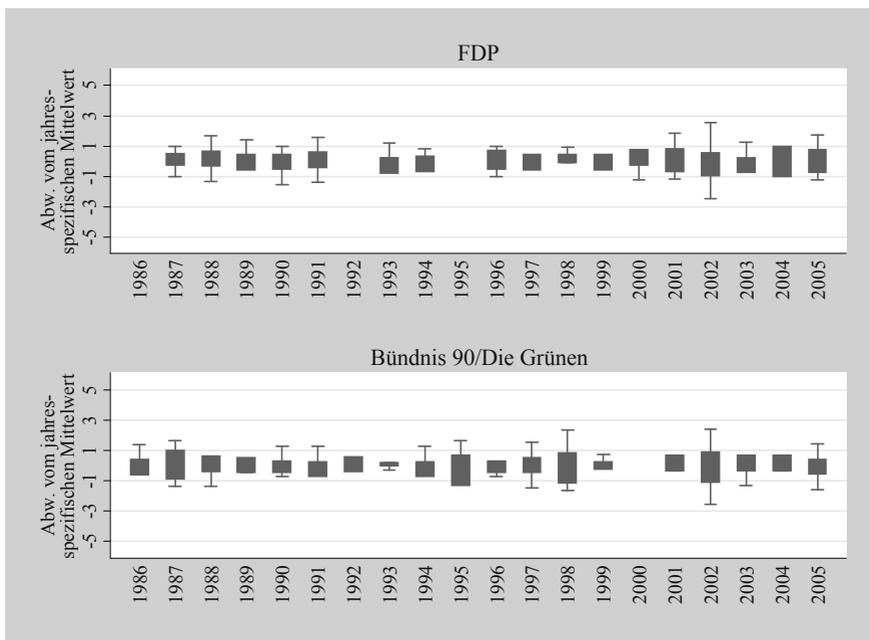


Anmerkungen: Eigene Berechnungen und Darstellung auf Grundlage der recherchierten Politbarometer-Projektionen. Abgebildet sind Box-Plots, wobei die grauen Kästen jeweils das 25- bzw.-75-Prozent-Quartil repräsentieren. Die Whiskers geben jeweils das Quartil plus den 1,5-fachen Interquartilsabstand an. Etwaige Ausreißer wurden bei der Berechnung berücksichtigt, jedoch nicht dargestellt. Die Nulllinie repräsentiert den jahresspezifischen Mittelwert der in den Rohdaten ermittelten Stimmenanteile nach Partei.

Entgegen der Vermutung zeigt sich, dass auch hinsichtlich der Projektionen für keine Partei ein eindeutiger Trend hin zu steigender Volatilität zu erkennen ist. Auffällig sind die Lücken für einzelne Jahre bei SPD, FDP, Bündnis 90/Die Grünen sowie – besonders häufig – der PDS/Die Linke. Dies resultiert daraus, dass im betreffenden Jahr die Projektion des Stimmenanteils für die jeweilige

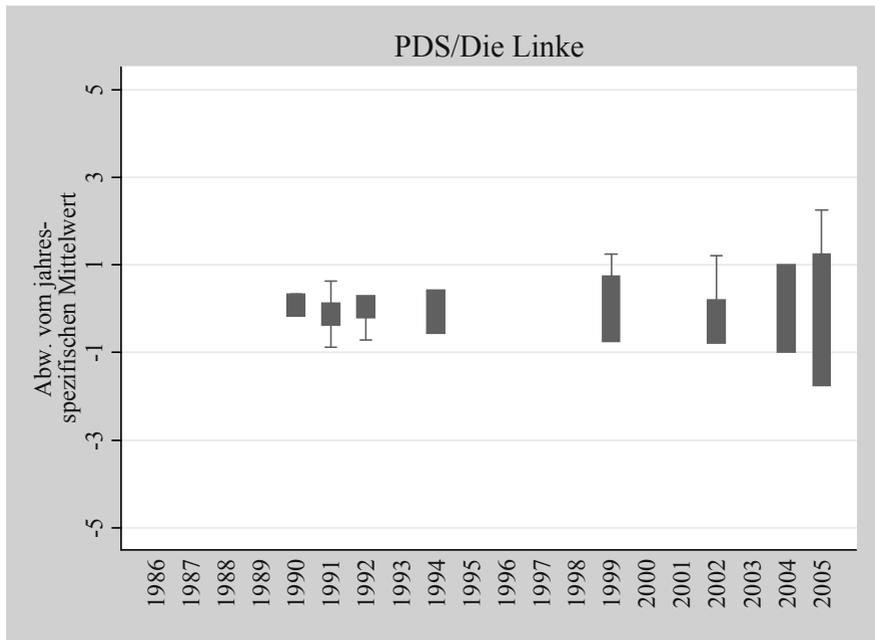
Partei keine Schwankung aufweist. Abbildung 3.5 ist dies ebenfalls zu entnehmen. Hier erkennt man, dass für die PDS/Die Linke von 1995 bis 1998 durchgehend vier Prozent Stimmenanteil von der Forschungsgruppe Wahlen projiziert wurde.

Abbildung 3.7: Volatilität der Stimmung bei der FDP und Bündnis 90/Die Grünen nach Jahren, 1986-2005, Politbarometer-Projektionen



Anmerkungen: Eigene Berechnungen und Darstellung auf Grundlage der recherchierten Politbarometer-Projektionen. Abgebildet sind Box-Plots, wobei die grauen Kästen jeweils das 25- bzw-75-Prozent-Quartil repräsentieren. Die Whiskers geben jeweils das Quartil plus den 1,5-fachen Interquartilsabstand an. Etwaige Ausreißer wurden bei der Berechnung berücksichtigt, jedoch nicht dargestellt. Die Nulllinie repräsentiert den jahresspezifischen Mittelwert der in den Rohdaten ermittelten Stimmenanteile nach Partei.

Abbildung 3.8: Volatilität der Stimmung bei der PDS/Die Linke nach Jahren, 1990-2005, Politbarometer-Projektionen



Anmerkungen: Eigene Berechnungen und Darstellung auf Grundlage der recherchierten Politbarometer-Projektionen. Abgebildet sind Box-Plots, wobei die grauen Kästen jeweils das 25- bzw. 75-Perzent-Quartil repräsentieren. Die Whiskers geben jeweils das Quartil plus den 1,5-fachen Interquartilsabstand an. Etwaige Ausreißer wurden bei der Berechnung berücksichtigt, jedoch nicht dargestellt. Die Nulllinie repräsentiert den jahresspezifischen Mittelwert der in den Rohdaten ermittelten Stimmenanteile nach Partei.

Schließlich werden auch die Zeitreihen aus Abbildung 3.5 mit zeitreihenanalytischen Verfahren näher untersucht. Dabei wird analog zu den Zeitreihen auf Grundlage der Rohdaten vorgegangen. Zunächst ist auffällig, dass die Autokorrelationen für alle Zeitreihen im Fall der Projektionen für die kleinen Parteien darauf hinweisen, dass keine Stationarität vorliegt (nicht dargestellt). Die graphische Analyse der partiellen Autokorrelationsplots zeigt für die CDU/CSU sowie die SPD, dass zwei *lags* berücksichtigt werden müssen. Für die übrigen Parteien ist jeweils nur ein *lag* relevant (nicht dargestellt). Wiederum wird diese graphische Analyse mit Dickey-Fuller-Tests überprüft, wobei jeweils eine Konstante ungleich

Null, die zuvor ermittelte Anzahl an *lags* sowie gegebenenfalls eine Trendkomponente berücksichtigt wurde. Die Tests bestätigen die Befunde: Die Zeitreihen der kleinen Parteien müssen einmal differenziert werden, um Stationarität herzustellen, und auch die *lag*-Struktur entspricht den Vermutungen. Trends weisen die im Weiteren betrachteten stationären Zeitreihen nicht auf (nicht dargestellt).

Tabelle 3.6: Ergebnisse der ARIMA-Modelle für die Stimmenanteile von CDU/CSU, SPD, FDP, Bündnis 90/Die Grünen und PDS/Die Linke, 1986-2005, Politbarometer-Projektionen

	CDU/CSU	SPD	FDP (erstes Differenzial)	Bündnis 90/ Die Grünen (erstes Differenzial)	PDS/Die Linke (erstes Differenzial)
AR(1)	1,157*** (0,047)	1,166*** (0,057)	-0,130* (0,052)	-0,172** (0,060)	-0,099** (0,041)
AR(2)	-0,237*** (0,046)	-0,222*** (0,054)			
Konstante	40,989***	37,105***	0,008 (0,036)	0,009 (0,036)	0,033 (0,033)
Wald Chi ²	1.674,33***	2.700,96***	6,19*	8,20**	5,73*
BIC	773,604	765,798	465,714	487,219	210,446
N	238	238	237	237	182

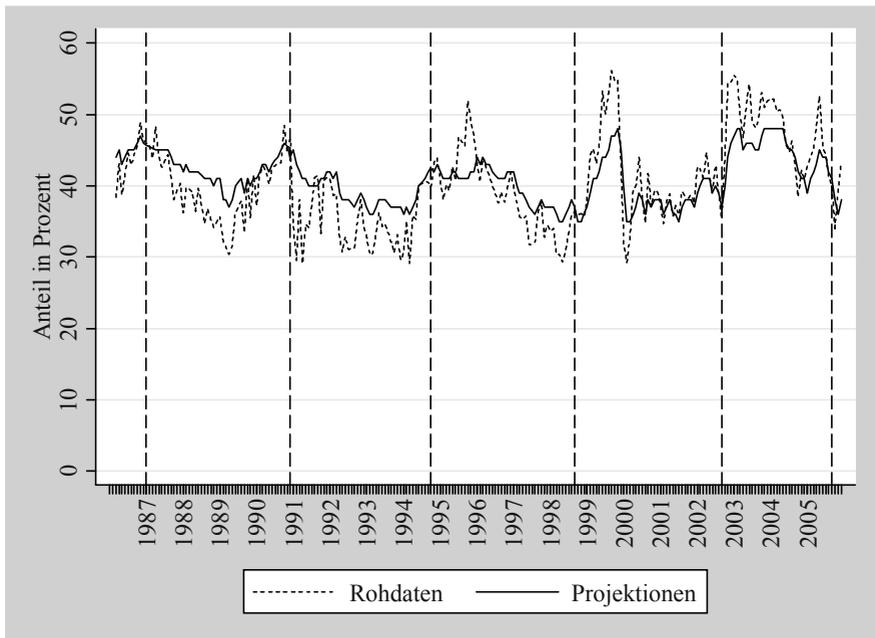
Anmerkungen: Dargestellt sind die unstandardisierten Koeffizienten sowie die Standardfehler (in Klammern) von ARIMA-Modellen mit der jeweils in der Tabelle angegebenen Anzahl an autoregressiven Termen; die Zeitreihen für die FDP, Bündnis 90/Die Grünen sowie PDS/Die Linke wurde einmal differenziert, um Stationarität herzustellen; *** $p \leq 0,001$, ** $p \leq 0,01$, * $p \leq 0,05$.

Auf Grundlage dieser Ergebnisse werden entsprechend spezifizierte ARIMA-Modelle zur Beschreibung der einzelnen Zeitreihen geschätzt. Dabei wurde darauf geachtet, dass die in Tabelle 3.6 dargestellten Modelle jeweils das angemessenste Modell gemäß des Bayesianischen Informationskriteriums (BIC) abbilden. Interessant sind die hier abzulesenden Differenzen zu Tabelle 3.3 mit den Ergebnissen der ARIMA-Modelle auf Grundlage der Politbarometer-Rohdaten. Während dort nur bei Bündnis 90/Die Grünen auf eine Zunahme des Stimmenanteils in der Tendenz eine Abnahme folgte, ist dieser Regressionseffekt bei den Projektionen auch bei der FDP sowie der PDS/Die Linke festzustellen. Bei den großen Parteien

CDU/CSU und SPD bleibt der signifikant positive AR(1)-Effekt bestehen, jedoch zeigen sich – anders als bei den Rohdaten – signifikant negative AR(2)-Effekte.

Für die Rohdaten wurde gezeigt, dass mit zeitlicher Nähe zur Wahl der Abstand der gemessenen Stimmenanteile von CDU/CSU und SPD abnimmt (Tabelle 3.4). Bei den Projektionen lässt sich ein entsprechender Zusammenhang aus Abbildung 3.5 nicht ablesen. Dies deutet darauf hin, dass die politische Gewichtung insbesondere in Zwischenwahlzeiten deutlich ausfällt. Eine – analog zu der in Tabelle 3.4 dargestellten – OLS-Regression zeigt auch keinen Zusammenhang mehr an zwischen dem zeitlichen Abstand zur Wahl und der Differenz der Stimmenanteile zwischen CDU/CSU und SPD.

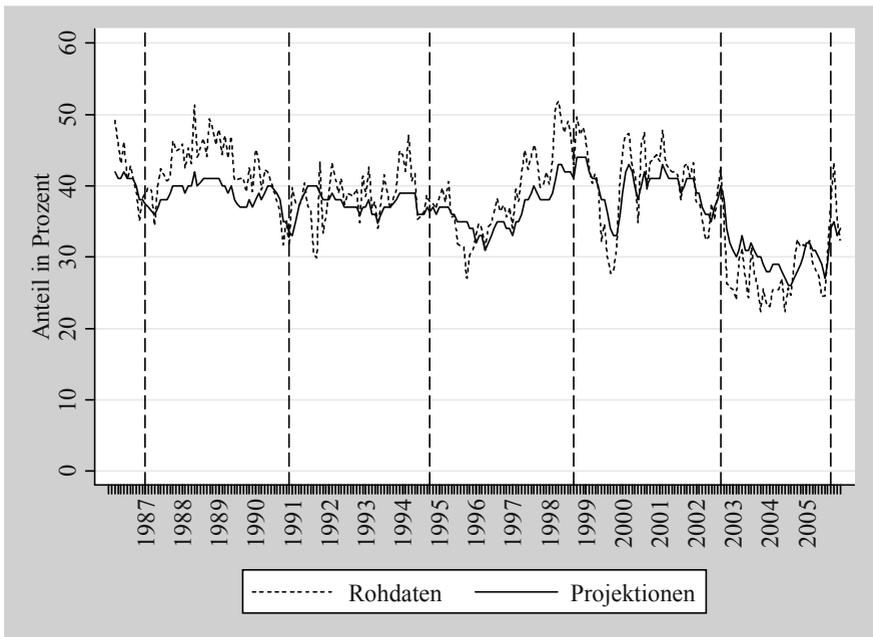
Abbildung 3.9: Differenzen der Rohdaten und der Projektionen der Politbarometer-Sonntagsfragen 1986-2005, CDU/CSU



Anmerkungen: Eigene Darstellung auf Grundlage der Politbarometer-Rohdaten sowie der veröffentlichten Politbarometer-Projektionen. Bis einschließlich März 1990 beziehen sich die dargestellten Werte auf die alten Bundesländer, ab April 1990 auf das gesamte Bundesgebiet. Die vertikal gestrichelten Linien sind die Zeitpunkte der Bundestagswahlen im betrachteten Zeitraum.

Nun wird parteispezifisch die aus den Rohdaten ermittelte Stimmung mit den Projektionen miteinander verglichen. Hierbei zeigen sich für die CDU/CSU (Abbildung 3.9) sowie die SPD (Abbildung 3.10) teilweise eine erhebliche Glättung der Zeitreihen insbesondere, so die Vermutung anhand der Darstellungen, in wahlfernen Zeiten. Für die CDU/CSU beträgt die durchschnittliche Differenz von Rohdaten und Projektionen $-0,8$ Prozentpunkte mit einer Standardabweichung von $4,0$. Die maximale absolute Abweichung liegt bei $13,4$ Prozentpunkten im Februar 1991. Für die SPD stellt sich dies ähnlich dar. Die mittlere Differenz beträgt $1,3$ Prozentpunkte mit einer Standardabweichung von $3,7$ und die maximale absolute Abweichung beträgt $10,1$ Prozentpunkte (September 1991 sowie März 1998).

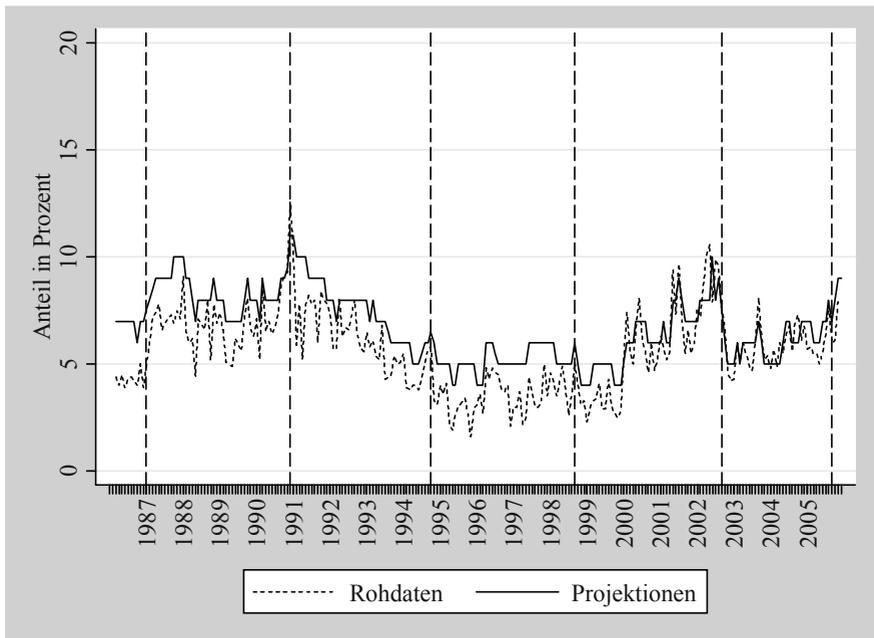
Abbildung 3.10: Differenzen der Rohdaten und der Projektionen der Politbarometer-Sonntagsfragen 1986-2005, SPD



Anmerkungen: Eigene Darstellung auf Grundlage der Politbarometer-Rohdaten sowie der veröffentlichten Politbarometer-Projektionen. Bis einschließlich März 1990 beziehen sich die dargestellten Werte auf die alten Bundesländer, ab April 1990 auf das gesamte Bundesgebiet. Die vertikal gestrichelten Linien sind die Zeitpunkte der Bundestagswahlen im betrachteten Zeitraum.

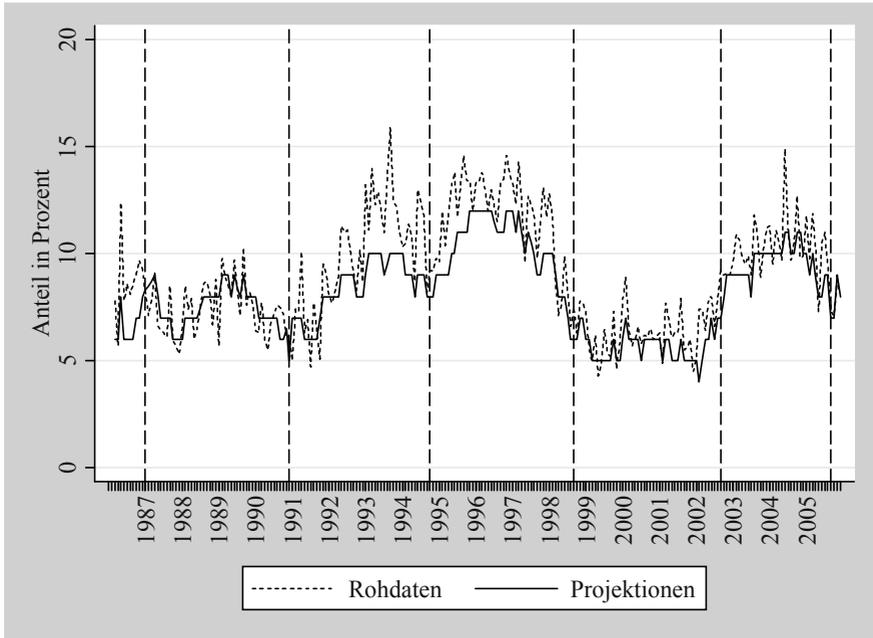
Mit Blick auf die kleinen Parteien ergibt sich folgendes Bild: Wie bereits erwähnt zeigt sich, dass für die FDP in der Projektion fast immer höhere Stimmenanteile berichtet werden als in den Rohdaten ermittelt (Abbildung 3.11). Die durchschnittliche Differenz zwischen Rohdaten und Projektionen über alle betrachteten Jahre hinweg beträgt dabei -1,2 Prozentpunkte mit einer Standardabweichung von 1,1. Hier zeigt sich der – relativ gesehen zum durchschnittlich ermittelten Stimmenanteil der FDP von 5,6 Prozent (Rohdaten) beziehungsweise 6,8 Prozent (Projektionen) – beachtliche Einfluss der politischen Gewichtung besonders deutlich. Bei der FDP beträgt die maximale absolute Differenz 4,8 Prozentpunkte, beobachtet im April 1991.

Abbildung 3.11: Differenzen der Rohdaten und der Projektionen der Politbarometer-Sonntagsfragen 1986-2005, FDP



Anmerkungen: Eigene Darstellung auf Grundlage der Politbarometer-Rohdaten sowie der veröffentlichten Politbarometer-Projektionen. Bis einschließlich März 1990 beziehen sich die dargestellten Werte auf die alten Bundesländer, ab April 1990 auf das gesamte Bundesgebiet. Die vertikal gestrichelten Linien sind die Zeitpunkte der Bundestagswahlen im betrachteten Zeitraum.

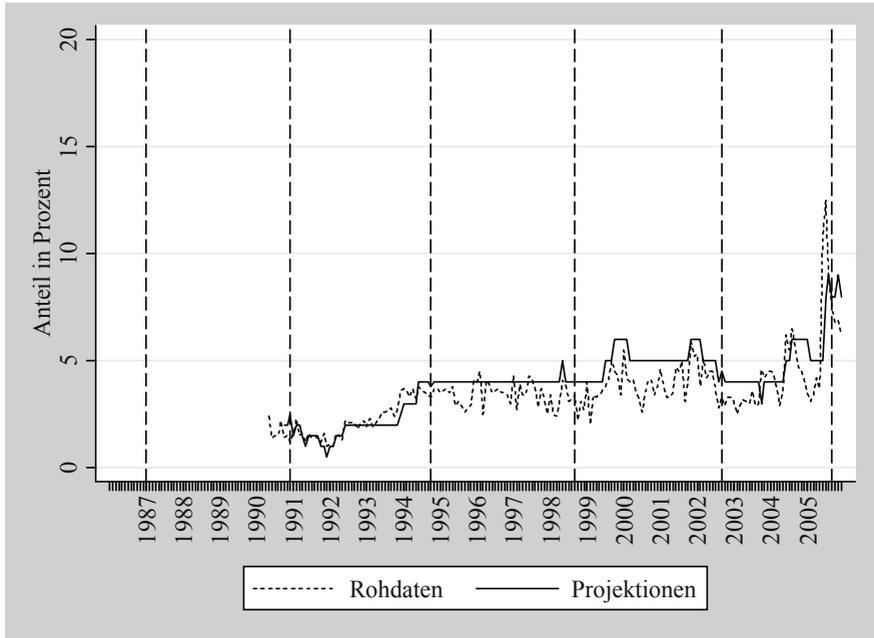
Abbildung 3.12: Differenzen der Rohdaten und der Projektionen der Politbarometer-Sonntagsfragen 1986-2005, Bündnis 90/Die Grünen



Anmerkungen: Eigene Darstellung auf Grundlage der Politbarometer-Rohdaten sowie der veröffentlichten Politbarometer-Projektionen. Bis einschließlich März 1990 beziehen sich die dargestellten Werte auf die alten Bundesländer, ab April 1990 auf das gesamte Bundesgebiet. Die vertikal gestrichelten Linien sind die Zeitpunkte der Bundestagswahlen im betrachteten Zeitraum.

Nun wird parteispezifisch die aus den Rohdaten ermittelte Stimmung mit den Projektionen miteinander verglichen. Hierbei zeigen sich für die CDU/CSU (Abbildung 3.9) sowie die SPD (Abbildung 3.10) teilweise eine erhebliche Glättung der Zeitreihen insbesondere, so die Vermutung anhand der Darstellungen, in wahlfernen Zeiten.

Abbildung 3.13: Differenzen der Rohdaten und der Projektionen der Politbarometer-Sonntagsfragen 1990-2005, PDS/Die Linke

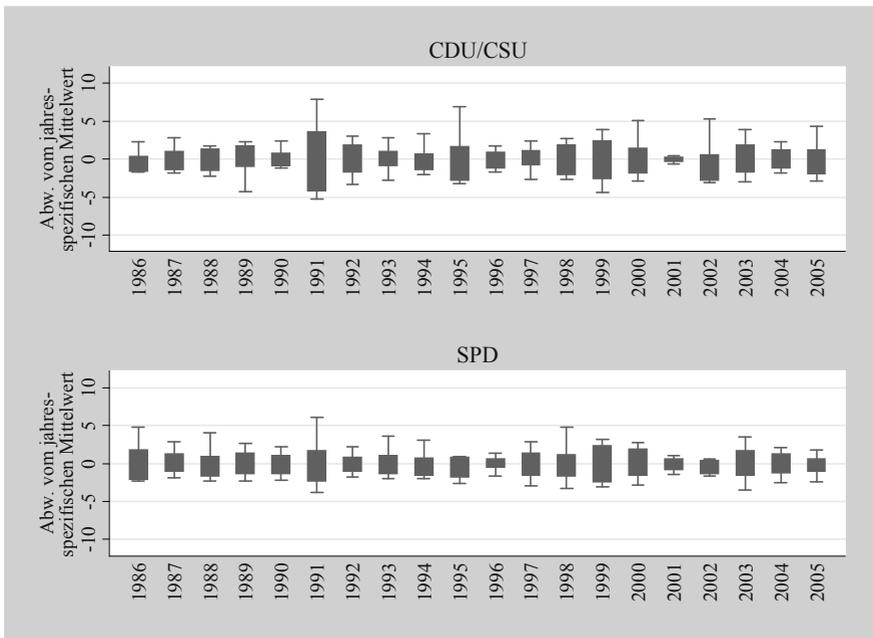


Anmerkungen: Eigene Darstellung auf Grundlage der Politbarometer-Rohdaten sowie der veröffentlichten Politbarometer-Projektionen. Die vertikal gestrichelten Linien sind die Zeitpunkte der Bundstagswahlen im betrachteten Zeitraum.

Für die PDS/Die Linke sind in den ersten Jahren nach der Wiedervereinigung nur geringe Abweichungen zwischen Rohdaten und Projektionen festzustellen. Diese vergrößern sich jedoch etwa Mitte der 1990er Jahre und gleichen sich dem Niveau der anderen Parteien an. Im Schnitt beträgt die Differenz über den gesamten Zeitraum -0,5 Prozentpunkte mit einer Standardabweichung von 1,0. Die größte absolute maximale Abweichung ist – wie vermutet – im Vergleich zu den übrigen Parteien in jüngerer Zeit zu beobachten, nämlich im Juni 2005 mit 5,8 Prozentpunkten. Dies ist jedoch als Ausreißer zu werten, da nur noch im Folgemonat Juli mit 4,5 Prozentpunkten eine ähnlich hohe Abweichung zu beobachten war – alle anderen Abweichungen liegen unter 2,2 Prozentpunkten.

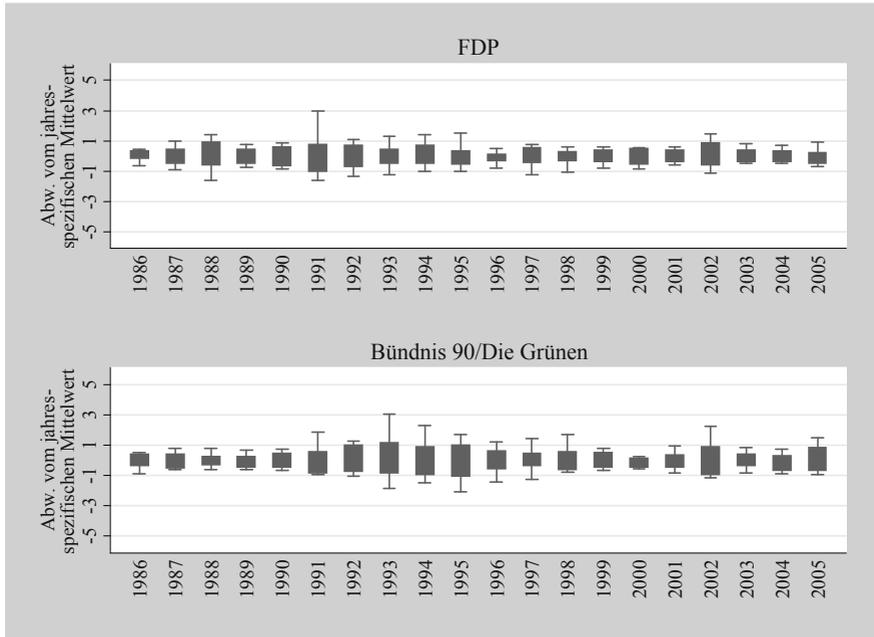
Bildet man aus den in den Abbildungen 3.9 bis 3.13 dargestellten Zeitreihen jeweils die Differenzen, entstehen für jede Partei neue Zeitreihen, welche mit entsprechenden statistischen Instrumentarien analysiert werden können. Damit wird untersucht, ob sich im Zeitverlauf Änderungen hinsichtlich des Ausmaßes der politischen Gewichtung feststellen lassen.

Abbildung 3.14: Volatilität der absoluten Differenzen zwischen Politbarometer-Rohdaten und -Projektionen bei der CDU/CSU und der SPD nach Jahren, 1986-2005



Anmerkungen: Eigene Berechnungen und Darstellung auf Grundlage der recherchierten Politbarometer-Projektionen sowie der Politbarometer-Rohdaten. Abgebildet sind Box-Plots, wobei die grauen Kästen jeweils das 25- bzw-75-Prozent-Quartil repräsentieren. Die Whiskers geben jeweils das Quartil plus den 1,5-fachen Interquartilsabstand an. Etwaige Ausreißer wurden bei der Berechnung berücksichtigt, jedoch nicht dargestellt. Die Nulllinie repräsentiert den jahresspezifischen Mittelwert der absoluten Abweichung der jeweiligen Stimmenanteile aus Rohdaten und Projektionen.

Abbildung 3.15: Volatilität der absoluten Differenzen zwischen Politbarometer-Rohdaten und -Projektionen bei der FDP und Bündnis 90/Die Grünen nach Jahren, 1986-2005

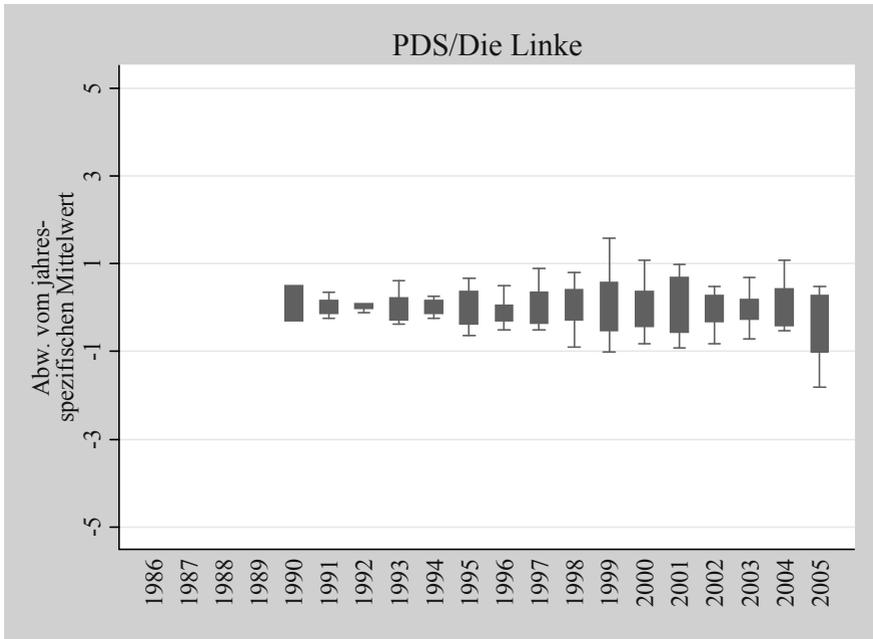


Anmerkungen: Eigene Berechnungen und Darstellung auf Grundlage der recherchierten Politbarometer-Projektionen sowie der Politbarometer-Rohdaten. Abgebildet sind Box-Plots, wobei die grauen Kästen jeweils das 25- bzw-75-Prozent-Quartil repräsentieren. Die Whiskers geben jeweils das Quartil plus den 1,5-fachen Interquartilsabstand an. Etwaige Ausreißer wurden bei der Berechnung berücksichtigt, jedoch nicht dargestellt. Die Nulllinie repräsentiert den jahresspezifischen Mittelwert der absoluten Abweichung der jeweiligen Stimmenanteile aus Rohdaten und Projektionen.

Gängige Auffassung – auch der Meinungsforschungsinstitute selbst – ist, dass Projektionen von Stimmenanteilen im Zeitverlauf schwieriger geworden sind (vgl. Anholz 2001: 102). Als Gründe hierfür werden bereits thematisierte Prozesse wie *partisan dealignment*, *last minute swings* oder auch eine Personalisierung der Politik angeführt (vgl. Abschnitt 3.1). Dementsprechend müsste man entweder erwarten, dass das Ausmaß der politischen Gewichtung im Zeitablauf zugenommen hat, oder aber die Projektionen in Bezug auf das Wahlergebnis zunehmend schlechter geworden sind. Die erste Frage wird im Folgenden

beantwortet, die Frage etwaiger Trends im Hinblick auf die Prognosegüte wird in Abschnitt 3.4.2 aufgegriffen.

Abbildung 3.16: Volatilität der absoluten Differenzen zwischen Politbarometer-Rohdaten und -Projektionen bei der PDS/Die Linke nach Jahren, 1990-2005



Anmerkungen: Eigene Berechnungen und Darstellung auf Grundlage der recherchierten Politbarometer-Projektionen sowie der Politbarometer-Rohdaten. Abgebildet sind Box-Plots, wobei die grauen Kästen jeweils das 25- bzw-75-Prozent-Quartil repräsentieren. Die Whiskers geben jeweils das Quartil plus den 1,5-fachen Interquartilsabstand an. Etwaige Ausreißer wurden bei der Berechnung berücksichtigt, jedoch nicht dargestellt. Die Nulllinie repräsentiert den jahresspezifischen Mittelwert der absoluten Abweichung der jeweiligen Stimmenanteile aus Rohdaten und Projektionen.

Betrachtet man zunächst die Volatilität der politischen Gewichtung im Zeitablauf, so ergeben sich die in Abbildungen 3.14 bis 3.16 dargestellten Ergebnisse. Demnach ist für keine Partei ein Trend im Hinblick auf eine zunehmende oder abnehmende Schwankung der absoluten Differenzen zwischen Rohdaten und

Projektionen im Zeitverlauf festzustellen. Damit ist noch nicht belegt, ob ein genereller Trend bezüglich des Ausmaßes der politischen Gewichtung besteht. Diese Befunde offenbaren jedoch keinen erkennbaren Trend der Streuung des Ausmaßes der politischen Gewichtung innerhalb eines Jahres im Zeitverlauf.

Anhand der graphischen Darstellungen der Differenzen-Zeitreihen, können leichte Trends im Zeitablauf für die CDU/CSU und die FDP ausgemacht werden. Dies bestätigen auch durchgeführte Dickey-Fuller-Tests. Allerdings muss hier eine starke Konfundierung mit einer Regierungsbeteiligung der jeweiligen Partei vermutet werden. So zeigt sich für die CDU/CSU, dass die Projektionen die in den Rohdaten ermittelten Stimmenanteile bis 1998 fast durchgehend nach oben gewichten, danach tendenziell nach unten. Im Betrachtungszeitraum von 1986 bis 1998 stellte die CDU/CSU mit Helmut Kohl den Kanzler, 1998 bis 2005 regierte die SPD mit Kanzler Gerhard Schröder. Für die SPD und auch in abgeschwächter Form für die FDP zeigen sich analoge Befunde.

Im Folgenden werden die absoluten Differenzen betrachtet, um zu vermeiden, dass sich mögliche positive und negative Abweichungen gegenseitig aufheben. Dies ist vor allem im Hinblick auf den Effekt der Regierungsbeteiligung von Parteien zu sehen: Liegt diese vor, werden tendenziell die Stimmenanteile eher nach oben gewichtet. Bei Oppositionsparteien werden die Stimmenanteile hingegen im Durchschnitt nach unten gewichtet. Die zeitreihenspezifischen Analysen zeigen, dass alle Zeitreihen stationär sind und eine AR(1)-Struktur aufweisen. Weiterhin zeigen die Dickey-Fuller-Tests für die FDP einen leichten, jedoch signifikanten Abwärtstrend und für die PDS/Die Linke einen signifikanten Aufwärtstrend (nicht dargestellt). Dieser bleibt im Übrigen bestehen, auch wenn man die beiden Ausreißer (Juni und Juli 2005) aus der Analyse ausschließt.

Analog zum bisherigen Vorgehen wurden entsprechend spezifizierte ARIMA-Modelle berechnet, deren Ergebnisse in Tabelle 3.7 abgebildet sind. Ergebnis dieser Analysen ist zunächst, dass kein eindeutiger Trend hinsichtlich der Bedeutungszunahme der politischen Gewichtung im Zeitverlauf zu erkennen ist. Insbesondere für die beiden großen Parteien ist dies nicht zu konstatieren. Allerdings zeigt sich für die FDP und die PDS/Die Linke eine Zunahme des Ausmaßes der absoluten Abweichungen der Rohdaten von den Projektionen.

Tabelle 3.7: Ergebnisse der ARIMA-Modelle für die absoluten Differenzen der Stimmenanteile von CDU/CSU, SPD, FDP, Bündnis 90/Die Grünen und PDS/Die Linke von Politbarometer-Rohdaten und -Projektionen, 1986-2005

	CDU/CSU	SPD	FDP	Bündnis 90/ Die Grünen	PDS/Die Linke
AR(1)	0,517*** (0,053)	0,421*** (0,064)	0,893*** (0,053)	0,323*** (0,052)	0,979*** (0,044)
MA(1)			-0,636*** (0,079)		-0,633*** (0,065)
MA(2)					-0,218*** (0,058)
Konstante	3,250*** (0,347)	3,189*** (0,263)	1,392*** (0,174)	1,220*** (0,116)	0,877 (0,411)
Wald Chi ²	94,65***	42,67***	389,66***	38,78***	1.102,21***
BIC	1.061,447	1.022,120	555,337	668,039	381,419
N	238	238	238	238	183

Anmerkungen: Dargestellt sind die unstandardisierten Koeffizienten sowie die Standardfehler (in Klammern) von ARIMA-Modellen mit der jeweils in der Tabelle angegebenen Anzahl an autoregressiven und moving-average-Termen; *** $p \leq 0,001$, ** $p \leq 0,01$, * $p \leq 0,05$.

Im Hinblick auf die PDS/Die Linke sollte dieses Ergebnis nicht überinterpretiert werden, auch wenn der Befund robust ist gegenüber den beiden diagnostizierten Ausreißern. Zum einen liegen hier deutlich weniger Datenpunkte vor und zum anderen ist durch die Fusion der PDS mit der WASG eine neue Partei entstanden. Dadurch erscheint die Fortschreibung der PDS-Stimmenanteile zumindest fraglich. In Bezug auf die FDP kann vermutet werden, dass die Zunahme der absoluten Abweichungen im Zeitverlauf eine zunehmende Rationalisierung des Wählerverhaltens widerspiegelt. Damit ist der steigende Anteil strategischer Wähler beziehungsweise Stimmensplitter gemeint, wovon insbesondere die FDP hinsichtlich der Zweitstimmen profitiert. Weiterhin ist bei allen Parteien der deutliche positive Effekt der AR(1)-Komponente zu erkennen, der angibt, dass die absoluten Abweichungen gleichsinnig mit den absoluten Abweichungen des Vormonats zusammenhängen.

Der graphischen Analyse folgend, kann vermutet werden, dass die Differenzen zwischen Rohdaten und Projektionen mit dem Wahltermin zusammenhängen. Dies wird nun noch näher untersucht. Hierzu werden die in Tabelle 3.7 geschätzten ARIMA-Modelle um den zeitlichen Abstand zur Wahl in Monaten ergänzt (Tabelle 3.8).

Tabelle 3.8: Einfluss des zeitlichen Abstands zu einer Bundestagswahl auf die absoluten Differenzen zwischen den jeweiligen Stimmenanteilen aus Politbarometer-Rohdaten und -Projektionen nach Parteien, 1986-2005

	CDU/CSU	SPD	FDP	Bündnis 90/ Die Grünen	PDS/Die Linke
Zeitl. Abstand zur Wahl in Monaten	0,042** (0,016)	0,022 (0,014)	0,009 (0,005)	-0,011 (0,006)	-0,001 (0,006)
AR(1)	0,476*** (0,055)	0,420*** (0,063)	0,892*** (0,054)	0,323*** (0,052)	0,979*** (0,045)
MA(1)			-0,633*** (0,080)		-0,634*** (0,065)
MA(2)					-0,218*** (0,058)
Konstante	2,336*** (0,513)	2,706*** (0,421)	1,199*** (0,211)	1,462*** (0,170)	0,893* (0,435)
Wald Chi ²	83,32***	44,77***	380,15***	36,20***	1.088,82***
BIC	1.060,706	1.025,114	558,032	670,133	386,607
N	238	238	238	238	183

Anmerkungen: Dargestellt sind die unstandardisierten Koeffizienten sowie die Standardfehler (in Klammern) von ARIMA-Modellen mit der jeweils in der Tabelle angegebenen Anzahl an autoregressiven und moving-average-Termen; *** $p \leq 0,001$, ** $p \leq 0,01$, * $p \leq 0,05$.

Als Ergebnis dieser Analysen lässt sich festhalten, dass der Einfluss der politischen Gewichtung nur bei der CDU/CSU signifikant mit dem zeitlichen Abstand zur Wahl variiert (Tabelle 3.8). Hier zeigt sich, dass mit jedem Monat näher an der Wahl die absolute Differenz zwischen den aus den Rohdaten ermittelten

Stimmenanteilen und den Projektionsstimmenanteilen abnimmt. In der Tendenz findet man dies auch für die SPD ($p = 0,105$) sowie die FDP ($p = 0,096$). Insgesamt kann daraus jedoch gefolgert werden, dass – mit Ausnahme für die CDU/CSU – das Ausmaß der politischen Gewichtung nicht mit dem zeitlichen Abstand zur Wahl variiert.

Zusammenfassend ist aus diesen Analysen *erstens* festzuhalten, dass ein bedeutender Einfluss der politischen Gewichtung festgestellt werden konnte. *Zweitens* ist hervorzuheben, dass weder in Bezug auf die ermittelten Stimmenanteile aus den Rohdaten, noch aus den Projektionen für die einzelnen Parteien deutliche Trendentwicklungen abzulesen sind. Einzig für die PDS/Die Linke ergibt sich ein stabiler Aufwärtstrend, welcher aufgrund der geringen vorliegenden Datenpunkte und der Fusion der WASG und der PDS zu Die Linke nicht überbetont werden sollte. Die durchgeführten Analysen zeigen *drittens*, dass in Bezug auf die politische Gewichtung ebenfalls keine einheitlichen Trends vorliegen. So ist im Zeitablauf weder eine zunehmende jahresspezifische Volatilität des Gewichtungseinflusses festzustellen, noch ein genereller Trend einer Relevanzzunahme der politischen Gewichtung. Letzteres zeigte sich nur für die FDP sowie für die PDS/Die Linke. Dieser Befund sollte hinsichtlich der PDS/Die Linke aufgrund der angesprochenen Einschränkungen vorsichtig interpretiert werden. Für die FDP ist dieses Ergebnis möglicherweise als Reaktion auf die zunehmende Zahl von Wählern zu interpretieren, die ihre Stimmen splitten und damit überproportional häufig ihre Zweitstimme am Wahltag der FDP bei vor allem vorliegender CDU/CSU-Präferenz vergeben. *Viertens* wurde schließlich herausgearbeitet, dass es nur für die CDU/CSU ein bedeutender Zusammenhang zwischen dem zeitlichen Abstand zu einer Bundestagswahl und dem Ausmaß der politischen Gewichtung besteht.

Im Folgenden werden die Prognosegüte der Sonntagsfragen sowie mögliche Einflussfaktoren hierauf untersucht. Die bisherigen Analysen zum Ausmaß der politischen Gewichtung anhand der Politbarometer-Daten zeigen, dass die mit den Rohdaten ermittelten Stimmenanteile teilweise erheblich von den berichteten Stimmenanteilen der Projektionen abweichen. Für die weiteren Analysen bedeutet dies zweierlei: Zum einen zeigt sich die Unerlässlichkeit der Berücksichtigung von *house effects*, da die jeweiligen Gewichtungsverfahren vermutlich institutspezifisch variieren und nicht bekannt sind. Zum anderen wird die Identifikation von weiteren Einflussfaktoren auf die Prognosegüte erschwert, da aufgrund des

Ausmaßes der politischen Gewichtung entsprechende Effekte erheblich sein müssen, um aufgedeckt werden zu können.

Da im Weiteren nur diejenigen Wahlabsichtsumfragen berücksichtigt werden, welche in den zwölf Monaten vor einer Bundestagswahl veröffentlicht wurden, sind in Tabelle 3.9 die absoluten mittleren Differenzen zwischen Rohdaten und Projektionen nur für diesen Zeitraum vor einer Wahl dargestellt, um auch einen quantitativen Eindruck vom Ausmaß der politischen Gewichtung zu erhalten.

Tabelle 3.9: Mittlere Differenzen zwischen den Stimmenanteilen ermittelt aus den Politbarometer-Rohdaten und den -Projektionen nach Parteien bei Betrachtung von zwölf Monaten vor einer Bundestagswahl, 1986-2005

	CDU/CSU	SPD	FDP	Bündnis 90/ Die Grünen	PDS/ Die Linke
Mittelwert	-1,0	1,8	-1,1	1,1	-0,5
Standardabweichung	2,8	3,1	1,2	1,3	1,4
Minimum	-7,4	-4,5	-3,1	-1,6	-2,5
Maximum	7,6	10,1	2,6	4,4	5,8
N	75	75	75	75	54

Anmerkung: Eigene Darstellung auf Grundlage der Politbarometer-Rohdaten sowie der veröffentlichten Politbarometer-Projektionen.

3.4.2 Die Prognosegüte von Wahlabsichtsumfragen

In diesem Abschnitt wird auf die Prognosegüte von Wahlabsichtsumfragen eingegangen. Diese wird anhand der in Abschnitt 3.3.2 für die Deskription als sinnvoll erachteten Maßzahlen – mittlere absolute und mittlere absolute prozentuale Abweichung, das auf dem Herfindahl-Index beruhende Maß ν sowie das A -Maß – dargestellt. In Anschluss an Antholz (2001) wird erstmals eine systematische Bestandsaufnahme der Güte von Sonntagsfragen für alle Bundestagswahlen von 1949 bis 2009 vorgelegt. Im Rahmen dieser deskriptiven Betrachtung werden

ausschließlich Wahlabsichtsfragen, die sich auf das gesamte Bundesgebiet beziehen, berücksichtigt.

Hierbei werden neben einer Darstellung der Prognosegüte bereits weitergehende Fragen beantwortet, etwa nach möglichen Trends im Zeitablauf und den häufig kolportierten parteispezifischen Verzerrungen einzelner Meinungsforschungsinstitute (vgl. zu Hinweisen hierzu bspw. Antholz 2001: 151ff.). Zunächst wird auf die generelle Prognosegüte von Sonntagsfragen über den gesamten Zeitraum eingegangen. Die Deskription wird dann weiter ausdifferenziert und die Prognosegüte im Zeitablauf, nach Region, nach Instituten und Parteien analysiert.

Die Sonntagsfragen werden auch in bundestagswahlfernen Zeiten im Rahmen der politischen Meinungsdauerbeobachtung wie beispielsweise durch das Politbarometer in regelmäßigen Abständen erhoben und publiziert. Diese Ergebnisse sind ebenfalls in dem für die Arbeit erfassten Datensatz enthalten. Da es sich bei Sonntagsfragenergebnissen, welche etwa kurz nach einer Bundestagswahl veröffentlicht werden, wohl kaum um eine Vorhersage des zukünftigen Bundestagswahlergebnisses handeln kann, wird hier die Auswahl beschränkt auf Wahlabsichtsumfragen, die maximal 365 Tage vor einer Bundestagswahl veröffentlicht wurden. Auch dieser Zeitraum mag zu lange erscheinen, doch eine zentrale Einflussvariable, so die Vermutung, stellt der zeitliche Abstand zur Wahl dar. Um diese Hypothese angemessen prüfen zu können, erscheint es sinnvoll, einen nicht zu kurzen Zeitraum vor einer Wahl zu betrachten. In Bezug auf die Ausdehnung des Betrachtungszeitraums grenzt sich diese Arbeit im Übrigen von bisherigen wesentlich ab, da bisher immer deutlich kürzere Zeitabschnitte vor einer Wahl betrachtet wurden (vgl. Abschnitt 2.1.6). Auch um Vergleiche mit anderen Prognoseansätzen herstellen zu können, ist eine zeitliche Erweiterung zwingend, denn Prognosen aus theoretischen Ansätzen werden in der Regel bereits mehrere Monate vor einer Wahl aufgestellt. Ein fairer Vergleich verschiedener Ansätze muss entsprechend den *lead* der jeweiligen Vorhersagen berücksichtigen.

In Bezug auf die tagesgenaue Erfassung der Publikation der Wahlabsichtsumfragen sind jedoch einige Anmerkungen notwendig. Es ist üblich, dass die Ergebnisse politischer Meinungsumfragen in verschiedenen Medien zu unterschiedlichen Zeitpunkten veröffentlicht werden. Bei der Datensammlung wurde versucht, sicherzustellen, immer das Erstveröffentlichungsdatum der jeweiligen Sonntagsfrage zu ermitteln, um den Vorhersagezeitpunkt möglichst exakt wiederzugeben. Weiterhin ist anzumerken, dass zu 493 beziehungsweise 13,7 Prozent

der Sonntagsfragen kein Veröffentlichungstag ermittelt werden konnte. Für diese Fälle wurde ein Wert von 15, also zur Mitte des Monats, eingesetzt, um auch diese Fälle weiterhin nutzen zu können.

Da vermutet wird, dass diese fehlenden Werte eine selektive Auswahl der Sonntagsfragen darstellen, werden im Folgenden Analysen mit und ohne die imputierten Werte durchgeführt. Im Hinblick auf die zeitliche Einschränkung auf ein Jahr vor der Wahl ergibt sich hieraus kein Problem, denn der Veröffentlichungsmonat ist in allen Fällen bekannt.

Im Hinblick auf die Berechnung der mittleren absoluten Abweichung (MAE) sowie der mittleren absoluten prozentualen Abweichung (MAPE) stellt sich die Frage, wie viele Parteien berücksichtigt werden sollen. In der Regel werden im Rahmen der Berichterstattung nur Stimmenanteile von Parteien berücksichtigt, denen eine realistische Chance zugesprochen wird, die Fünf-Prozent-Hürde zu überspringen. Hier werden verschiedene Vorgehensweisen gewählt. Zunächst werden MAE und MAPE unter Einschluss der sonstigen Parteien berechnet. Der Vorteil dieser Betrachtung wird darin gesehen, dass zum einen bei allen berücksichtigten Sonntagsfragenergebnissen die Basis nicht variiert, nämlich immer 100 Prozent der Stimmen abgedeckt werden.³⁷ Zum anderen ist der Stimmenanteil der sonstigen Parteien implizit gegeben und erscheint für die spätere Zusammensetzung des Bundestages nicht unbedeutend. Allerdings wird damit insbesondere der MAPE aufgrund seiner Eigenschaft der „Bestrafung“ von Abweichungen bei Parteien mit geringem Stimmenanteil unbotmäßig erhöht. Deshalb werden auch alternativ hierzu die Maße ohne Berücksichtigung der sonstigen Parteien und ohne Einbezug der Republikaner betrachtet.

Die mittlere absolute Abweichung (MAE), die mittlere absolute prozentuale Abweichung (MAPE) sowie der Vergleichsquotient ν der Verteilungen sollten jeweils nur miteinander in Beziehung gesetzt werden, sofern jeweils dieselbe Anzahl an Parteien die Berechnungsbasis stellt (vgl. Abschnitt 3.3.2). Da dies hier nicht der Fall ist, wird in Tabelle 3.10 neben der jeweiligen Gesamtprognosegüte nach der Anzahl der Parteien differenziert. Nicht berücksichtigt wird hierbei die Koinzidenz der Parteienanzahl und der Wahlperiode – die mögliche Veränderung der Vorhersagegüte im Zeitablauf wird jedoch noch detailliert untersucht.

37 Dies ist im Übrigen auch die Vorgehensweise zur Berechnung der Herfindahl-Indizes für die Vorhersagen beziehungsweise die tatsächlichen Stimmenanteile.

Tabelle 3.10: Prognosegüte der umfragebasierten Vorhersagen nach berücksichtigter Parteienanzahl, 1949-2009

Anzahl Parteien	Kennwerte	Mittlere absolute Abweichung (MAE)	Mittlere absolute prozentuale Abweichung (MAPE)	Auf dem Herfindahl-Index basierender Vergleichsquotient (v)
3	Mittelwert	3,07	11,32	0,004
	Std.-abw.	1,11	4,79	0,045
	N	6	6	6
4	Mittelwert	2,72	20,93	-0,003
	Std.-abw.	1,68	17,14	0,070
	N	120	120	120
5	Mittelwert	1,87	27,47	-0,053
	Std.-abw.	0,90	15,41	0,052
	N	62	62	62
6	Mittelwert	2,08	18,58	-0,038
	Std.-abw.	0,93	7,55	0,063
	N	750	750	750
7	Mittelwert	1,71	26,21	-0,034
	Std.-abw.	0,83	10,47	0,055
	N	132	132	132
Gesamt	Mittelwert	2,10	20,26	-0,034
	Std.-abw.	1,06	10,47	0,064
	N	1.070	1.070	1.070

Anmerkungen: Die Angaben beziehen sich auf alle erfassten Ergebnisse von Sonntagsfragen, welche maximal ein Jahr vor einer Bundestagswahl veröffentlicht wurden. Berücksichtigt wurden jeweils auch die Stimmenanteile sonstiger Parteien.

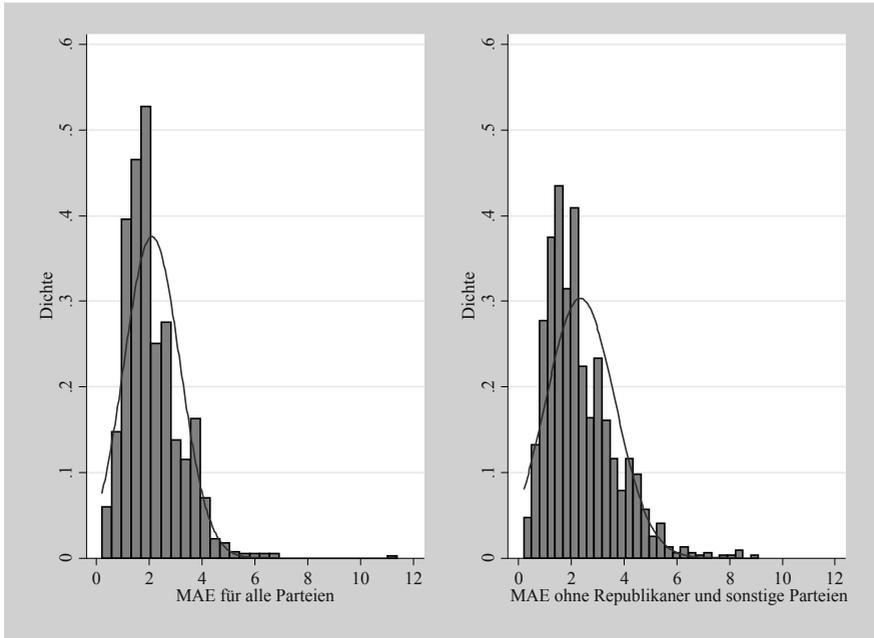
Insgesamt zeigt sich, dass die Sonntagsfragen im Schnitt für jede Partei um 2,1 Prozentpunkte mit einer Standardabweichung von 1,1 Prozentpunkten vom späteren Wahlergebnis abweichen. Der Unterschied zur ausschließlichen Betrachtung der nicht imputierten Werte fällt sehr gering aus. Hier beträgt der MAE ebenfalls 2,1 Prozentpunkte mit einer Standardabweichung von 1,0 Prozentpunkten. Die mittlere absolute prozentuale Abweichung über alle Vorhersagen beträgt 20,3 Prozent mit einer Standardabweichung von 10,5 Prozent. Der Vergleichsquotient v liegt mit einem Mittelwert von $-0,03$ nahe bei Null. Daraus lässt sich schließen, dass insgesamt die tatsächliche Stimmenanteilkonzentration durch die Vorhersagen gut approximiert wird. Das negative Vorzeichen gibt an, dass die Konzentration der Stimmenanteile durch die Vorhersagen etwas überschätzt wird.

Zur graphischen Veranschaulichung der Verteilung der mittleren absoluten Abweichungen sind in Abbildung 3.17 die Histogramme für den MAE über alle Parteien dargestellt (linke Seite) und für den MAE ohne Berücksichtigung der Republikaner und sonstiger Parteien (rechte Seite). Auffällig ist in der linken Grafik ein deutlicher Ausreißer. Der MAE von 11,4 Prozentpunkten bezieht sich auf die Vorhersage des Wahlergebnisses von 1949. Ob es sich hierbei um einen – gemessen an heutigen Standards der Umfrageforschung – legitimerweise zu berücksichtigenden Datenpunkt handelt, kann durchaus bezweifelt werden, zumal sich die Angaben noch auf die Erststimmen beziehen.³⁸ In den multivariaten Analysen wird die Vorhersage für die Wahl 1949 nicht berücksichtigt.

Abbildung 3.17 ist weiterhin zu entnehmen, dass die Verteilung der MAE leicht rechtsschief ist und somit mehr kleinere Abweichungen zu beobachten sind als man bei Normalverteilung erwarten würde. Dies trifft insbesondere auf die rechte Grafik zu, welche nur die bei der politischen Gewichtung der Meinungsforschungsinstitute berücksichtigten Parteien umfasst. Dies könnte man dahingehend interpretieren, dass die politische Gewichtung die Vorhersage verbessert, da man ansonsten normalverteilte Abweichungen erwarten würde.

38 Der Parlamentarische Rat konnte sich 1948/1949 nicht auf eine verfassungsrechtliche Festschreibung des Wahlsystems festlegen. Uneinigkeit bestand insbesondere darüber, ob ein Mehrheits- oder Verhältniswahlrecht eingeführt werden solle. Schließlich einigten sich die Ratsmitglieder darauf ein Bundeswahlgesetz zu verabschieden, welches nur für die erste Bundestagswahl am 14. August 1949 galt. Bei dieser Wahl konnte nur eine Stimme abgegeben werden (Nohlen 2007: 327; Woyke 2005: 49).

Abbildung 3.17: Histogramme für die mittlere absolute Abweichung der umfragebasierten Vorhersagen, 1949-2009

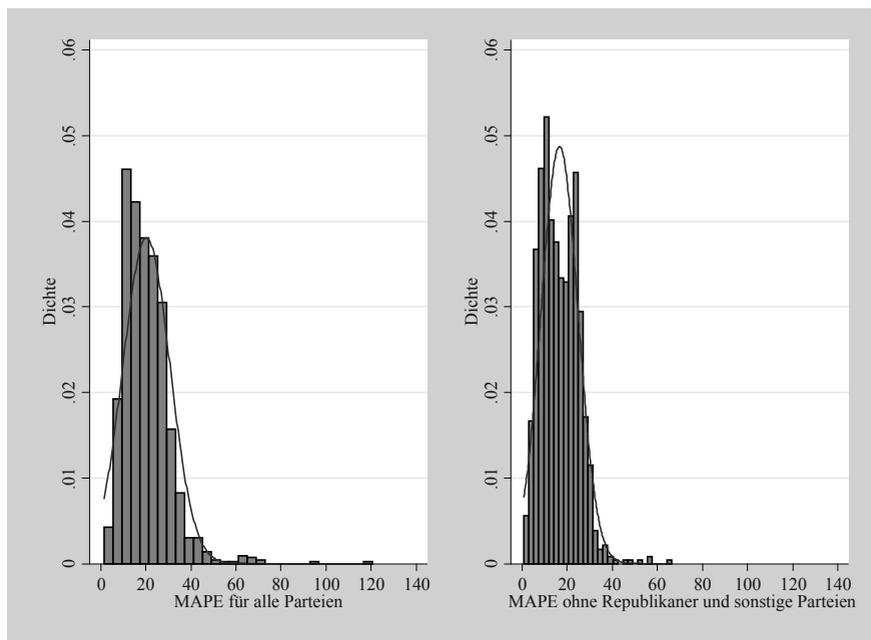


Anmerkungen: Die Darstellungen beziehen sich auf alle erfassten Ergebnisse von Sonntagsfragen, welche maximal ein Jahr vor einer Bundestagswahl veröffentlicht wurden.

Zieht man das skalenunabhängige Maß heran, so erhält man eine mittlere Abweichung der Vorhersagen vom tatsächlichen Wahlergebnis von 20,3 Prozent mit einer Standardabweichung von 10,5 Prozent. In diesem Fall sinkt der MAPE leicht, sofern man die imputierten Fälle ausschließt. Es ergibt sich ein Wert von 19,6 Prozent mit einer Standardabweichung von 9,2 Prozent. Damit sind – wie schon beim MAE – keine nennenswerten Abweichungen zwischen den beiden Berechnungsarten festzustellen. Dies gilt auch bei Differenzierung nach der Parteienanzahl, weshalb im Weiteren auf diese Differenzen nur eingegangen wird, sofern der zeitliche Abstand zur Wahl thematisiert wird.

Um die Verteilung zu visualisieren, sind in Abbildung 3.18 die Histogramme für die mittleren prozentualen Abweichungen einmal für alle Parteien (linke Seite) und ohne Berücksichtigung der Republikaner sowie sonstiger Parteien (rechte Seite) dargestellt.

Abbildung 3.18: Histogramme für die mittlere prozentuale Abweichung der umfragebasierten Vorhersagen, 1949-2009



Anmerkungen: Die Darstellungen beziehen sich auf alle erfassten Ergebnisse von Sonntagsfragen, welche maximal ein Jahr vor einer Bundestagswahl veröffentlicht wurden.

Wie bereits bei Analyse des MAE ist auch hier im linken Teil der Grafik ein deutlicher Ausreißer zu erkennen. Hierbei handelt es sich um eine Vorhersage des IfD Allensbach zur Bundestagswahl 1972. Diese weist insofern eine Besonderheit auf, als die sonstigen Parteien zusammen nur 0,9 Prozent der Stimmen erhielten. Nur 1983 erhielten die sonstigen Parteien mit 0,4 Prozent der Stimmen einen geringeren Anteil als 1972. Hinzu kommt, dass 1983 Bündnis 90/Die Grünen

erstmalig in den Bundestag einzogen und somit bei der Berechnung des Stimmenanteils sonstiger Parteien vier Parteien einzeln ausgewiesen wurden, während 1973 nur drei Parteien aufgenommen wurden.

Durch die beachtliche Abweichung der Vorhersage des Stimmenanteils sonstiger Parteien (5,0 Prozent) vom tatsächlichen Stimmenanteil (0,9 Prozent), ergibt sich die insgesamt hohe mittlere prozentuale Abweichung von 120,7 Prozent. Dies verdeutlicht einmal mehr den großen Einfluss auch eher geringer absoluter Abweichungen bei Parteien mit geringen Stimmenanteilen. Aus diesem Grund wird der MAPE noch differenzierter für einzelne Parteien betrachtet.

Berechnet man den MAPE nur für die beiden großen Parteien, so ergibt sich eine mittlere prozentuale Abweichung von 8,6 Prozent. Betrachtet man hingegen nur die kleinen Parteien FDP, Bündnis 90/Die Grünen, PDS/Die Linke und sonstige Parteien, dann springt der MAPE auf einen Wert von 26,8 Prozent (Tabelle 3.11).

Dies zeigt, dass offenbar die Stimmenanteile der kleinen Parteien relativ gesehen deutlich schlechter vorhergesagt werden als die Stimmenanteile von CDU/CSU und SPD. Daher ist bereits an dieser Stelle auch für empirische Evaluationen der Prognosegüte zumindest im Rahmen bundesdeutscher Wahlen festzuhalten, dass neben der üblicherweise ausschließlich verwendeten mittleren absoluten Abweichung auch der prozentuale Fehler berücksichtigt werden sollte. Dies gilt umso mehr vor dem Hintergrund der Bedeutung kleinerer Parteien für die späteren Mehrheitsverhältnisse im Bundestag und damit ihr Einfluss auf sich ergebende Koalitionsoptionen.

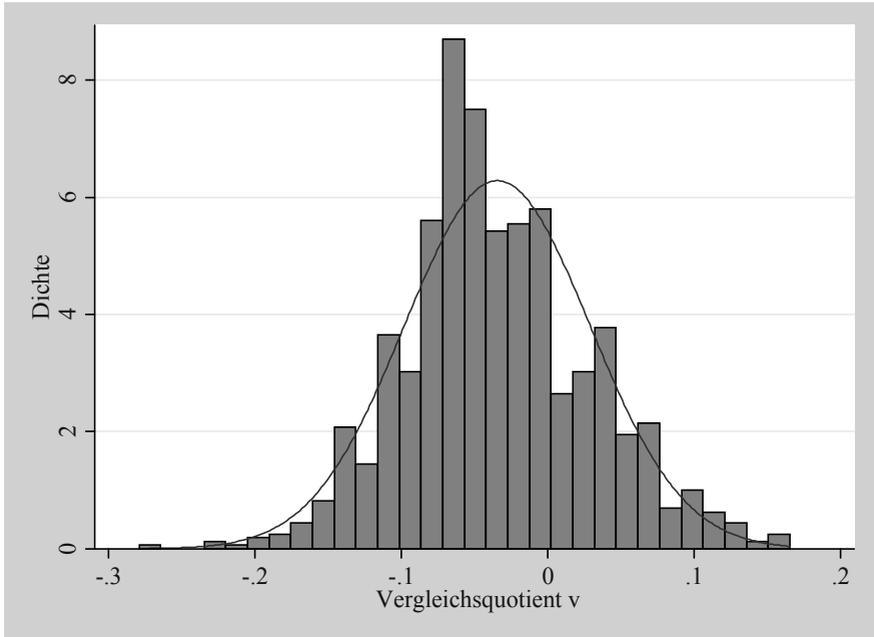
Tabelle 3.11: Mittlere absolute prozentuale Abweichungen der umfragebasierten Vorhersagen nach berücksichtigten Parteien, 1949-2009

	Mittelwert	Standardabweichung	N
Alle Parteien	20,26	10,47	1.070
Ohne Sonstige	16,62	8,19	1.070
Ohne Sonstige und Republikaner	17,73	9,17	1.070
Nur CDU/CSU und SPD	8,59	6,45	1.070
FDP, Bündnis 90/Die Grünen, PDS/Die Linke und sonstige Parteien	26,76	17,55	1.070

Anmerkungen: Die Angaben beziehen sich auf alle erfassten Ergebnisse von Sonntagsfragen, welche maximal ein Jahr vor einer Bundestagswahl veröffentlicht wurden.

Die Verteilung des Vergleichquotienten v ist in Abbildung 3.19 dargestellt. Hier zeigt sich einmal mehr die insgesamt gute Annäherung der Konzentration der tatsächlichen Zweitstimmenanteile durch die wahlabsichtsfragenbasierten Vorhersagen. Weiterhin ist der Abbildung zu entnehmen, dass v annähernd normalverteilt ist. Ein entsprechender Signifikanztest zeigt allerdings eine signifikante Abweichung der Verteilung von einer Normalverteilung.³⁹

³⁹ Durchgeführt wurde der von D'Agostino et al. (1990) vorgeschlagene und in Stata implementierte „skewness and kurtosis test for normality“ (StataCorp 2007: 289ff.).

Abbildung 3.19: Histogramm für Vergleichsquotienten v , 1949-2009

Anmerkungen: Die Darstellungen beziehen sich auf alle erfassten Ergebnisse von Sonntagsfragen, welche maximal ein Jahr vor einer Bundestagswahl veröffentlicht wurden.

In Abschnitt 3.3.2 wurde gezeigt, dass sowohl MAE als auch MAPE von der Anzahl inkludierter Parteien abhängen, jedoch unterschiedlich darauf reagieren. Entsprechend kann man auch einen zwar starken, jedoch nicht perfekten Zusammenhang der Maßzahlen erwarten. Eine Korrelationsanalyse bestätigt diese Vermutung: Der Bravais-Pearson-Korrelationskoeffizient r zwischen MAE und MAPE beträgt bei Berücksichtigung jeweils aller Parteien 0,41 und ist höchst signifikant. Der stärkste Zusammenhang ergibt sich – wie angesichts der jeweiligen Eigenschaften der Maßzahlen zu erwarten – zwischen dem MAPE für die großen Parteien und dem MAE ohne Berücksichtigung der Republikaner und sonstiger Parteien. Der Bravais-Pearson-Korrelationskoeffizient r steigt in diesem Fall auf 0,87 und ist höchst signifikant. Entsprechend schwach fällt andererseits die Korrelation zwischen dem MAPE für die kleinen Parteien und dem

MAE ohne Berücksichtigung der Republikaner sowie sonstiger Parteien aus ($r = 0,16$, höchst signifikant). Dies ist als Indiz zu werten, dass es zwingend erscheint, im Rahmen von vergleichbaren Evaluationsstudien nicht nur absolute Maßzahlen wie den MAE zu berücksichtigen.

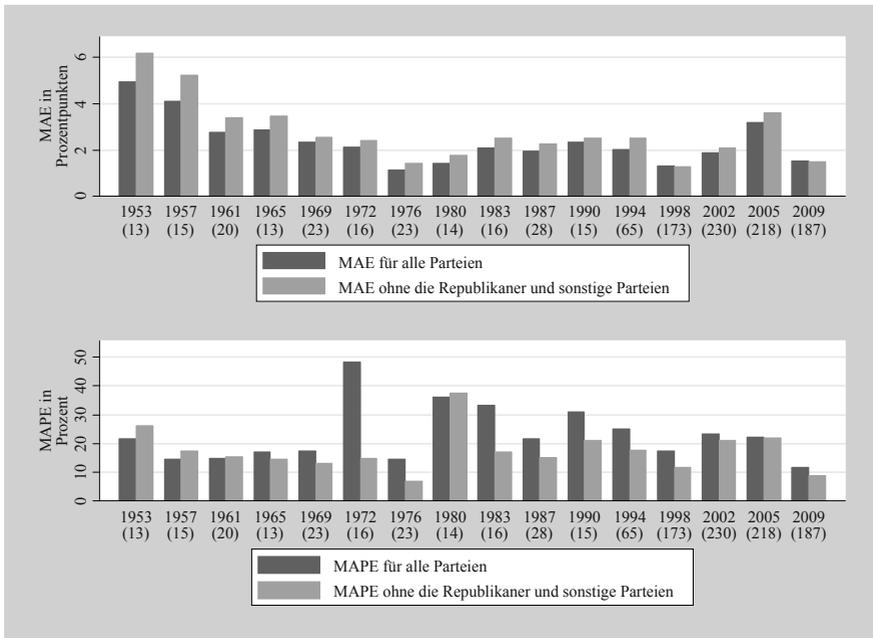
Der Vergleichsquotient ν beruht auf einem Vergleich der Konzentration der Verteilungen der vorhergesagten mit den tatsächlichen Stimmenanteilen. Damit liegt dieser Maßzahl zur Erfassung der Prognosegüte eine andere Logik als dem MAE und dem MAPE zugrunde. Entsprechend fallen die Bravais-Pearson-Korrelationskoeffizienten zwischen diesen Maßzahlen geringer aus. Der lineare Zusammenhang zwischen ν und dem MAE bei Berücksichtigung aller Parteien ist zwar höchst signifikant, beträgt jedoch nur $r = 0,14$. Die Korrelation mit dem MAPE für alle Parteien beträgt $r = 0,18$ und ist ebenfalls höchst signifikant. Diese schwachen Zusammenhänge konnten aufgrund der unterschiedlichen Messung der Prognosequalität durch die beiden Maße erwartet werden. Für eine vollständige Abdeckung der Prognosegüte ist deshalb die Berücksichtigung verschiedener Maßzahlen notwendig.

Im nächsten Schritt wird die durchschnittliche Prognosegüte der Wahlabsichtsumfragen im Zeitablauf betrachtet. Die methodischen und theoretischen Überlegungen ließen vermuten, dass im Zeitverlauf eine sinkende beziehungsweise eine zunächst steigende und dann wieder sinkende Prognosegüte zu beobachten sei (vgl. Abschnitt 3.1). Bei einer wahlspezifischen Betrachtung der mittleren absoluten sowie mittleren prozentualen Abweichungen sollte man folglich einen Trend zu steigenden Fehlern oder den angesprochenen umgekehrt u-förmigen Verlauf ablesen können.

Abbildung 3.20 bestätigt diesen postulierten kurvilinearen Zusammenhang nicht. Bezogen auf den MAE ist von 1953 bis 1976 eine stetige Zunahme der Prognosegüte zu verzeichnen. Diese verbleibt danach bis 1998 auf einem relativ niedrigen Niveau und für die beiden Bundestagswahlen 2002 und 2005 ist eine Verschlechterung der Annäherung erkennbar, jedoch waren die Sonntagfragen zur Bundestagswahl 2009 wieder vergleichsweise genau. Einschränkend muss jedoch gesagt werden, dass diese Interpretation keineswegs zwingend ist. So beruht die Berechnung der durchschnittlichen Abweichungen auf sehr unterschiedlichen Fallzahlen und es liegen erst für die letzten Bundestagswahlen auch Umfrageergebnisse vor, die einen großen zeitlichen Abstand zum Wahltermin aufweisen.

Ein weiterer Hinweis auf die nicht vorhandene Evidenz im Hinblick auf einen Zeittrend ergibt die Betrachtung der mittleren absoluten prozentualen Abweichung (untere Grafik). Hier ist ebenfalls kein klarer Trend im Zeitablauf zu erkennen. Auffällig hingegen ist der Ausreißer beim MAPE für alle Parteien im Jahr 1972. Dies ist auf die bereits erwähnte große Abweichung bei den sonstigen Parteien zurückzuführen.

Abbildung 3.20: Mittlere absolute und mittlere absolute prozentuale Abweichungen der umfragebasierten Vorhersagen nach Wahljahr, 1953-2009



Anmerkungen: Die Angaben beziehen sich auf alle erfassten Ergebnisse von Sonntagsfragen, welche maximal ein Jahr vor einer Bundestagswahl veröffentlicht wurden. Die Bundestagswahl 1949 wurde ausgeschlossen, da hierzu nur eine Vorhersage vorliegt. In den Klammern unter der Jahreszahl, welche die jeweilige Bundestagswahl indiziert, sind die der Berechnung zugrunde liegenden Fallzahlen dargestellt.

Ein Vergleich der beiden Darstellungen in Abbildung 3.20 verdeutlicht noch einmal die Auswirkung der Anwendung der verschiedenen Maße. Während beim MAE durchgehend die Werte ohne Berücksichtigung der Republikaner sowie der sonstigen Parteien über denjenigen bei Berücksichtigung aller Parteien liegt, verhält es sich beim MAPE genau anders herum.

Betrachtet man die Maßzahl v wahlspezifisch, so ergeben sich die in Tabelle 3.12 dargestellten Befunde. Demnach ist ebenfalls kein eindeutiger Trend im Zeitablauf zu erkennen. Gleichwohl sind auch hier deutlich wahlspezifische Schwankungen festzustellen. So ist beispielsweise bei den Wahlen in der jüngeren Vergangenheit nur 2002 die Konzentration der Stimmenanteile entgegen der sonstigen Tendenz im Mittel leicht unterschätzt worden.

Tabelle 3.12: Vergleichsquotient v nach Wahljahr, 1953-2009

Bundestagswahl	Mittelwert	Standardabweichung	N
1953	0,108	0,035	13
1957	0,030	0,035	15
1961	-0,077	0,048	20
1965	-0,050	0,033	13
1969	0,026	0,045	23
1972	-0,029	0,035	16
1976	0,011	0,027	23
1980	-0,019	0,017	14
1983	-0,027	0,040	16
1987	-0,085	0,051	28
1990	-0,063	0,035	15
1994	-0,011	0,049	65
1998	-0,054	0,031	173
2002	0,036	0,041	230
2005	-0,086	0,050	218
2009	-0,063	0,030	187

Anmerkungen: Die Angaben beziehen sich auf alle erfassten Ergebnisse von Sonntagsfragen, welche maximal ein Jahr vor einer Bundestagswahl veröffentlicht wurden.

Weiterhin ist eine nach Parteien differenzierte Betrachtung interessant. Dies erscheint lohnend, da damit untersucht werden kann, ob bestimmte Parteien systematisch über- oder unterschätzt werden. Zunächst können die absoluten und prozentualen Abweichungen parteispezifisch betrachtet werden. Das Ergebnis ist in Tabelle 3.13 abgetragen. Demnach ist erkennbar, dass es in Bezug auf den MAE keine wesentlichen Unterschiede zwischen den einzelnen Parteien gibt, sofern man deren Größe in Betracht bezieht. Ergiebiger ist hier der Vergleich der mittleren absoluten prozentualen Abweichungen. Hier zeigt sich, dass die Wahlabsichtsumfragen insbesondere bei den Republikanern sowie der PDS/Die Linke große Prognosefehler aufweisen. Insgesamt wird hier noch einmal besonders die relativ gute Vorhersage der Stimmenanteile der beiden großen Parteien deutlich, während die Fehler bei allen kleinen Parteien deutlich größer ausfallen. Die besonders großen Abweichungen bei der PDS/Die Linke und auch den Republikanern könnte man auf die geringen Erfahrungswerte mit diesen Parteien zurückführen, da ja anzunehmen wäre, dass mit zunehmender Erfahrung der Meinungsforschungsinstitute mit den jeweiligen Parteien deren politische Gewichtung besser wird. Mit Blick auf Bündnis 90/Die Grünen erscheint diese Vermutung nicht bestätigt, denn dann sollte man auch Unterschiede zwischen FDP und Bündnis 90/Die Grünen beobachten können. Dies ist nicht der Fall.

Die Betrachtung der absoluten Abweichungen ermöglicht es nicht, zu beurteilen, ob einzelne Parteien spezifisch über- oder unterschätzt werden. Um dieser Frage nachgehen zu können, muss das von Martin et al. (2005) vorgeschlagene *A*-Maß (vgl. Abschnitt 3.3.2) herangezogen werden. Das *A*-Maß nimmt den Wert Null an, wenn eine perfekte Übereinstimmung zwischen der Vorhersage und dem tatsächlichen Ergebnis vorliegt. Je größer sich der Wert absolut von Null unterscheidet, desto größer ist der Prognosefehler. Schließlich kann man anhand des Vorzeichens mögliche Verzerrungen aufzeigen.

Tabelle 3.13: Parteispezifische mittlere absolute und mittlere absolute prozentuale Abweichungen der umfragebasierten Vorhersagen, 1949-2009

		CDU/ CSU	SPD	FDP	Bündnis 90/Die Grünen	PDS/ Die Linke	Die Repub- likaner	Sons- tige Parteien
MAE	Mittelwert	3,20	2,97	1,87	1,55	1,68	1,09	1,13
	Std.- abw.	2,96	2,25	1,24	1,10	1,20	0,78	1,03
MAPE	Mittelwert	8,55	8,63	21,25	20,53	26,72	64,18	31,53
	Std.- abw.	8,02	6,67	14,06	20,09	20,13	50,99	33,43
N		1.070	1.070	1.064	944	870	144	1.070

Anmerkungen: Die Angaben beziehen sich auf alle erfassten Ergebnisse von Sonntagsfragen, welche maximal ein Jahr vor einer Bundestagswahl veröffentlicht wurden.

Zunächst liegt es nahe, zu analysieren, ob sich hinsichtlich der beiden großen Parteien entsprechende systematische Abweichungen feststellen lassen. Hier ergibt sich ein A -Wert von 0,05 mit einer Standardabweichung von 0,20. Das positive Vorzeichen indiziert eine Verzerrung zugunsten von CDU/CSU im Vergleich zur SPD. Dies bedeutet, dass im Durchschnitt über den gesamten Zeitraum der Stimmenanteil von CDU/CSU etwas zu hoch vorhergesagt wird, während der von der SPD gemessen an ihren Ergebnissen zu niedrig eingestuft wird. Allerdings erscheint die Abweichung sehr gering, zumal wenn man sich die Höhe der Standardabweichung vor Augen führt.

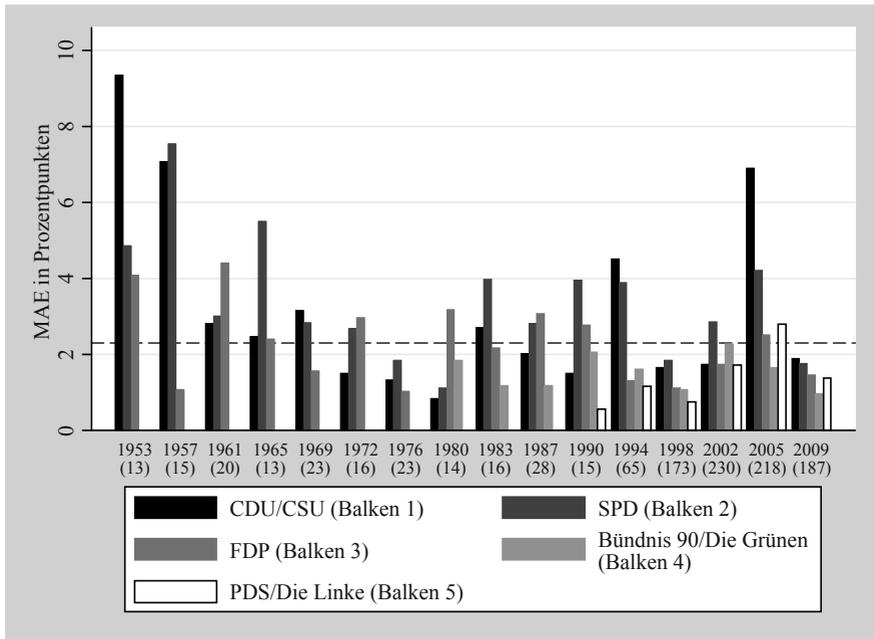
Ebenfalls ist eine leichte, wenn auch ebenso unbedeutende Überschätzung der Stimmenanteile von CDU/CSU und FDP gegenüber der SPD sowie Bündnis 90/Die Grünen festzustellen. Weitert man diese Betrachtung etwas aus und fragt sich, ob eher rechte oder linke Parteien überschätzt werden, so ergibt sich ein A -Wert von 0,03 (Standardabweichung: 0,16), welcher vernachlässigbar ist. Vergleicht man große mit kleinen Parteien, so zeigt sich eine geringe Überschätzung der Stimmenanteile großer Parteien und eine geringe Unterschätzung der Stimmenanteile kleiner Parteien (A -Maß: 0,08; Standardabweichung: 0,19). Interessant erscheint schließlich noch eine vergleichende Betrachtung von FDP und

Bündnis 90/Die Grünen. In Abschnitt 3.4.1 wurde gezeigt, dass die politische Gewichtung bei diesen beiden Parteien systematisch ausfällt. Während die Stimmenanteile von Bündnis 90/Die Grünen nach unten gewichtet werden, verhält es sich bei der FDP genau anders herum. Doch wie sieht die Konsequenz in Bezug auf die Prognosegüte aus? Das entsprechende A -Maß nimmt den Wert von $-0,10$ (Standardabweichung: $0,43$) an. Dies deutet auf eine schwache Überschätzung des Stimmenanteils von Bündnis 90/Die Grünen und entsprechend einer Unterschätzung des Stimmenanteils der FDP hin.

Anhand der Analysen der Politbarometer-Zeitreihen wurde bereits gezeigt, dass die Abweichungen insbesondere bei den beiden großen Parteien wahlspezifisch variieren. Deshalb wird die parteispezifische Betrachtung weiter nach Bundestagswahl ausdifferenziert. Abbildung 3.21 zeigt zunächst die parteispezifischen mittleren absoluten Abweichungen. Auffällig sind hier insbesondere die Wahlen vor 1970 sowie auch die Bundestagswahl von 2005. Für die Wahlen 1953 und 1957 sind zumindest für die beiden großen Parteien erhebliche Abweichungen festzustellen.

Angesichts der nur geringen Anzahl an Sonntagsfragenergebnissen, der zum damaligen Zeitpunkt noch jungen deutschen Meinungsforschung sowie der mangelnden Erfahrungswerte überrascht dies wenig. Für CDU/CSU und die SPD sind in der Folge keine so großen Abweichungen mehr festzustellen. Einzig bei der Bundestagswahl von 1965 wurde der Stimmenanteil der SPD nur relativ schlecht vorhergesagt und die Prognosegüte hinsichtlich des Stimmenanteils von CDU/CSU im Rahmen der Bundestagswahl von 2005 weist eine unterdurchschnittliche Annäherung auf. Bei den kleinen Parteien sind insbesondere die deutlichen Schwankungen der Prognosegüte des Stimmenanteils der FDP auffällig. Im Rahmen der Bundestagswahl 2009 wurden den Stimmenanteile aller Parteien überdurchschnittlich gut vorhergesagt. Insgesamt muss festgehalten werden, dass auch für einzelne Parteien keine klaren Trends hinsichtlich der Entwicklung der Prognosegüte festgestellt werden können.

Abbildung 3.21: Mittlere absolute Abweichungen der umfragebasierten Vorhersagen nach Partei und Bundestagswahl, 1953-2009

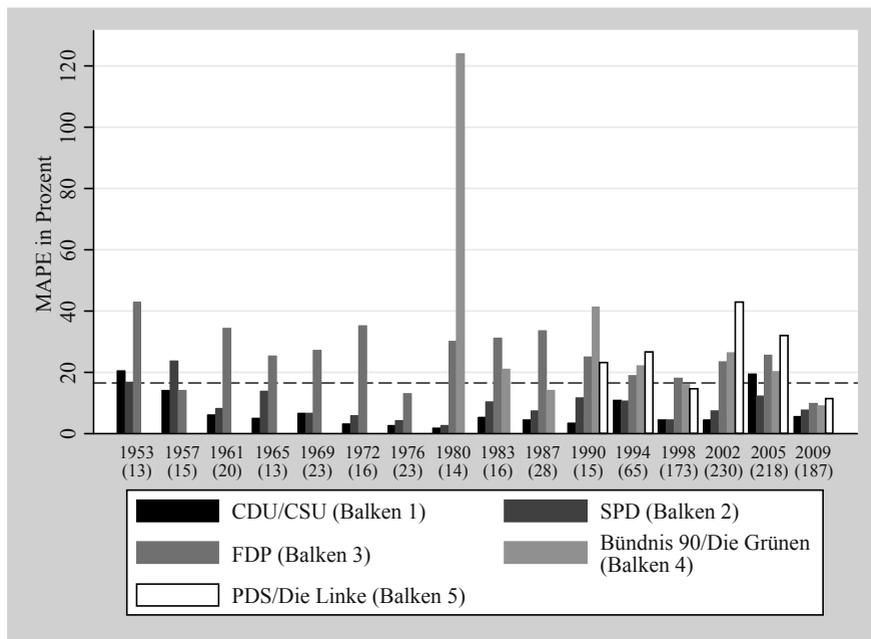


Anmerkungen: Die Angaben beziehen sich auf alle erfassten Ergebnisse von Sonntagsfragen, welche maximal ein Jahr vor der jeweiligen Bundestagswahl veröffentlicht wurden. In Klammern unter der Jahreszahl der Bundestagswahl sind die Fallzahlen angegeben. Die schwarze gestrichelte horizontale Linie entspricht der durchschnittlichen mittleren Abweichung über den gesamten dargestellten Zeitraum und bezieht sich auf die dargestellten Parteien.

Um beurteilen zu können, wie die Vorhersagen der einzelnen Parteien zueinander abschneiden, wird die mittlere absolute prozentuale Abweichung herangezogen (Abbildung 3.22). Hier ist ein sehr markanter Ausreißer festzustellen: Der Stimmenanteil der damals neu gegründeten Partei Die Grünen wurde für die Bundestagswahl 1980 miserabel vorhergesagt. Im Schnitt weichen die 14 Sonntagsfragenergebnisse um 124,1 Prozent vom tatsächlichen Stimmenanteil ab. Allerdings muss erwähnt werden, dass alle Umfragen die Grünen unter fünf Prozent sahen. Sie erreichten schließlich 1,5 Prozent, die Vorhersagen sahen die Partei im Schnitt bei 3,6 Prozent. Erkennbar ist jedoch auch, dass bereits bei der

darauf folgenden Bundestagswahl die mittlere absolute prozentuale Abweichung nur mehr 21,3 Prozent beträgt und damit nur geringfügig über dem Durchschnitt (horizontale gestrichelte schwarze Linie).

Abbildung 3.22: Mittlere absolute prozentuale Abweichungen der umfragebasierten Vorhersagen nach Partei und Bundestagswahl, 1953-2009



Anmerkungen: Die Angaben beziehen sich auf alle erfassten Ergebnisse von Sonntagsfragen, welche maximal ein Jahr vor der jeweiligen Bundestagswahl veröffentlicht wurden. In Klammern unter der Jahreszahl der Bundestagswahl sind die Fallzahlen angegeben. Die schwarzgestrichelte horizontale Linie entspricht der durchschnittlichen mittleren prozentualen Abweichung über den gesamten dargestellten Zeitraum und bezieht sich auf die dargestellten Parteien.

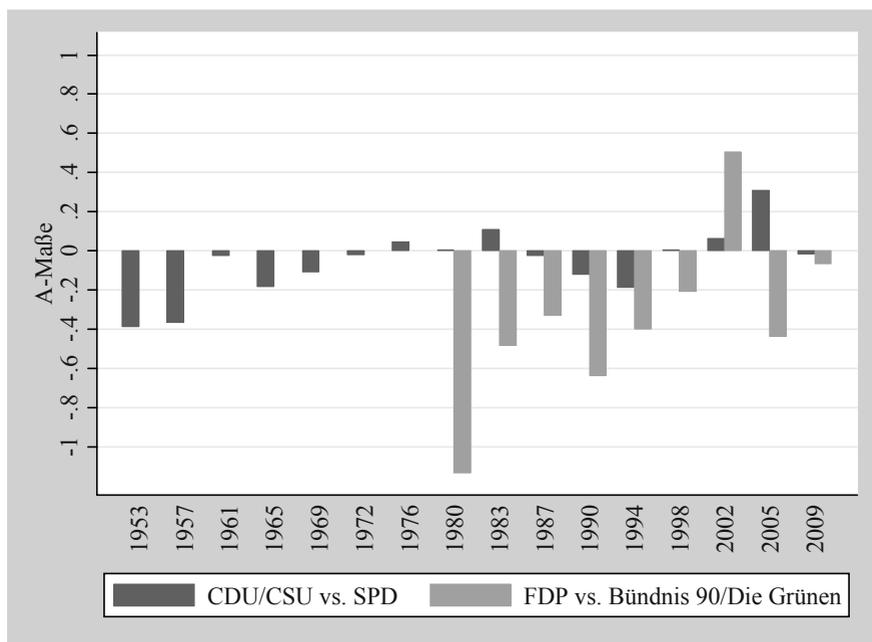
Weiterhin zeigt sich hier, dass die Stimmenanteile der beiden großen Parteien insgesamt sehr passabel vorhergesagt werden, während die relativen Fehler bei allen kleinen Parteien deutlich größer sind. Vor dem Hintergrund der Fünf-Prozent-Hürde und der in aller Regel gegebenen Notwendigkeit von Koalitionsregierungen erscheinen diese großen Abweichungen bemerkenswert. Dieser

Befund verdeutlicht einmal mehr die Notwendigkeit der Entwicklung wissenschaftlicher Prognoseinstrumente, um auch bei diesen Parteien möglichst bessere Vorhersagen zu erzielen als mit den vorliegenden umfragebasierten Instrumenten.

Unbeantwortet ist bisher die Frage nach der Richtung der Abweichung. Dazu wird das A -Maß herangezogen. In Abbildung 3.23 sind die richtungsspezifischen Abweichungen für die CDU/CSU gegenüber der SPD sowie von der FDP gegenüber Bündnis 90/Die Grünen dargestellt. Hintergrund der Auswahl dieser Vergleichspaare ist der Befund aus Abschnitt 3.4.1. Danach variiert bei den beiden großen Parteien der Einfluss der politischen Gewichtung wahl-spezifisch und zwar in der Art, dass jeweils die in den Rohdaten ermittelten Stimmenanteile der Kanzlerpartei hoch, die der großen Oppositionspartei herunter gewichtet werden. Bei der FDP und Bündnis 90/Die Grünen verhält es sich anders. Hier wurde festgestellt, dass der Stimmenanteil der FDP systematisch nach oben gewichtet wird, der von Bündnis 90/Die Grünen nach unten.

Diese Analysen zeigen, dass bis 1972 tendenziell die Stimmenanteile der SPD überschätzt wurden. Die einzige merkliche Unterschätzung des SPD-Stimmenanteils ist für die Bundestagswahl 2005 festzustellen. Bei der FDP im Vergleich mit Bündnis 90/Die Grünen zeigt sich, dass der Stimmenanteil der FDP fast durchgehend unterschätzt wird. Eine Ausnahme stellt lediglich die Wahl von 2002 dar. Als mögliche Gründe hierfür kommen Wahlspezifika in Frage. Der Wahlkampf wurde maßgeblich bestimmt von der Positionierung Deutschlands zum Irakkrieg und in der heißen Phase auch durch das Krisenmanagement des Elbehochwassers. Insbesondere FDP-Spitzenkandidat Guido Westerwelle wurde hier Versagen attestiert, was zum unerwartet schlechten Abschneiden der Partei ebenso beigetragen haben könnte wie die so genannte Flugblatt-Affäre um Jürgen Möllemann (für Analysen zur Wahl vgl. etwa die Aufsätze in Brettschneider et al. 2004; Jesse 2003).

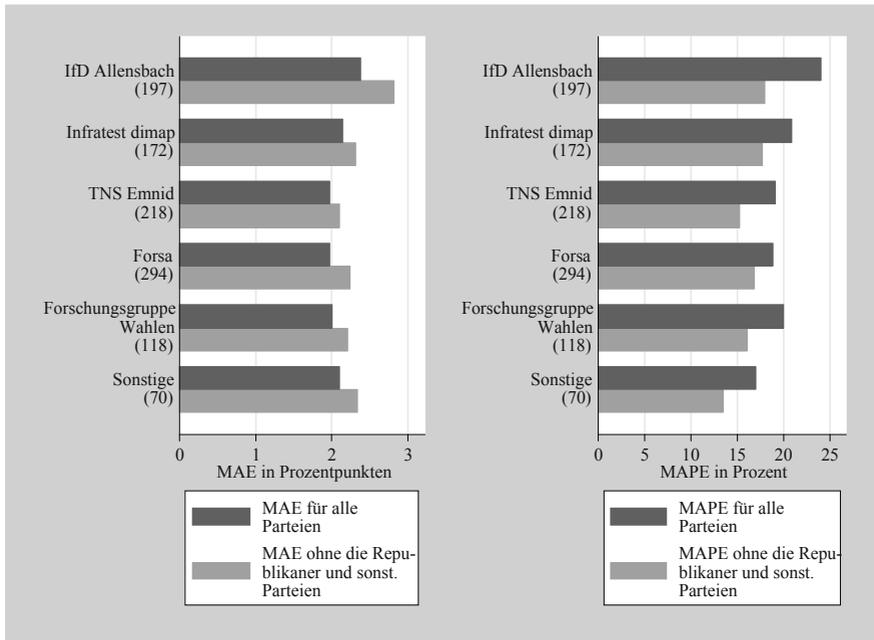
Abbildung 3.23: Verzerrungen der Vorhersagen von Wahlabsichtsumfragen nach Partei und Bundestagswahl, 1953-2009



Anmerkungen: Die Angaben beziehen sich auf alle erfassten Ergebnisse von Sonntagsfragen, welche maximal ein Jahr vor der jeweiligen Bundestagswahl veröffentlicht wurden.

Als letzter Analyseschritt dieser deskriptiven Betrachtung der Vorhersagegüte von Sonntagsfragen wird die Performanz der einzelnen Institute einander gegenüber gestellt. Ein erster Vergleich zeigt sowohl für den MAE als auch den MAPE, dass zwar einzelne Differenzen bestehen, diese jedoch kein erhebliches Ausmaß annehmen (Abbildung 3.24).

Abbildung 3.24: Institutsspezifischer Vergleich der Prognosegüte von umfragebasierten Vorhersagen, MAE und MAPE, 1953-2009



Anmerkungen: Die Angaben beziehen sich auf alle erfassten Ergebnisse von Sonntagsfragen, welche maximal ein Jahr vor der jeweiligen Bundestagswahl veröffentlicht wurden. Da zur Bundestagswahl 1949 nur ein Umfrageergebnis vorliegt, wurde diese Wahl nicht berücksichtigt.

Dies zeigt sich im Übrigen auch bei Betrachtung der institutsspezifischen Vergleichsquotienten v (Tabelle 3.14). Hier sind zwar scheinbar deutliche Unterschiede zwischen dem IfD Allensbach und Infratest dimap zu erkennen, diese sind jedoch vor dem Hintergrund der beträchtlichen Standardabweichungen (jeweils ca. 0,07) nicht überzubewerten. Überdies ist festzustellen, dass alle Institute tendenziell die Konzentration der Stimmenanteile leicht überschätzen.

Tabelle 3.14: Institutsspezifischer Vergleich der Prognosegüte von umfragebasierten Vorhersagen, Vergleichsquotient v , 1953-2009

Institut	Mittelwert	Standardabweichung	N
IfD Allensbach	-0,015	0,073	197
Infratest dimap	-0,045	0,064	172
TNS Emnid	-0,030	0,060	218
Forsa	-0,039	0,058	294
Forschungsgruppe Wahlen	-0,047	0,057	118
Sonstige	-0,034	0,061	70

Anmerkungen: Die Angaben beziehen sich auf alle erfassten Ergebnisse von Sonntagsfragen, welche maximal ein Jahr vor der jeweiligen Bundestagswahl veröffentlicht wurden. Da zur Bundestagswahl 1949 nur ein Umfrageergebnis vorliegt, wurde diese Wahl nicht berücksichtigt.

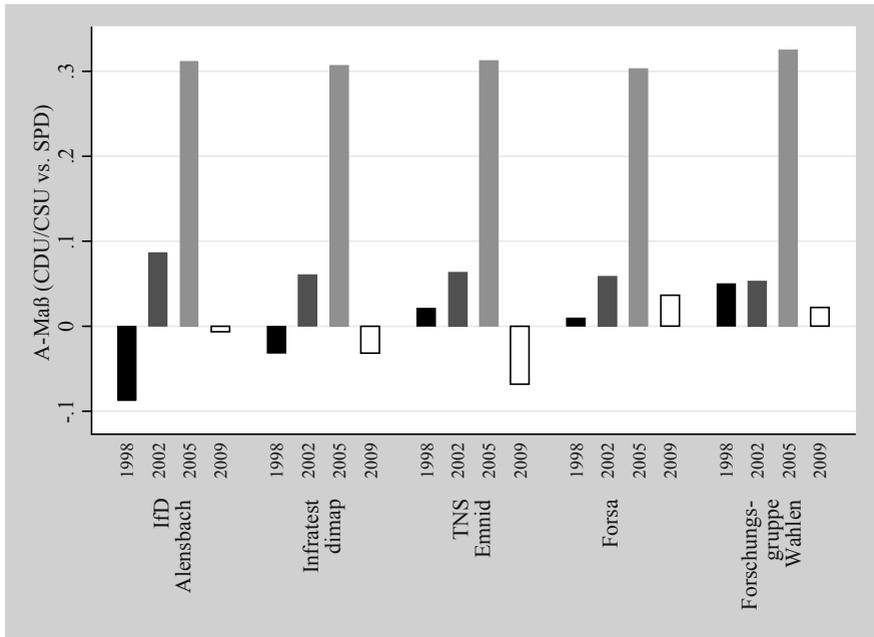
Damit bestätigen diese Befunde das Ergebnis von Antholz (2001: 161), der ebenfalls eher geringe institutsspezifische Abweichungen der durchschnittlichen Prognosefehler findet. Dennoch ist darauf hinzuweisen, dass diese bivariate Betrachtung um eine multivariate Analyse ergänzt werden muss, da möglicherweise unter Kontrolle auf Drittvariablen relevante Institutsunterschiede aufgedeckt werden können.

Die Aktivität der Institute schwankt über die Zeit. So sind das IfD Allensbach und TNS Emnid die einzigen Institute, welche zumindest seit 1953 bereits in der politischen Meinungsforschung aktiv sind. Infratest dimap beispielsweise wurde erst in den 1990er Jahren gegründet und Infas betätigt sich nicht mehr auf diesem Gebiet. Um deshalb mögliche parteispezifische Verzerrungen einzelner Institute darzustellen, wird nur auf die letzten drei Bundestagswahlen rekurriert, so dass ein fairer Vergleich ermöglicht wird. Die Ergebnisse sind Abbildung 3.25 zu entnehmen.

Demnach ist deutlich erkennbar, dass von parteispezifischen Verzerrungen zugunsten von CDU/CSU oder SPD bei einzelnen Instituten nicht auszugehen ist. Für die Bundestagswahlen 2002 und 2005 weisen alle Institute eine Überschätzung des Stimmenanteils von CDU/CSU auf – und dies auch in jeweils ähnlichen Ausmaßen. Größere Unterschiede sind nur für die Bundestagswahlen 1998 und 2009 festzustellen. Während das IfD Allensbach sowie Infratest dimap

den Stimmenanteil der SPD 1998 überschätzen, sagen TNS Emnid, Forsa sowie die Forschungsgruppe Wahlen einen zu hohen Stimmenanteil für CDU/CSU voraus. Bei der Bundestagswahl 2009 überschätzen Forsa und die Forschungsgruppe Wahlen leicht den Stimmenanteil von CDU/CSU.

Abbildung 3.25: Institutsspezifische Verzerrungen der umfragebasierten Vorhersagen der Stimmenanteile von CDU/CSU und SPD, 1998-2009



Anmerkungen: Die Angaben beziehen sich auf alle erfassten Ergebnisse von Sonntagsfragen der dargestellten Institute, welche maximal ein Jahr vor der jeweiligen Bundestagswahl veröffentlicht wurden.

Eine abschließende Beurteilung über das Vorliegen bewusster oder unbewusster Über- beziehungsweise systematischer Unterschätzung bestimmter Parteistimmenanteile ist auf Grundlage dieser Analysen nicht möglich, jedoch zeigen sich – wie vermutet – keine entsprechenden Hinweise.

Zusammenfassend ist zu konstatieren, dass *erstens* die Wahlabsichtsumfragen gemessen an der mittleren absoluten Abweichung eine durchaus passable Prognosegüte aufweisen. *Zweitens* ist jedoch anhand der Analysen auf Grundlage der mittleren absoluten prozentualen Abweichung deutlich geworden, dass die Stimmenanteile kleinerer Parteien nur unzureichend vorhergesagt werden können. Dies bestätigen auch die parteispezifischen Betrachtungen. Weiterhin konnte *drittens* kein eindeutiger Trend hinsichtlich der Prognosegüte im Zeitablauf festgestellt werden. Allenfalls auf Grundlage des MAE könnte eine zunächst steigende und in den letzten Jahren wieder abnehmende Vorhersagegüte vermutet werden – Auswertungen auf Basis des MAPE stützen diesen Befund jedoch nicht. *Viertens* ist festzuhalten, dass die häufig geäußerte Vermutung, dass einzelne Institute systematisch eine der beiden großen Parteien über- oder unterschätzen – zumindest für die jüngsten Bundestagswahlen – nicht bestätigt werden konnte. Demnach stützen diese Analysen die in dieser Arbeit herausgearbeitete Hypothese, wonach aufgrund des Konkurrenzdrucks und der unmittelbaren Beobachtbarkeit demoskopischer Arbeit in Bezug auf die Sonntagsfragen langfristig eine solche systematische Verzerrung nicht aufrecht erhalten werden könnte (*Hypothese 4b*).

3.4.3 Einflussfaktoren auf die Prognosegüte von Wahlabsichtsfragen

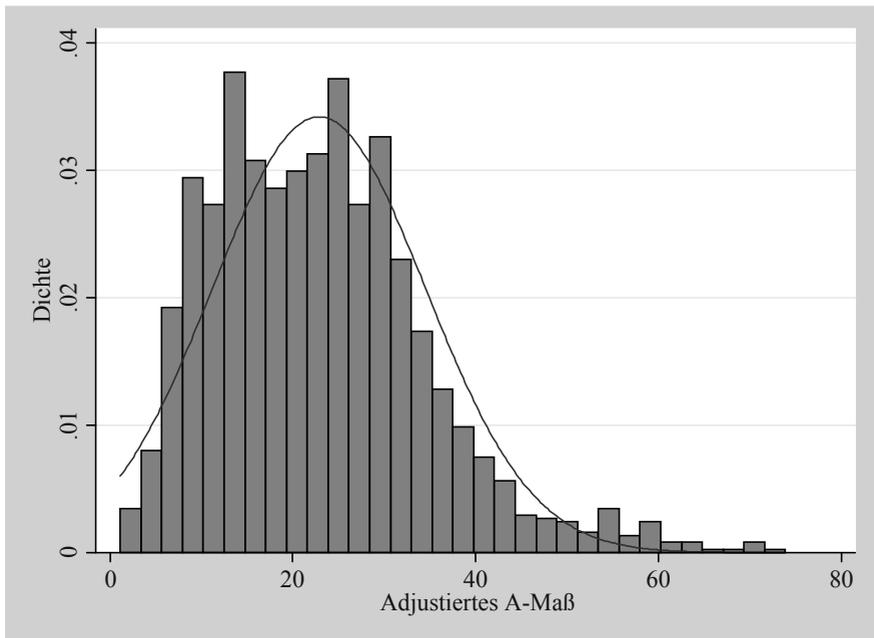
Für die Analyse einflussreicher Faktoren auf die Prognosegüte von Wahlabsichtsumfragen wird das adjustierte A -Maß als Gütekriterium herangezogen (vgl. Abschnitt 3.3.2.4). Dieses Maß ermöglicht den Vergleich der Abweichungen auch im Fall variierender Anzahl an Parteien. Nur so sind für Deutschland angemessene Vergleiche über die Zeit möglich. Weiterhin weist dieses Maß nicht die Eigenschaft auf, Abweichungen bei großen oder kleinen Parteien überproportional zu berücksichtigen. Dies ist ein zentrales Manko des MAE sowie des MAPE, weshalb deren Anwendung im Rahmen der Analyse potenzieller Einflussfaktoren gerade vor dem Hintergrund des deutschen Parteiensystems weniger sinnvoll erscheint.

Das adjustierte A -Maß repräsentiert den Durchschnitt der absoluten A -Maße für jede Partei im Verhältnis zu allen anderen. Schließlich wird das Maß mit 100 multipliziert. Bei der Berechnung wurde auf die Inklusion der Partei Die Repub-

likaner sowie sonstiger Parteien verzichtet. Das Maß ist dahingehend zu interpretieren, dass ein Wert von Null als eine perfekte Übereinstimmung der Prognose mit dem tatsächlichen Wahlergebnis aufzufassen ist. Je größer das adjustierte A -Maß wird, desto geringer die Prognosegüte. Die Verteilung der abhängigen Variable ist in Abbildung 3.26 dargestellt.

Das Histogramm zeigt, dass die abhängige Variable annähernd normalverteilt aber leicht rechtsschief ist. Gleichwohl ist eine durchaus beachtliche Streuung der Abweichungen der Sonntagsfragenergebnisse vom tatsächlichen Ergebnis festzustellen. Dies ist auch Tabelle 3.15 zu entnehmen, welche eine Übersicht über die Merkmalsverteilung aller im Rahmen dieser Analyse relevanten Variablen enthält.

Abbildung 3.26: Histogramm adjustiertes A -Maß, 1953-2009



Anmerkungen: Die Darstellung bezieht sich auf alle erfassten Ergebnisse von Sonntagsfragen, welche maximal ein Jahr vor der jeweiligen Bundestagswahl veröffentlicht wurden. Da für die Bundestagswahl 1949 nur eine umfragebasierte Vorhersage vorliegt, wird diese in den folgenden Analysen nicht berücksichtigt.

Entgegen der bisherigen Betrachtungen werden nun alle vorhandenen Wahlabsichtsfragenergebnisse, welche maximal 365 Tage vor einer Bundestagswahl veröffentlicht wurden, in die Analyse mit einbezogen – die Beschränkung der Betrachtung der auf das gesamte Bundesgebiet bezogenen Sonntagfragen wird folglich aufgehoben.

Aufgrund der vorliegenden hierarchischen Datenstruktur sowie der variierenden Fallzahlen werden die als einflussreich vermuteten Faktoren nicht analog zur Reihenfolge der Hypothesenentwicklung in Abschnitt 3.1 aufgegriffen. Weiterhin sind zur angemessenen Überprüfung der Hypothesen Verfahren der Mehrebenenanalyse anzuwenden (siehe hierzu Abschnitt 3.3.3.2). Diesen Überlegungen folgend werden *random-intercept*-Modelle geschätzt, welche auf Ebene 1 die Wahlabsichtsumfrageergebnisse und auf Ebene 2 die Bundestagswahlen umfassen. Die Institute werden als Dummy-Variablen einbezogen und so als weitere Ebene mit fixen Effekten betrachtet.⁴⁰

In Bezug auf den vermuteten Trend über die Zeit erscheint erklärungsbedürftig, warum in diesem Rahmen nicht wie in Abschnitt 3.4.1 auf zeitreihenanalytische Verfahren zurückgegriffen wird. Hier werden nur diejenigen Umfrageergebnisse berücksichtigt, welche maximal ein Jahr vor einer Bundestagswahl veröffentlicht wurden. Insofern liegen allenfalls für jede einzelne Wahl Zeitreihen vor. Diese müssten entsprechend getrennt analysiert werden, womit der Einfluss von Wahlspezifika nicht explizit erfasst werden könnte. Zweitens werden die Wahlabsichtsumfragen der einzelnen Institute in unterschiedlichen Intervallen veröffentlicht. Während beispielsweise die Forschungsgruppe Wahlen monatlich den Politbarometer durchführt, veröffentlicht Forsa wöchentlich Ergebnisse des *tracking polls*. Entsprechende Analysen sind allenfalls für einzelne Institute möglich – analog zur Betrachtung in Abschnitt 3.4.1. Ein solcher Blick wird jedoch dem Anspruch einer generellen Identifikation von Einflussfaktoren auf die Prognosegüte umfragebasierter Vorhersagen nicht gerecht. Zudem sind entsprechende institutsinterne Analysen aufgrund mangelnder Varianz (etwa der Stichprobengröße) hinsichtlich der theoretisch als einflussreich vermuteten Faktoren nicht möglich.

40 Sämtliche Berechnungen wurden mit Stata 10.1 durchgeführt. Eine Darstellung der Mehrebenenanalyse mit Stata liefern Rabe-Hesketh und Skrondal (2008).

Tabelle 3.15: Merkmalsverteilung der verwendeten Variablen, 1953-2009

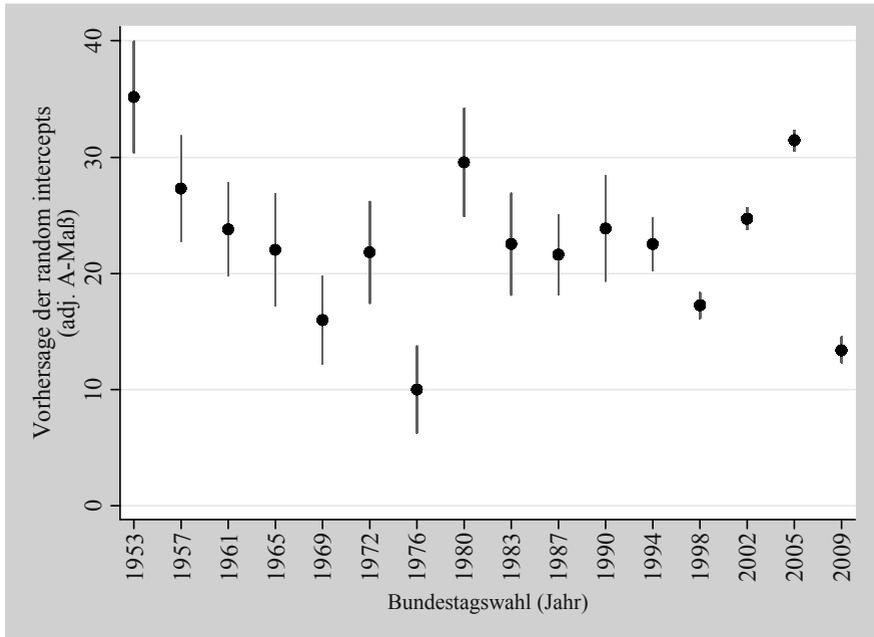
	Mittelwert	Standard- abweichung	Minimum	Maximum	N
Adjustiertes A-Maß	22,87	11,66	1,12	73,94	1.642
IfD Allensbach (1 = ja)	0,19		0	1	1.642
Infratest dimap (1 = ja)	0,29		0	1	1.642
TNS Emnid (1 = ja)	0,19		0	1	1.642
Forsa (1 = ja)	0,22		0	1	1.642
Forschungsgruppe Wahlen (1 = ja)	0,07		0	1	1.642
Sonstige Institute (1 = ja)	0,04		0	1	1.642
Neue Bundesländer (1998-2009) (1 = ja)	0,17		0	1	1.642
Alte Bundesländer (1949-1987) (1 = ja)	0,11		0	1	1.642
Alte Bundesländer (1990-2009) (1 = ja)	0,17		0	1	1.642
Gesamtdeutschland (1990-2009) (1 = ja)	0,54		0	1	1.642
Zeitl. Abstand zur Wahl in Tagen	150,41	106,90	1	365	1.642
Wahlbeteiligung in Prozent	78,23	5,03	64,8	91,1	1.642
Stichprobengröße	2.022,30	1.047,23	276	5.983	685
Befragungszeit in Tagen	5,68	2,90	1	32	643
Am Wochenende befragt (1 = ja)	0,45		0	1	643

Anmerkungen: Die Angaben beziehen sich auf alle erfassten Ergebnisse von Sonntagsfragen, welche maximal ein Jahr vor der jeweiligen Bundestagswahl veröffentlicht wurden. Da für die Bundestagswahl 1949 nur eine umfragebasierte Vorhersage vorliegt, wird diese in den folgenden Analysen nicht berücksichtigt.

Zunächst wird der Frage nachgegangen, ob die Mehrebenenmodellierung notwendig ist. Dies ist positiv zu beantworten wie das Nullmodell (Tabelle 3.16,

Modell 1) zeigt.⁴¹ Zur graphischen Verdeutlichung der Unterschiede zwischen den einzelnen Wahlen sind in Abbildung 3.27 die geschätzten *random intercepts* nach Bundestagswahl dargestellt.

Abbildung 3.27: Geschätzte random intercepts des adjustierten A-Maßes für Bundestagswahlen, 1953-2009



Anmerkungen: Die Abbildung beruht auf einer Maximum-Likelihood-Schätzung des random-intercept-Modells 1 aus Tabelle 3.16 auf das adjustierte A-Maß. Dargestellt sind die geschätzten random intercepts inkl. 95-Prozent-Konfidenzintervall.

Deutlich zu erkennen sind die erheblichen Differenzen zwischen den durchschnittlichen Prognosefehlern je nach Wahl. Ebenfalls auffällig sind die markant unterschiedlich großen Konfidenzintervalle. Dies ist auf unterschiedliche Fall-

41 Die in Tabelle 3.16 dargestellten Modelle 1 bis 4 wurden zur Kontrolle auch mit *fixed-effects*-Spezifikation berechnet. Diese Schätzungen bestätigen die im Folgenden berichteten Ergebnisse.

zahlen innerhalb der Gruppen zurückzuführen. Für die Bundestagswahlen 1998, 2002, 2005 und 2009 liegen mit Abstand am meisten umfragebasierte Vorhersagen vor. Bemerkenswert ist zudem, dass die Ergebnisse der jüngsten Bundestagswahl 2009 im Durchschnitt nach der Bundestagswahl von 1976 am besten approximiert wurden.

Auch die Differenzierung nach einzelnen Instituten sollte beibehalten werden, wie Modell 2 in Tabelle 3.16 zu entnehmen ist. Demnach weicht in dieser Betrachtung nur die Prognosegüte von Forsa nicht signifikant von derjenigen des IfD Allensbach ab. Gleichwohl sind die Differenzen zwischen den Instituten – wie bereits in der bivariaten Betrachtung gesehen (Abschnitt 3.4.2) – eher gering. Die Unterschiede der durchschnittlichen Prognosegüte der Institute bleiben auch bei Kontrolle auf Drittvariablen gering und verschwinden zumeist (Modelle 3 bis 5). In Modell 5 weist lediglich TNS Emnid im Durchschnitt eine signifikant bessere Prognosegüte auf als das IfD Allensbach, wohingegen Forsa eine hoch signifikant schlechtere Prognosegüte aufweist. Aufgrund der Variation dieser Effekte mit der Berücksichtigung verschiedener Kontrollvariablen, der nur knapp signifikanten und geringen Effekte auch in bivariater Betrachtung sollten diese Unterschiede jedoch nicht überbewertet werden. Auffällig ist hier insbesondere die erhebliche Abhängigkeit der Effekte der einzelnen Institute je nach betrachteten Bundestagswahlen.

In Bezug auf die vermuteten *house effects* ist festzuhalten, dass die Institute zumindest teilweise auch unter Kontrolle weiterer Variablen voneinander hinsichtlich ihrer durchschnittlichen Prognosegüte abweichen, diese Differenzen jedoch gering ausfallen. Auch ein durchgeführter *Likelihood-Ratio*-Test zeigt, dass das Modell inklusive der *house effects* dem Nullmodell vorzuziehen ist (LR-Chi²-Wert = 61,23, höchst signifikant). Insofern kann *Hypothese 4a* eingeschränkt bestätigt werden. Vor dem Hintergrund des bedeutenden Einflusses der politischen Gewichtung auf die berichteten Vorhersagen (Abschnitt 3.4.1) verwundern diese geringen Unterschiede. Dies kann als Hinweis gedeutet werden, dass entweder die eingesetzten Verfahren sich relativ stark ähneln, oder aber, dass die Meinungsforschungsinstitute sich jeweils untereinander von den Vorhersagen der anderen Institute beeinflussen lassen.

Gemäß *Hypothese 4b* wird vermutet, dass keine systematischen, parteispezifischen Verzerrungen einzelner Institute zu beobachten sind. Aufgrund der Notwendigkeit der Verwendung des adjustierten *A*-Maßes im Rahmen der multi-

variater Betrachtung ist eine zufriedenstellende Analyse dieser Beziehung hier nicht möglich. Die bivariaten Betrachtungen (Abschnitt 3.4.2) liefern jedoch keine Anhaltspunkte für entsprechende systematische Über- oder Unterschätzung der Stimmenanteile einzelner Parteien von bestimmten Instituten. Vielmehr zeigen sich mit der jeweiligen Bundestagswahl variierende Effekte. Insofern kann von einer Bestätigung von *Hypothese 4b* ausgegangen werden.

Auf Basis methodischer Erwägungen wurde argumentiert, dass umfragebasierte Vorhersagen, welche auf Quotenstichproben beruhen, ungenauer sein sollten als solche, die auf Zufallsstichproben fußen. Es wurde bereits angemerkt, dass ein stringenter Test dieser Vermutung aufgrund fehlender Informationen nicht verfolgt werden kann. Jedoch ist bekannt, dass das IfD Allensbach konsequent auf Quotenstichproben zurückgreift, während Infratest dimap, Forsa und auch die Forschungsgruppe Wahlen ausschließlich Zufallsstichproben ziehen. Insofern müsste man erwarten, dass die letzt genannten Institute signifikant niedrigere Prognosefehler aufweisen als das IfD Allensbach. Dies ist, wie Tabelle 3.16 zu entnehmen ist, nicht der Fall.

Dennoch kann von einer Falsifizierung von *Hypothese 4b* nicht zwingend ausgegangen werden. Der Einfluss der politischen Gewichtung dürfte weitaus bedeutender sein als der Stichprobenverfahrenseffekt und angesichts der Tatsache, dass das IfD Allensbach die weitaus größten Erfahrungen im Hinblick auf die politische Meinungsforschung aufweist, könnte man vermuten, dass dieses methodische Manko durch möglicherweise adäquatere Gewichtungsverfahren ausgeglichen werden kann.

Hypothese 6 zufolge sollten die Wahlresultate für die neuen Bundesländer schlechter vorhergesagt werden können als die für die alten Bundesländer. Im Datensatz sind für die Zeit nach der Wiedervereinigung Umfrageergebnisse für die neuen, die alten sowie Gesamtdeutschland enthalten. Die Analysen zeigen, dass im Durchschnitt unter Kontrolle auf die Wahl sowie die Institute die Vorhersagen für die alten Bundesländer nach 1990 die größten Abweichungen aufweisen. Allerdings sind die Unterschiede zu den umfragebasierten Vorhersagen für die neuen Bundesländer in Modell 5 nicht signifikant. Eine höchst signifikant bessere Prognosegüte weisen die Vorhersagen für Gesamtdeutschland seit 1990 auf. Unterschiede zwischen den durchschnittlichen Abweichungen für die neuen Bundesländer und für die alten Bundesländer vor 1990 sind indes nicht festzustellen

Tabelle 3.16: Einflussfaktoren auf die Prognosegüte von Sonntagsfragen, 1953-2009

	Modell 1	Modell 2	Modell 3	Modell 4	Modell 5
IfD Allensbach (1 = ja)		<i>Ref.</i>	<i>Ref.</i>	<i>Ref.</i>	<i>Ref.</i>
Infratest dimap (1 = ja)		1,725*	1,297	0,545	0,567
		(0,779)	(0,754)	(0,694)	(0,692)
TNS Emnid (1 = ja)		-2,662**	-1,031	-1,699*	-1,723*
		(0,834)	(0,816)	(0,750)	(0,749)
Forsa (1 = ja)		0,482	2,711**	2,082**	2,177**
		(0,831)	(0,826)	(0,759)	(0,759)
Forschungsgruppe Wahlen (1 = ja)		-2,844**	0,280	-0,454	-0,413
		(1,089)	(1,085)	(0,997)	(0,995)
Sonstige (1 = ja)		-4,118**	1,929	-0,188	-0,170
		(1,267)	(1,232)	(1,135)	(1,133)
Neue Bundesländer (1998-2005) (1 = ja)			<i>Ref.</i>	<i>Ref.</i>	<i>Ref.</i>
Alte Bundesländer (1990-2005) (1 = ja)			2,469***	2,150**	-0,542
			(0,749)	(0,688)	(1,431)
Gesamtdeutschland (1990-2005) (1 = ja)			-4,910***	-4,666***	-6,851***
			(0,666)	(0,612)	(1,194)
Alte Bundesländer (1949-1987) (1 = ja)			0,551	0,469	-5,623
			(3,238)	(2,953)	(4,150)
Zeitl. Abstand zur Wahl in Tagen				0,095***	0,095***
				(0,007)	(0,007)
Quadr. Zeitl. Abstand zur Wahl in Tagen				-0,000***	-0,000***
				(0,000)	(0,000)
Wahlbeteiligung in Prozent					0,501*
					(0,235)
Konstante	22,668***	23,426***	23,420***	16,454***	-21,026
	(1,707)	(1,720)	(2,687)	(2,490)	(17,830)
$\sigma(u)$	6,549	5,460	6,528	5,961	6,168
	(1,259)	(1,240)	(1,257)	(1,141)	(1,217)
$\sigma(e)$	9,506	9,330	8,955	8,219	8,205
	(0,167)	(0,164)	(0,157)	(0,144)	(0,144)
ρ	0,322***	0,324***	0,347***	0,345***	0,361***
	(0,084)	(0,084)	(0,088)	(0,087)	(0,091)
N	1.642	1.642	1.642	1.642	1.642
Zahl der Gruppen	16	16	16	16	16

Anmerkungen: Random-intercept-Maximum-Likelihood-Schätzungen für die abhängige Variable „adjustiertes A -Maß“, wobei der Wert Null einer perfekten Vorhersage entspricht und größere Werte eine schlechtere Prognosegüte anzeigen. Ebene 1: Umfragebasierte Vorhersagen; Ebene 2: Bundestagswahlen. Standardfehler in Klammern. Signifikant für $p < 0,05$ (*), $p < 0,01$ (**) und $p < 0,001$ (***) bei zweiseitigem Test.

Dieser Befund ist möglicherweise darauf zurückzuführen, dass seit der Wiedervereinigung nur für die publizierten Vorhersagen für Gesamtdeutschland die politische Gewichtung eingesetzt wird. Dies ist anhand der vorliegenden Daten nicht zu prüfen. Insofern wird vermutet, dass für die neuen und alten Bundesländer nach der Wiedervereinigung die ungewichteten Stimmenanteile berichtet werden. Ein Vergleich mit den Rohdaten der Forschungsgruppe Wahlen stützt diesen Befund, denn die aus den Rohdaten ermittelten Stimmenanteile weisen im Durchschnitt größere Prognosefehler auf als die Projektionen der Forschungsgruppe Wahlen. Trifft dies zu, ließe sich in Bezug auf umfragebasierte Voraussagen aus diesen Befunden zumindest ablesen, dass die Voraussagen vor der Wiedervereinigung im Durchschnitt größere Abweichungen zeigen als nach der Wiedervereinigung. Ein entsprechender Test zeigt jedoch, dass dieser Effekt nicht signifikant ist.

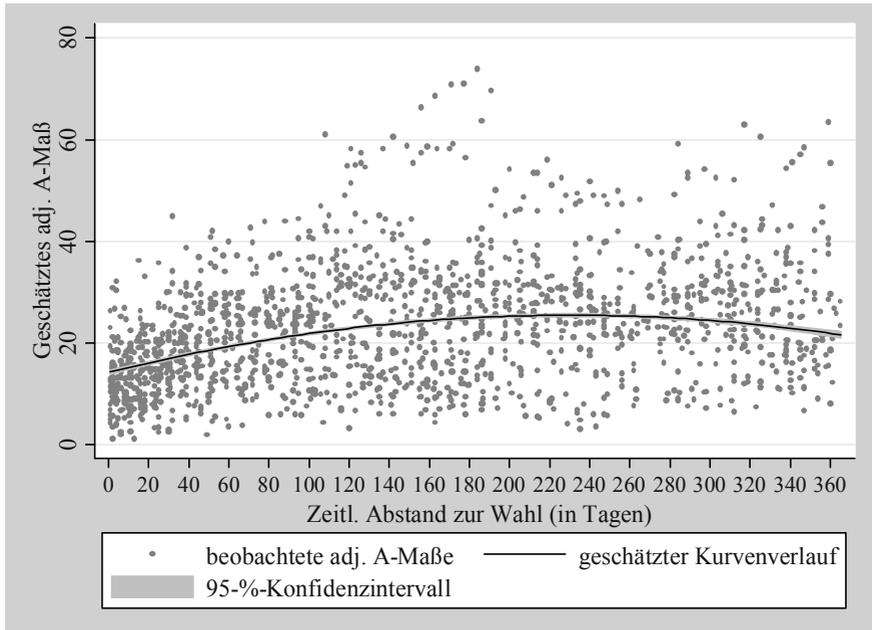
In Bezug auf den in *Hypothese 6* postulierten Zusammenhang muss festgehalten werden, dass dieser nicht bestätigt werden kann und die umfragebasierten Vorhersagen für die neuen Länder nicht schlechter abschneiden als diejenigen für die alten Bundesländer. Dieses Ergebnis ist jedoch insofern kritisch zu sehen, als davon ausgegangen werden muss, dass die erfassten Stimmenanteile für die neuen Bundesländer nicht politisch gewichtet wurden. Allerdings trifft dies dann auch für die Vorhersagen nur für die alten Bundesländer seit 1990 zu – und auch hier sind keine Unterschiede zwischen den Landesteilen hinsichtlich der Prognosegüte festzustellen.

Einen wesentlichen Einfluss, so die theoretischen Überlegungen, übt der zeitliche Abstand zur Wahl aus. Vermutet wird, dass mit zunehmender zeitlicher Nähe zum Wahltermin die Approximation des Wahlergebnisses besser wird. Weiterhin wird postuliert, dass die Prognosegüte mit abnehmender Grenzrate zunimmt (*Hypothese 2*). Der Test dieser Hypothese zeigt zwar wie erwartet einen höchst signifikanten positiven Zusammenhang zwischen dem zeitlichen Abstand zur Wahl und dem adjustierten A -Maß, jedoch weist der Kurvenverlauf nicht den vermuteten u-förmigen Zusammenhang auf, sondern einen umgekehrt u-förmigen Verlauf.⁴² Um dies zu verdeutlichen, wurde der Kurvenverlauf visualisiert

42 Um zu untersuchen, ob ein Modell mit dem quadrierten Term einem Modell mit ausschließlich linearem Term vorzuziehen ist, wurden zwei entsprechende Modelle spezifiziert und mit Hilfe

(Abbildung 3.28). Demnach steigt zwar die Prognosegüte mit zeitlicher Nähe zur Wahl. Der Wendepunkt für den Zusammenhang zwischen dem zeitlichen Abstand zur Wahl und dem adjustierten A -Maß liegt bei etwa 200 Tagen – das heißt, im Durchschnitt weisen die Sonntagsfragen, welche maximal ein Jahr vor einer Bundestagswahl veröffentlicht werden, 200 Tage vor der Wahl die größten Abweichungen auf.⁴³

Abbildung 3.28: Einfluss des zeitlichen Abstands zur Wahl auf die Prognosegüte



Anmerkungen: Der abgebildete geschätzte Kurvenverlauf resultiert aus der Schätzung, welche in Tabelle 3.16, Modell 5 dargestellt ist.

- eines *Likelihood-Ratio*-Tests miteinander vergleichen. Der Test zeigt, dass das Modell mit quadriertem Term höchst signifikant demjenigen nur mit linearem Term vorzuziehen ist.
- 43 Das Maximum errechnet sich durch Ableiten nach dem zeitlichen Abstand zur Wahl und Nullsetzen des Modells 5 in Tabelle 3.16 zugrunde liegenden Schätzformel. Dies ergibt:

$$x_z = \frac{-\beta_1}{2\beta_2}$$

Die Mehrebenenbetrachtung zeigt, dass die Varianz auf Bundestagswahlebene in Bezug auf das adjustierte A -Maß erheblich ist (vgl. auch Abbildung 3.27). Es wurde argumentiert, dass ein möglicher Effekt dieser Ebene die Höhe der Wahlbeteiligung darstellt. Es zeigt sich ein signifikanter positiver Effekt der Höhe der Wahlbeteiligung auf den Prognosefehler. Je höher die Wahlbeteiligung, desto größer die durchschnittliche Abweichung vom späteren Ergebnis. Weiterführende Analysen zeigen jedoch, dass dieser Effekt bei Variation der berücksichtigten Wahlen nicht robust ist. Die Wahlbeteiligung an der Bundestagswahl 2009 war mit 70,8 Prozent die bisher niedrigste bei Bundestagswahlen; die Vorhersage der Stimmenanteile durch die Sonntagsfragen war jedoch überdurchschnittlich genau. Berechnet man Modell 5 unter Ausschluss der Bundestagswahl 2009, so bestätigen sich zwar die Effekte aller anderen Variablen, nicht jedoch für die Höhe der Wahlbeteiligung. Bis 2005 findet man – wie theoretisch erwartet – einen höchst signifikanten negativen Effekt. *Hypothese 5* kann damit allenfalls als eingeschränkt bestätigt angesehen werden, wenn man die Bundestagswahl als „Ausreißer“ in Bezug auf die bisher ungewöhnliche Kombination hoher Prognosegüte und niedrige Wahlbeteiligung wertet. Zudem ist darauf hinzuweisen, dass aufgrund der vielen Beobachtungspunkte für die Bundestagswahl 2009 diese ein durchaus erhebliches Gewicht bei der Schätzung erfährt.

Auch die Varianz auf Ebene 2 bleibt erheblich, woraus ersichtlich ist, dass eine befriedigende Erklärung weiterer theoretischer Überlegungen und Identifikation einflussreicher wahlspezifischer Einflussfaktoren bedarf. Zugleich muss auf die erhebliche Veränderung der Effektgrößen der Regionsvariablen bei Hinzunahme der Wahlbeteiligung eingegangen werden. Bei Kontrolle auf diese verringert sich der Koeffizient des Dummys für die alten Bundesländer nach der Wiedervereinigung von 2,2 auf -0,5. Auch der Unterschied zwischen den neuen Bundesländern und den alten Bundesländern im Zeitraum vor der Wiedervereinigung verändert sich erheblich.

Dies ist vor dem Hintergrund des starken Zusammenhangs der Höhe der Wahlbeteiligung mit diesen Dummys zu sehen. Im Datensatz wurde die Wahlbeteiligung für die Zeit nach der Wiedervereinigung getrennt für die neuen und alten Bundesländer erfasst. Dadurch, dass die Wahlbeteiligung in den neuen Bundesländern deutlich niedriger ist als in den alten Bundesländern und in Westdeutschland vor der Wiedervereinigung noch einmal höher, verstärken sich diese Differenzen. Bei Konstanthalten der Höhe der Wahlbeteiligung werden die

Wahlresultate in den neuen Bundesländern folglich deutlich besser vorhergesagt – auch im Vergleich zu den Vorhersagen für Gesamtdeutschland. Dies ist ein weiteres Indiz dafür, dass *Hypothese 6* nicht zutrifft.

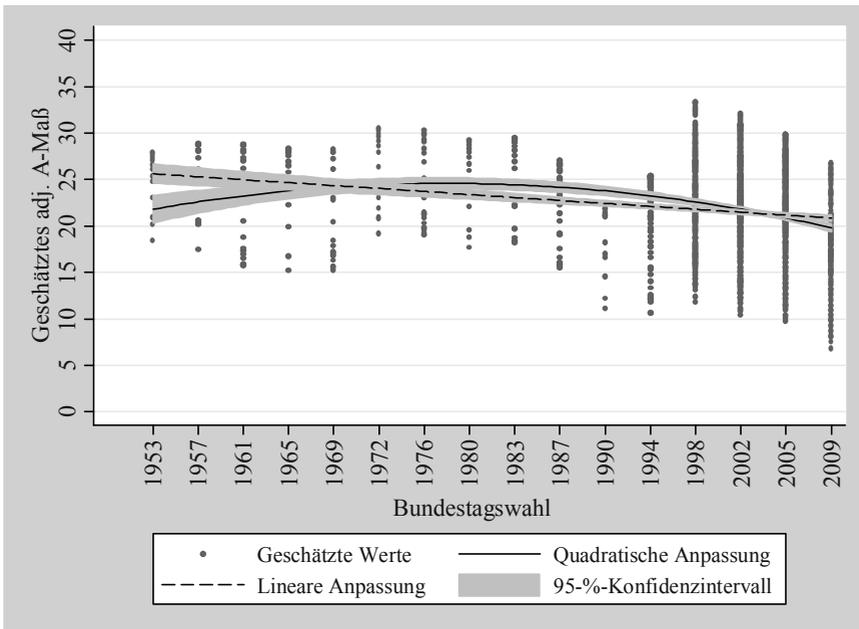
Eine zentrale Frage, die in diesem Kontext zu beantworten ist, ist diejenige, ob ein Trend im Hinblick auf die Prognosegüte vorliegt. Die theoretischen Überlegungen ergaben zwei Vermutungen: Im Zeitverlauf nimmt die Prognosegüte ab (*Hypothese 7a*) sowie: Es liegt ein umgekehrt u-förmiger Zusammenhang zwischen dem Zeitverlauf und der Prognosegüte vor (*Hypothese 7b*). Diese Hypothesen können auf Grundlage der Modelle in Tabelle 3.16 nicht beantwortet werden. Für die Beantwortung dieser Frage wurden zunächst die geschätzten adjustierten *A*-Maße aus Modell 5 berechnet. Weiterhin wurde eine Variable gebildet, die jeder Bundestagswahl eine fortlaufende Nummer zuweist, um die zeitliche Abfolge der Wahlen abzubilden – der wahre zeitliche Abstand zwischen den Bundestagswahlen wird hierbei ausgeblendet. Schließlich werden diese beiden Variablen miteinander in Beziehung gesetzt und mittels einer OLS-Regression der Kurvenverlauf geschätzt.

Zur graphischen Veranschaulichung sind in Abbildung 3.29 zum einen die geschätzten adjustierten *A*-Maße aus Modell 5 in Tabelle 3.16 dargestellt (graue Punkte). Zum anderen sind die lineare sowie die quadratische Anpassungslinie durch diese Punktwolke abgebildet. Die lineare Anpassungslinie bestätigt die Abnahme der Prognosegüte im Zeitverlauf nicht (*Hypothese 7a*). Die kurvilineare Anpassungslinie zeigt zudem entgegen der Erwartung einen u-förmigen Zusammenhang. Aufgrund der graphischen Analyse kann man vermuten, dass die kurvilineare Modellierung nicht angemessener ist als die lineare. Dies bestätigen entsprechende Modellvergleiche (nicht dargestellt). Als Ergebnis dieser Analyse kann festgehalten werden, dass die *Hypothesen 7a und 7b* als nicht bestätigt gelten können.

Kritisch einzuwenden ist hier jedoch erstens, dass diese einfache Vorgehensweise eine zeitreihenanalytische Betrachtung nicht ersetzen kann. Dies erscheint auf Grundlage von 16 Beobachtungszeitpunkten zu diesem Zeitpunkt wenig sinnvoll. Zweitens zeigt sich auch hier die durchgehend festgestellte Variation nach Bundestagswahl. Führt man dieselben Analysen unter Ausschluss der Bundestagswahl 2009 durch, bestätigt sich *Hypothese 7b*. Hier zeigt sich ein deutlicher umgekehrt u-förmiger Zusammenhang. Auch diese Analyse legt nahe, dass die Bundestagswahl 2009 aus Sicht der Demoskopien ein durchaus bemerk-

kenswertes Jahr war: Entgegen dem Trend, der bis zum Jahr 2005 galt, wonach sowohl ein zeitlicher Trend als auch die niedrige Wahlbeteiligung ein schwieriges Umfeld für die Vorhersage der Stimmenanteile boten, gelang den Demoskopern mit den Wahlabsichtsfragen eine gute Approximation des Wahlergebnisses.

Abbildung 3.29: Die Prognosegüte von umfragebasierten Vorhersagen im Zeitverlauf, 1953-2009



Anmerkungen: Die geschätzten Werte resultieren aus Modell 5 in Tabelle 3.16.

Bisher vernachlässigt wurden die methodischen Überlegungen zum Einfluss der Stichprobengröße sowie dem Befragungszeitraum. Die hierzu korrespondierenden *Hypothesen 1, 3a* und *3b* werden nun in gesonderten Analysen getestet. Dies ist notwendig, da nur für einen Teil der erfassten umfragebasierten Vorhersagen diese Informationen zur Verfügung stehen (vgl. auch Tabelle 3.15). Diese Informationen erstrecken sich zudem nicht auf den gesamten Zeitraum, sondern es liegen hierzu Informationen erst seit der Bundestagswahl 1994 vor. Aufgrund

dieser geringen Gruppenzahl erscheint die Anwendung von Mehrebenenmodellen nicht sinnvoll. Deshalb wurden OLS-Regressionen spezifiziert, welche auf die fixen Effekte der jeweiligen Bundestagswahl kontrollieren. Bedingt durch den Verzicht auf die Anwendung von *random-intercept*-Modellen kann die bereits als einflussreich identifizierte Höhe der Wahlbeteiligung in diesen Schätzungen nicht berücksichtigt werden. Die Ergebnisse sind in Tabelle 3.17 dargestellt.

Zunächst wird – analog zum bisherigen Vorgehen – ein Modell geschätzt, welches nur die Effekte der Bundestagswahlen enthält (Modell 1, Tabelle 3.17). Hier wird noch einmal der bereits belegte Befund des zentralen Einflusses von Wahlspezifika unterstrichen. Allein die vier berücksichtigten Bundestagswahlen erklären 49,1 Prozent der beobachteten Varianz des adjustierten A -Maßes. Weiterhin ist zu sehen, dass die Ergebnisse der Bundestagswahl 2009 durch die Meinungsforschungsinstitute höchst signifikant besser approximiert wurden als die der Bundestagswahlen von 1994. Die Ergebnisse der Bundestagswahl 2005 hingegen wurden deutlich schlechter vorhergesagt als die der Bundestagswahl 1994.

Im nächsten Schritt werden für die Wahlen 1994 bis 2009 die Befunde für die *random-intercept*-Modelle mit Ausnahme des Effekts der Wahlbeteiligung repliziert (Modell 2, Tabelle 3.18). Diese Analyse reproduziert im Wesentlichen die Befunde der mehrebenenanalytischen Betrachtung (Modell 4, Tabelle 3.16). Demnach sind die Unterschiede nach Instituten eher gering. Auffällig hingegen ist das umgedrehte Vorzeichen für den Prognosefehler in Gesamtdeutschland. Während das *random-intercept*-Modell hier eine signifikant bessere Approximation im Vergleich zu den neuen Bundesländern ausweist, zeigt diese Analyse eine tendenziell etwas schlechtere Approximation.

Der kurvilenare Effekt des zeitlichen Abstands zur Wahl wird in dieser Analyse ebenfalls reproduziert. Die Berechnung standardisierter Regressionskoeffizienten (nicht dargestellt) bestätigt zudem den bedeutenden Einfluss des zeitlichen Abstands zur Wahl. Der lineare Term ist der stärkste Einflussfaktor, gefolgt vom quadrierten zeitlichen Abstand zur Wahl.

Tabelle 3.17: Einflussfaktoren auf die Prognosegüte umfragebasierter Vorhersagen, 1994-2009

	Modell 1	Modell 2	Modell 3	Modell 4	Modell 5
Bundestagswahl 1994 (1 = ja)	<i>Ref.</i>	<i>Ref.</i>	<i>Ref.</i>	<i>Ref.</i>	<i>Ref.</i>
Bundestagswahl 1998 (1 = ja)	2,900 (2,866)	-2,739 (1,977)	-2,831 (1,944)	0,786 (2,248)	0,723 (2,254)
Bundestagswahl 2002 (1 = ja)	-2,691* (1,284)	-1,923 (1,124)	-1,994 (1,141)	0,571 (1,626)	0,155 (1,677)
Bundestagswahl 2005 (1 = ja)	6,725*** (1,177)	6,006*** (1,010)	5,975*** (1,007)	10,027*** (1,538)	10,152*** (1,578)
Bundestagswahl 2009 (1 = ja)	-13,196*** (1,096)	-12,266*** (1,008)	-12,265*** (1,007)	-8,316*** (1,541)	-8,299*** (1,575)
IFD Allensbach (1 = ja)		<i>Ref.</i>	<i>Ref.</i>	<i>Ref.</i>	<i>Ref.</i>
Infratest dimap (1 = ja)		3,537** (1,208)	3,362** (1,275)	3,796* (1,533)	4,883** (1,615)
TNS Emnid (1 = ja)		0,771 (1,181)	1,177 (1,337)	1,251 (1,354)	1,488 (1,317)
Forsa (1 = ja)		3,046** (1,150)	3,190** (1,170)	3,551* (1,403)	4,777** (1,637)
Forschungsgruppe Wahlen (1 = ja)		2,464* (1,253)	2,305 (1,269)	3,623* (1,558)	4,917** (1,740)
Sonstige (1 = ja)		2,734* (1,279)	2,485 (1,333)	2,814 (1,454)	2,873* (1,422)
Neue Bundesländer (1998-2009) (1 = ja)		<i>Ref.</i>	<i>Ref.</i>	<i>Ref.</i>	<i>Ref.</i>
Alte Bundesländer (1994-2009) (1 = ja)		19,622*** (2,206)	20,097*** (2,546)	25,330*** (2,258)	25,418*** (2,269)
Gesamtdeutschland (1994-2009) (1 = ja)		2,400 (1,568)	3,076 (2,100)	5,800*** (1,589)	6,519*** (1,644)
Zeitl., Abstand zur Wahl in Tagen		0,092*** (0,008)	0,093*** (0,008)	0,083*** (0,008)	0,082*** (0,008)
Quadr. zeitl. Abstand zur Wahl in Tagen		-0,000*** (0,000)	-0,000*** (0,000)	-0,000*** (0,000)	-0,000*** (0,000)
Stichprobenumfang			-0,000 (0,000)	-0,000 (0,000)	-0,001 (0,000)
Befragungszeit in Tagen				0,073 (0,122)	
Am Wochenende befragt (1 = ja)					1,915 (1,146)
Konstante	23,894*** (1,054)	11,530*** (2,215)	11,404*** (2,268)	5,347* (2,557)	3,871 (2,656)
R ²	0,491	0,707	0,708	0,763	0,765
N	665	665	665	594	594

Anmerkungen: OLS-Schätzungen für die abhängige Variable „adjustiertes A-Maß“, wobei der Wert Null einer perfekten Vorhersage entspricht und größere Werte eine schlechtere Prognosegüte anzeigen. Dargestellt sind die unstandardisierten Koeffizienten; robuste Standardfehler in Klammern. Signifikant für $p < 0,05$ (*), $p < 0,01$ (**) und $p < 0,001$ (***) bei zweiseitigem Test.

Die Modelle 3 bis 5 in Tabelle 3.17 nehmen die bisher nicht berücksichtigten Variablen in den Blick. Die Länge des Befragungszeitraums kann aufgrund von Multikollinearität nicht gemeinsam mit dem Dummy, ob sich der Befragungszeitraum auch über das Wochenende erstreckt, getestet werden, weshalb hierfür getrennte Modelle ausgewiesen sind. Die Analysen zeigen deutlich, dass keine der methodisch als einflussreich vermuteten Variablen einen signifikanten Effekt auf den Prognosefehler ausübt. Insbesondere der Stichprobenumfang spielt offenbar keine Rolle – dies widerspricht zwar *Hypothese 1*, geht jedoch konform mit den Ergebnissen aus vergleichbaren Studien für die USA (vgl. hierzu Abschnitt 3.2).

Dieser Nulleffekt, so könnte man vermuten, ist auf eine nur geringe Varianz der Stichprobengröße zurückzuführen. Im Rahmen dieser Analyse werden die Punktschätzer der umfragebasierten Vorhersagen verwendet und nicht die Vertrauensbereiche. Um letztere zu halbieren, muss dem Wurzel-N-Gesetz zufolge der Stichprobenumfang vervierfacht werden. Vor diesem Hintergrund erscheint die Streuung des Stichprobenumfangs gering (Tabelle 3.15). Weiterhin wird vermutet, dass der Einfluss des Stichprobenumfangs im Vergleich mit der politischen Gewichtung deutlich zu gering ist, um hier einen messbaren Effekt zu zeigen.

Auch die *Hypothesen 3a* und *3b* können mit den durchgeführten Analysen nicht gestützt werden. Beide Effekte sind äußerst gering und nicht signifikant. Auch in Bezug hierauf kann vermutet werden, dass diese überlagert werden durch die eingesetzten Gewichtungsverfahren. Diese Analysen abschließend fasst Tabelle 3.18 die zentralen Befunde hinsichtlich der Einflussfaktoren auf die Prognosegüte von umfragebasierten Vorhersagen zusammen.

Tabelle 3.18: Zusammenfassung der Ergebnisse zu den Einflussfaktoren auf die Prognosegüte umfragebasierter Vorhersagen

	Einflussfaktor	Vermuteter Effekt auf die Vorhersagegenauigkeit	Beobachteter Effekt auf die Vorhersagegenauigkeit
1	Zunahme der Stichprobengröße	steigt	kein Effekt
2	Zeitlicher Abstand zum Wahltermin sinkt	steigt mit abnehmender Grenzrate	steigt, jedoch besteht ein u-förmiger Zusammenhang
3a	Zunahme der Länge des Befragungszeitraums	steigt	kein Effekt
3b	Befragung an Werktagen und am Wochenende	steigt	kein Effekt
4a	house effects	liegen vor	liegen teilweise vor
4b	Institutsspezifischer Parteibias	kein Effekt	kein Effekt
5	Höhere Wahlbeteiligung	steigt	sinkt leicht (bei Berücksichtigung der Wahl von 2009); steigt (ohne Berücksichtigung der Wahl von 2009)
6	Vorhersagen für neue Bundesländer	sinkt	kein eindeutiger Effekt
7a	Zeitverlauf	sinkt	steigt leicht (bei Berücksichtigung der Wahl von 2009); sinkt (ohne Berücksichtigung der Wahl von 2009)
7b	Zeitverlauf	umgekehrt u-förmig	linearer Effekt (bei Berücksichtigung der Wahl von 2009); umgekehrt u-förmig (ohne Berücksichtigung der Wahl von 2009)

Anmerkung: Eigene Zusammenstellung.

4 Ein Prognosemodell für Bundestagswahlen

Aus den bisherigen Analysen ist für die Möglichkeiten der Prognose von Wahlergebnissen erstens festzuhalten, dass umfragebasierte Vorhersagen zwar zumindest für die beiden großen Parteien CDU/CSU und SPD eine gute Prognosequalität aufweisen. Damit sind sie aus empirischer Perspektive durchaus besser als ihr öffentlicher Ruf. Allerdings beruhen sie auf intersubjektiv nicht nachvollziehbaren Gewichtungungsverfahren der erhobenen Wahlabsicht, weshalb sie sich wissenschaftlich nicht fruchtbar nutzen lassen. Zweitens hat sich für experimentelle Märkte gezeigt, dass diese im Hinblick auf die empirische Prognosegüte mit den umfragebasierten Vorhersagen zumindest konkurrenzfähig sind. Für Deutschland sind jedoch weitere Analysen zu deren Leistungsfähigkeit notwendig. Dies betrifft vor allem die Beeinflussung der Wahlbörsen durch umfragebasierte Vorhersagen. Drittens wurde gezeigt, dass die Entwicklung von Prognosemodellen im Rahmen der empirischen Wahlforschung in Deutschland bisher noch nicht nachdrücklich verfolgt wurde.

Aus der prognostischen Performanz sowohl von Sonntagsfragen als auch von Wahlbörsen könnte man schließen, dass der Mangel an Prognosemodellen irrelevant erscheint. Hier wird jedoch die Auffassung vertreten, dass die Entwicklung entsprechender Modelle fruchtbar sein könnte. In den USA haben diese Bestrebungen maßgeblich die empirische Wahlforschung und hierbei insbesondere die Theoriebildung stimuliert. Insofern kann man auch für Deutschland durch den Ausbau theoriebasierter Prognosemodelle einen Anstoß für die Theoriebildung erwarten. Prognosen offenbaren deutlicher als post-hoc-Analysen Erklärungsdefizite und erlauben somit eine zielgerichtete Weiterentwicklung theoretischer Überlegungen.

Zwar liegt mit dem Instrumentarium experimenteller Märkte bereits ein wissenschaftliches Prognoseinstrument vor, dieses liefert jedoch keine Erkenntnisgewinne für die empirische Wahlforschung. Der Grund hierfür ist darin zu sehen, dass Wahlbörsen auf der These der effizienten Informationsaggregation durch

Wettbewerbsmärkte beruhen. Auf dieser Grundlage ist die Analyse von Informationsverarbeitungsprozessen möglich, jedoch nicht des Wählerverhaltens.

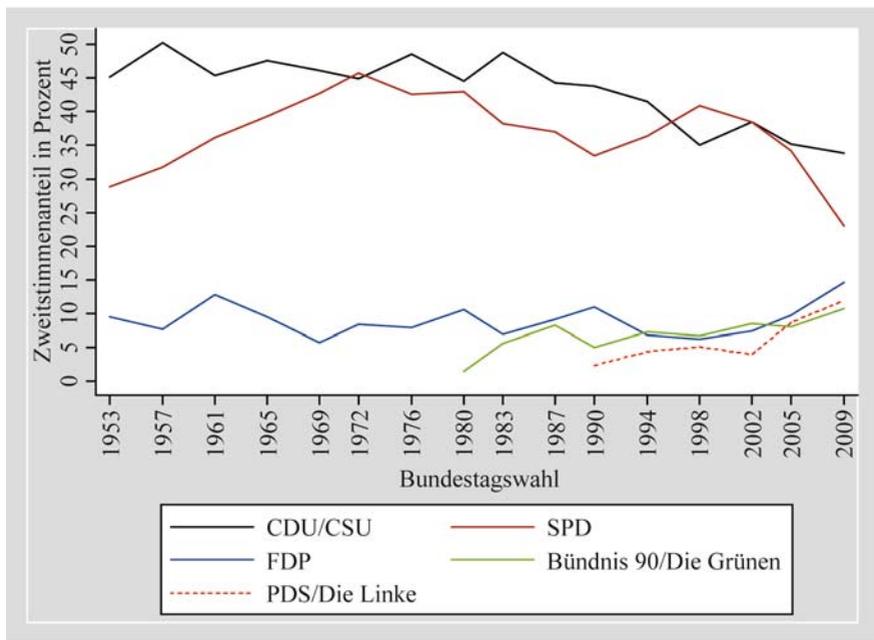
Weiterhin wird an dieser Stelle dafür plädiert, in den Sozialwissenschaften im Allgemeinen am Ziel der Prognose sozialer Phänomene festzuhalten. Die Vorhersage von Wahlergebnissen scheint hier als Betätigungsfeld besonders geeignet, da die empirische Wahlforschung dank ihres einheitlichen Forschungsprogramms und der kumulativen Ausrichtung profunde Erkenntnisse über das Wählerverhalten erarbeitet hat und das Ziel der Prognose des Wahlverhaltens auch zum Forschungsprogramm der Disziplin gezählt wird (Falter/Rattinger 1983: 321).

Schließlich erscheint es auch aus professionspolitischer Sicht lohnend, Prognosemodelle zu erarbeiten. Das öffentliche Interesse an Wahlen im Allgemeinen und auch an Wahlprognosen im Speziellen ist erheblich, wie beispielsweise die zunehmende Verbreitung von Sonntagsfragen (Brettschneider 1991, 1996) zeigt. Auch die wissenschaftliche wie öffentliche Resonanz auf das Modell von Gschwend und Norpoth (2001) sowie die Wahlbörsen unterstreichen dies. Insofern wäre dies aus Sicht der empirischen Wahlforschung eine lohnende Investition im Hinblick auf die öffentliche Wahrnehmung der Disziplin.

Im Rahmen dieses vierten Teils der Arbeit wird auf Grundlage theoretischer Überlegungen ein Vorschlag für ein erstes Prognosemodell präsentiert, das die Vorhersage von Zweitstimmenanteilen einzelner Parteien ermöglicht (Abbildung 4.1). Das gesuchte Modell soll dabei verschiedenen Anforderungen gerecht werden, die im Weiteren knapp skizziert werden.

Es wird ein möglichst einfaches Modell gesucht. Diese Zielsetzung ist mit den Befunden von Gigerenzer und Kollegen zu begründen, die gezeigt haben, dass brauchbare Prognosen nicht überkomplex sein dürfen (Czerlinski et al. 1999; Gigerenzer 2007; Goldstein/Gigerenzer 2008; Todd et al. 2008). Weiterhin muss in diesem Zusammenhang auf die schmale Datenbasis verwiesen werden. Es haben bisher 17 Bundestagswahlen stattgefunden, Bündnis 90/Die Grünen sind überdies erst seit 1983 im Deutschen Bundestag vertreten. Die PDS/Die Linke ist erst seit vier Legislaturperioden Mitglied des Deutschen Bundestages. Dies zwingt zum einen zur Beschränkung auf wenige Variablen und führt zum anderen – gerade im Hinblick auf die Vorhersage von Stimmenanteilen von Bündnis 90/Die Grünen sowie von PDS/Die Linke – zu vermutlich größeren Unsicherheiten.

Abbildung 4.1: Zweitstimmenanteile relevanter Parteien bei Bundestagswahlen, 1953-2009



Anmerkungen: Dargestellt sind die bei den angegebenen Wahlen zum Deutschen Bundestag erzielten Zweitstimmenanteile der Parteien CDU/CSU, SPD, FDP, Bündnis 90/Die Grünen sowie PDS/Die Linke. Die Ergebnisse beziehen sich bis 1990 auf das frühere Bundesgebiet, ab 1990 auf das gesamte Bundesgebiet. Die Verbindungslinien sollen die Lesbarkeit der Grafik erhöhen. Die Daten wurden den Berichten des Bundeswahlleiters entnommen (Der Bundeswahlleiter 2009a; Statistisches Bundesamt 2005).

Weiterhin soll ein intersubjektiv nachvollziehbares und ausbaufähiges Modell entwickelt werden, welches sich nach Möglichkeit auf prozessproduzierte Informationen stützt. Damit sollen explizit die Probleme der Umfrageforschung bei der Aggregation individueller Informationen vermieden werden. Gleichwohl wird damit nicht behauptet, dass die Entwicklung eines umfragebasierten Prognosemodells nicht viel versprechend sein könnte. Einen Vorschlag für das Vorgehen in diesem Fall haben Yu und Lam (1997) vorgelegt. Auch die Entwicklung möglicher zeitreihenanalytischer Modelle auf Grundlage der geäußerten Wahlabsichten ist denkbar und sollte in der Zukunft angestrebt werden (vgl. für

die USA z. B. Norpoth 1996).⁴⁴ Die empirischen Befunde zu den von Meinungsforschungsinstituten veröffentlichten Vorhersagen zeigen, dass sich Sonntagsfragen durchaus als Grundlage hierfür eignen. Die Gründe für die Verwendung prozessproduzierter Daten an dieser Stelle ist, dass diese im Zeitablauf mit einer höheren Validität erfasst werden können, in der Regel leichter zu erheben sind und die methodischen Probleme bei Befragungen (Non-Response, soziale Erwünschtheit etc.) nicht mit in die theoretischen Überlegungen einzufließen brauchen. Dies müsste im Rahmen der Entwicklung von Prognosemodellen auf Basis von Umfragedaten berücksichtigt werden.

Anzustreben ist zudem die Präsentation eines Modells, das möglichst früh vor einer anstehenden Wahl Prognosen ermöglicht. Dies bedingt, dass kurzfristige Faktoren und mögliche *last minute swings* nicht in hinreichendem Maße erfasst werden können. Damit orientiert sich das hier vorgeschlagene Modell zumindest in den Grundzügen an der Vorgehensweise von Gschwend und Norpoth (2001) für die Prognose des gemeinsamen Stimmenanteils der Regierungskoalitionsparteien und an den in Abschnitt 2.3.1 diskutierten regressionsbasierten Prognosemodellen für die USA.

Im Folgenden werden zunächst mögliche Modellparameter erörtert (Abschnitt 4.1). Die Ableitung der Modellparameter stützt sich auf theoretische Überlegungen, welche sowohl zentrale Einsichten der empirischen Wahlforschung berücksichtigen als auch auf statistische Gesetzmäßigkeiten zurückgreifen. Weiterhin werden Operationalisierungsvorschläge unterbreitet, welche auf ihre Brauchbarkeit hin untersucht werden. In Abschnitt 4.2 wird abschließend ein erster Modellvorschlag präsentiert, dessen empirische Performanz retrospektiv getestet und mit der Prognosegüte von wahlabsichtsfragenbasierten Umfragen verglichen wird.

44 Die Forschungsgruppe Wahlen veröffentlicht neben der politisch gewichteten Projektion auch ungewichtete „Stimmungen“, welche hierzu beispielsweise als Grundlage herangezogen werden könnten.

4.1 Die Modellvariablen

Grundsätzlich problematisch an der Konstruktion eines Modells zur Vorhersage von Wahlergebnissen auf Grundlage theoretischer Ansätze zum Wählerverhalten ist, dass es bisher nicht gelungen ist, ein „theoretisch fundiertes und empirisch überprüftes Modell des Wählerverhaltens, das mit hinreichender Präzision bestimmt für welche der fünf im Bundestag vertretenen Parteien ein Wähler zu stimmen gedenkt“ (Gschwend/Norpoth 2001: 475) vorzulegen.

Merrill III und Grofman (1999) haben zwar einen Vorschlag für ein integriertes Modell zur Erklärung des Wählerverhaltens unterbreitet, eine profunde empirische Prüfung des Erklärungsansatzes für Deutschland steht jedoch noch aus. Allerdings weisen die Befunde von Debus (2007) für ausgewählte Bundestagswahlen auf ein großes Erklärungspotenzial des Ansatzes hin. Das Modell von Merrill III und Grofman (1999) basiert wesentlich auf räumlichen Theorien des Wahlverhaltens (für einen Überblick vgl. bspw. Enelow/Hinich 1984, 1990; Hinich/Munger 1997), wonach insbesondere die *policy*-Distanzen zwischen Idealpositionen der Wähler und den Positionen der Parteien zu berücksichtigen sind. Ein retrospektiver Test des Ansatzes mit den Daten der vorliegenden deutschen Wahlstudien ist nicht möglich, da die hierfür notwendigen Informationen zumeist nicht erhoben wurden (zur Dokumentation von Datenproblemen der Wahlstudien vgl. etwa Schnell/Kohler 1995). Als Ausgangspunkt zur Ableitung relevanter Variablen für das Prognosemodell kann dieser Ansatz daher nicht herangezogen werden. Allerdings beruht dieses Modell im Wesentlichen auf dem Versuch der Integration zentraler Einsichten verschiedener Ansätze innerhalb eines einheitlichen theoretischen Rahmens.

Im Folgenden wird deshalb auf verschiedene theoretische Überlegungen zurückgegriffen, um hieraus zentrale Einflussfaktoren abzuleiten. Aus der vermuteten theoretischen Schwäche folgt die Erwartung, dass nicht alle wesentlichen Einflüsse erfasst werden können. Deshalb wird versucht, allgemeine statistische Gesetzmäßigkeiten zur Schätzung von Populationsanteilen in das Prognosemodell zu integrieren. Im Einzelnen werden hierbei die von Mosteller (1965) und Zipf (1965) begründeten Gesetzmäßigkeiten diskutiert.

4.1.1 Normalwahl

Als Normalwahl wird ein hypothetisches Wahlergebnis bezeichnet, das zustände käme, wenn alle Wähler gemäß ihrer langfristigen Parteibindung abstimmen würden. Das Konzept der Normalwahl wurde von Converse (1966) eingeführt und beruht auf dem sozialpsychologischen Ansatz zur Erklärung des Wählerverhaltens (Campbell et al. 1954, 1964; Miller/Shanks 1996). Das Konzept wurde zudem in einer theoretisch gehaltvolleren Variante von Fiorina (1981) im Rahmen politökonomischer Modelle reformuliert. Diese theoretische Fundierung wird im Folgenden aufgegriffen.

4.1.1.1 Die langfristige Parteibindung im sozialpsychologischen Ansatz

Das zentrale Postulat des sozialpsychologischen Ansatzes zur Erklärung des Wählerverhaltens ist die Existenz einer langfristig stabilen affektiven Bindung der Wähler an bestimmte Parteien. Converse (1969: 144) zufolge ist diese als „psychische Parteimitgliedschaft“ aufzufassen. Diese langfristige Parteibindung, Parteiidentifikation genannt, wird im Sozialisationsprozess geprägt und im Lebensverlauf durch die Teilnahme an Wahlen verstärkt. Das Konzept der Parteiidentifikation ist mit der Annahme des entscheidenden Einflusses von relevanten Bezugspersonen bereits im Kindesalter in der sozialpsychologischen Bezugsgruppentheorie verwurzelt (grundlegend hierzu z. B. Cartwright/Zander 1968; Merton/Kitt 1950) und wurde von Converse (1969) konkretisiert.⁴⁵

Die Parteiidentifikation wird im sozialpsychologischen Ansatz als exogen angenommen. Weiterhin wird postuliert, dass sie den anderen als einflussreich angesehenen Faktoren, vor allem der Sachfragen- und der Kandidatenorientierung vorgelagert ist. Schließlich ist die „Filterfunktion“ (Gabriel/Keil 2008) der Parteiidentifikation zu nennen, welche die Wahrnehmung von politischen Parteien und Kandidaten sowie von Sachfragen beeinflusst. Dem von Campbell et al. (1964: 31ff.) postulierten Kausalitätstrichter zufolge beeinflusst demnach die

45 Eine alternative Fundierung des Konzepts auf Grundlage von Tajfels (1981) Theorie der sozialen Identität hat Greene (1999) vorgelegt. Die Interpretation der Parteiidentifikation als *information shortcut* durch Popkin (1991) beruht auf dieser Basis.

gegebene Parteiidentifikation die Kandidaten- und Sachfragenorientierung und schließlich das Wählerverhalten.

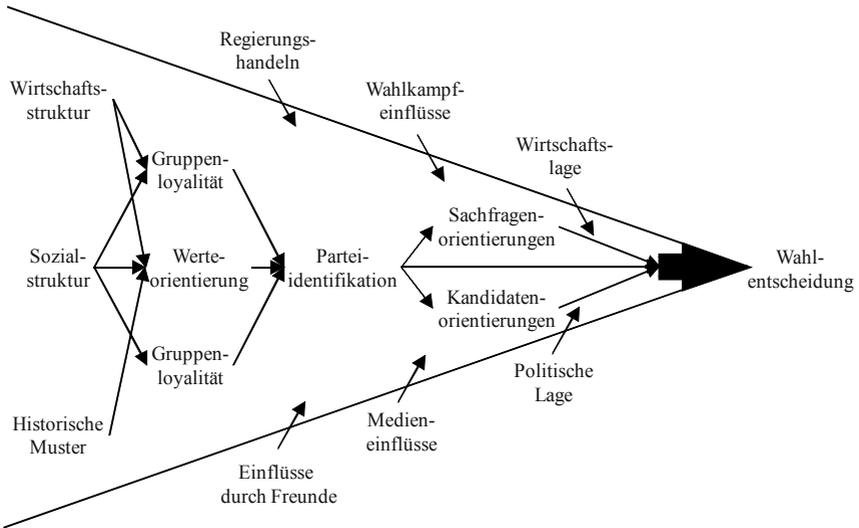
Im Hinblick auf die Möglichkeit der Übertragung des Konzepts auf Deutschland wurde eine Kontroverse geführt, diese jedoch bald mit dem Ergebnis beigelegt, dass eine fruchtbare Anwendung des sozialpsychologischen Ansatzes im Allgemeinen und des Konzepts der Parteiidentifikation im Speziellen möglich sei (Falter 1977, 1984; Falter et al. 2000). In der Folge hat sich ein Standardinstrument in der Umfrageforschung zur Erfassung der Parteiidentifikation durchgesetzt (z. B. Schoen/Weins 2005).⁴⁶ Zentral erscheint hierbei, dass diese Operationalisierung explizit versucht, die Unabhängigkeit der Parteiidentifikation von der Wahlentscheidung zu erfassen.

An dieser Konzeption wurde insbesondere die Annahme der Exogenität der Parteiidentifikation kritisiert. Markus und Converse (1979) griffen diese Kritik auf und entwickelten den Ansatz weiter, indem sie die Parteiidentifikation im Rahmen eines Pfadmodells endogenisieren. Diese Weiterentwicklungen des sozialpsychologischen Grundmodells führten zu einem erweiterten Kausalitätstrichter wie er in Abbildung 4.2 dargestellt ist.

In einem alternativen Modell der Parteiidentifikation hat Fiorina (1981: 89) die langfristige Parteibindung als „running tally of retrospective evaluations“ aufgefasst. Demnach unterliegt die Entwicklung der Parteiidentifikation einem Lernprozess, der bei Fiorina als Bayesianisches Updating modelliert wird (Pappi/Shikano 2007: 75ff.).

46 Die Formulierung der Frage nach der Parteiidentifikation lautet üblicherweise: „Viele Leute in der Bundesrepublik neigen längere Zeit einer bestimmten politischen Partei zu, obwohl sie ab und zu einmal eine andere Partei wählen. Wie ist es bei Ihnen? Neigen Sie – ganz allgemein gesprochen – einer bestimmten politischen Partei zu? Wenn ja, welcher?“ (Roth 1998: 43).

Abbildung 4.2: Der Kausalitätstrichter im sozialpsychologischen Ansatz zur Erklärung des Wählerverhaltens



Quelle: Eigene Darstellung nach Dalton (2005: 178).

Beiden Ansätzen gemeinsam ist, dass sie davon ausgehen, dass die frühere Parteiidentifikation die aktuelle beeinflusst. Allerdings sieht Fiorina (1981) anders als die Vertreter des sozialpsychologischen Ansatzes die Parteiidentifikation wesentlich durch die subjektive Leistungsbewertung der Parteien in der Vergangenheit beeinflusst und weniger geprägt durch die Primärsozialisation (vgl. hierzu auch Pappi/Shikano 2007: 44).

Mit dem Ansatz der Normalwahl verlagerte Converse (1966) das Konzept der Parteiidentifikation auf die Makroebene.⁴⁷ Eine Normalwahl liegt demnach vor, wenn das Wahlergebnis deckungsgleich mit den langfristigen Parteibindungen des Elektorats ausfällt. Ziel von Normwahlanalysen ist es, das Elektorat in verschiedene Untergruppen zu zerlegen, um tiefere Wahlanalysen zu ermöglichen

47 Eine alternative Begründung des Normwahlkonzepts als Ergebnis einer „politisierten Sozialstruktur“ haben Pappi und Shikano (2002) vorgelegt. Damit wird eine Verbindung zum *cleavage*-Ansatz hergestellt (vgl. hierzu Abschnitt 4.1.2.1).

(Falter/Rattinger 1983: 343). Weiterhin sind mit Normalwahlanalysen zumeist Klassifikationen von verschiedenen Wahlen verbunden, etwa ob generell eine Normwahlwahl oder eine kritische Wahl (hierzu: Key 1955) vorliegt.

Grundlegende Idee des Normwahlkonzepts ist weniger, dass Kurzfristfaktoren wie Sachfragen- oder Kandidatenorientierung keine Bedeutung bei der Wahlentscheidung der Wähler zukommt. Vielmehr – so die Überlegung – heben sich diese Einflüsse im Aggregat gegenseitig auf (Miller 1979). Converse' (1966: 10) Vorschlag zur Feststellung des Vorliegens einer Normalwahl ist eine Durchschnittsbildung über vorhergehende Wahlen. Die bei jeder Wahl auftretenden Abweichungen vom „wahren“ Normwahlwahlergebnis sollten sich durch die Mittelung über mehrere Wahlen hinweg ausgleichen. Grundlage der Mittelung stellt hierbei die in Umfragen gemessene Parteiidentifikation dar (Falter/Rattinger 1983: 326).

Für Deutschland haben Falter und Rattinger (1983, 1986) das Normwahlkonzept erstmals angewendet. Sie folgern aus ihren Analysen, dass es sich für die Untersuchung des deutschen Wahlsystems anwenden lässt. Sie zeigen, dass die Bundestagswahl 1983 entgegen der damalig vorherrschenden Interpretation zumindest für die CDU/CSU und die SPD als Normalwahl zu klassifizieren ist (Falter/Rattinger 1986).

In Bezug auf die individuelle Wahlentscheidung ist aus empirischer Sicht ebenfalls ein stabiler und dominanter Effekt der Parteiidentifikation festzustellen. Dieser variiert zwar hinsichtlich der Stärke wahlspezifisch, ist jedoch im Zeitablauf stabil geblieben (Gabriel 1997, 2001; Jagodzinski/Kühnel 1990; Kellermann 2008; Kellermann/Rattinger 2005). Für die Wahlentscheidung zugunsten der FDP zeigt sich, dass der Einfluss der Parteiidentifikation eine geringere Relevanz aufweist als bei den beiden großen Parteien.

Auf Makroebene ist für Deutschland ein Absinken der langfristigen Partei-bindungen zu konstatieren, wenn auch dieser *dealignment*-Prozess nicht dazu geführt hat, dass Parteibindungen irrelevant geworden wären.⁴⁸ So weisen immer noch rund 60 Prozent des deutschen Elektorats eine Parteiidentifikation auf

48 Für Analysen der Entwicklung der Parteiidentifikation auf Aggregatebene hat sich der Begriff *macropartisanship* etabliert (Green et al. 1998; MacKuen et al. 1989). Das Konzept der Normalwahl hingegen hat die Analyse von Wahlergebnissen in Abhängigkeit der Parteiidentifikation zum Ziel. International vergleichende Befunde zum *macropartisanship* liefern beispielsweise die Arbeiten von Box-Steffensmeier und Smith (1996) oder Dalton (2005: 177ff.).

(Arzheimer/Schoen 2005; Dalton 2005: 189; Dalton/Bürklin 2003; Schoen/Weins 2005).⁴⁹ Demnach ist, auch wenn auf individueller Ebene ein anhaltende Relevanz der Parteiidentifikation zu konstatieren ist, auf Makroebene davon auszugehen, dass Abweichungen vom Normalwahlergebnis häufiger und größer werden und somit langfristige Parteibindungen möglicherweise einen nicht so dominanten Einfluss auf Aggregat- wie auf Individualebene ausüben.

Aus diesen Überlegungen und empirischen Befunden folgt, dass ein wesentlicher Einfluss langfristiger Parteibindungen auf das Wahlergebnis zu erwarten ist. Dies gilt, wie mit dem Normalwahlkonzept hergeleitet, nicht nur für die individuelle Wahlentscheidung, sondern ist auch auf Aggregatenebene zu erwarten.

4.1.1.2 Operationalisierung

Als langfristige Parteibindung wird im Folgenden der durchschnittliche Zweitstimmenanteil einer Partei bei den letzten drei Bundestagswahlen aufgefasst. Diese Operationalisierung schließt an das theoretische Konzept Fiorinas (1981: 89) an, wonach die Parteiidentifikation als „running tally of retrospective evaluations“ interpretiert wird. Die Konzeption von Converse (1966) wird ebenfalls abgedeckt, wonach eine Approximation an ein Normalwahlergebnis über Durchschnittsbildung vergangener Wahlen erreicht wird. Hier wird die langfristige Parteibindung folglich als „Vorentscheidung (standing decision) in Form der durchschnittlichen Entscheidung bei zurückliegenden Wahlen“ (Pappi/Shikano 2002: 446) interpretiert und korrespondiert damit mit dem Vorgehen von Gschwend und Norpoth (2001: 482) für ihr Modell zur Vorhersage des gemeinsamen Stimmenanteils der amtierenden Regierungsparteien.

Diese Konzeption wirft in zweierlei Hinsicht Probleme auf. Erstens reduziert dieses Vorgehen die berücksichtigbare Fallzahl erheblich, da somit für die CDU/CSU, die SPD sowie die FDP erst Daten ab der Bundestagswahl 1961 vorliegen, für die PDS/Die Linke gar erst ab der Bundestagswahl 2002. Zweitens erscheint fraglich, ob eine Fortschreibung über die Zeit der Wiedervereinigung möglich erscheint. Gerade in Bezug auf das Ausmaß der Parteiidentifikation in

49 Mögliche Gründe für den Prozess der Auflösung von langfristigen Parteibindungen werden in Abschnitt 3.1 diskutiert.

den neuen Bundesländern ist festzuhalten, dass diese – erwartungsgemäß – deutlich geringer ausfällt als in den alten Bundesländern (Rattinger 1994).

Das im Rahmen dieser Arbeit verwendete Vorgehen bezieht im Fall von weniger als drei vorliegenden Wahlen in der Vergangenheit die jeweils vorliegenden Ergebnisse mit ein. So werden für die Bundestagswahl 1953 die Stimmenanteile von 1949 herangezogen und für die Bundestagswahl von 1957 der Durchschnitt der Ergebnisse von 1949 und 1953. Analog wird im Hinblick auf Bündnis 90/Die Grünen sowie die PDS/Die Linke zu späteren Zeitpunkten verfahren. Im Hinblick auf die Wiedervereinigung werden immer die Ergebnisse für das gesamte Bundesgebiet verwendet. Das heißt für die Bundestagswahl von 1990 werden als langfristige Parteibindungen die durchschnittlichen Ergebnisse aus den Bundestagswahlen von 1980, 1983 und 1987 für die alten Bundesländer verwendet und für die Vorhersage der Stimmenanteile für das gesamte Bundesgebiet genutzt. Dementsprechend ist eine Vorhersage des Stimmenanteils der PDS/Die Linke für die Bundestagswahl 1990 aufgrund fehlender Daten nicht möglich. Im Fall der Bundestagswahl 2009 argumentieren Norpoth und Gschwend (2009), dass die neue Partei Die Linke besonders zu Lasten der SPD gehe und daher weichen sie von ihrem festgelegten Vorgehen für diese Wahl ab und verwenden nur das letzte Ergebnis der SPD als langfristige Parteibindung. Dieser ad-hoc-Modifikation wird hier, um der Stringenz des Modells treu zu bleiben, nicht gefolgt, wenngleich vermutet werden kann, dass diese die Genauigkeit des Modells zumindest für die Wahl 2009 erhöht, das der durchschnittliche Stimmenanteil der SPD über die letzten drei Wahlen aufgrund ihrer Erfolge 1998 und 2002 höher liegt als der Stimmenanteil von 2005.

Ein hieraus ersichtlicher zentraler Nachteil des Modellvorschlags – wie auch dem von Gschwend und Norpoth (2001) – ist, dass Veränderungen des Parteiensystems nicht zufriedenstellend erfasst werden und neue Parteien nicht angemessen berücksichtigt werden können (vgl. Klein 2005b). Dies trifft auch für die Fusion der PDS mit der WASG zu. In diesem Fall werden die vorherigen Ergebnisse der PDS für die fusionierte Partei nicht fortgeschrieben. Für die Prognose des Stimmenanteils der Partei Die Linke wird hingegen das Ergebnis bei der Bundestagswahl 2005 herangezogen – die Wahl bei welcher der Parteizusammenschluss erstmals als Bündnis angetreten ist. Der Grund hierfür wird darin gesehen, dass bis zur Fusion der beiden Parteien, die PDS primär in den neuen Bundesländern verwurzelt war, während mit der WASG eine Absplitterung der

SPD in den alten Bundesländern hinzukam. Dadurch entstand gewissermaßen eine neue gesamtdeutsche Partei links von der SPD.

Nun werden die bivariaten Zusammenhänge zwischen der langfristigen Parteibindung und dem Zweitstimmenanteil für die einzelnen Parteien betrachtet (Tabelle 4.1). Die Korrelationen variieren stark nach betrachteter Partei. So ist für die CDU/CSU und die SPD eine moderate Abhängigkeit ihres Stimmenanteils von den vergangenen Wahlen zu konstatieren, während für die FDP ein geringer und Bündnis 90/Die Grünen sowie PDS/Die Linke ein starker Zusammenhang mit vergangenen Ergebnissen zu erkennen ist.

Tabelle 4.1: Zusammenhang der langfristigen Parteibindung mit dem Zweitstimmenanteil, 1953-2009

	Durchschnittliches Ergebnis der Partei bei den letzten drei Bundestagswahlen	N
CDU/CSU	0,345	16
SPD	0,382	16
FDP	-0,206	16
Bündnis 90/Die Grünen	0,554	8
PDS/Die Linke	0,882	5

Anmerkungen: Angegeben ist jeweils der Bravais-Pearson-Korrelationskoeffizient r . Auf Signifikanztest wird verzichtet, da es sich um die Ergebnisse von Vollerhebungen handelt.

Aus diesen bivariaten Analysen ist festzuhalten, dass sich nur teilweise Zusammenhänge in der erwarteten Stärke zeigen. Vor dem Hintergrund der Befunde von Falter und Rattinger (1983, 1986) aus den 1980er Jahren überraschen jedoch die Ergebnisse für die FDP nicht, für die CDU/CSU hingegen schon. Sie belegen, dass zwar in Bezug auf die Ergebnisse der beiden großen Parteien Normwahlergebnisse vorliegen, nicht jedoch im Hinblick auf die Wahlergebnisse der FDP. Vergegenwärtigt man sich einmal mehr die empirischen Befunde zur Relevanz der Parteiidentifikation auf die Wahlentscheidung auf Individualebene, dann wird deutlich, dass die berichtete Orientierung der Wähler an ihrer Parteineigung insbesondere für die CDU/CSU sowie die SPD getestet wurden und die

kleineren Parteien häufig unberücksichtigt blieben (so bspw. bei Kellermann/Rattinger 2005). Weiterhin zeigen Analysen zum strategischen Wählen, dass insbesondere die FDP Nutznießerin von Leihstimmen ist (Jesse 1988). Insofern muss davon ausgegangen werden, dass bei der FDP erstens größere Schwankungen auftreten und deren Ergebnisse aufgrund taktischen Wahlverhaltens bezogen auf ihre Wählerschaft mit einer langfristigen Parteibindung an die FDP systematisch zu hoch ausfallen.

Für die SPD ist weiterhin der bedeutende Einfluss der Bundestagswahl 2009 festzuhalten. Der Bravais-Pearson-Korrelationskoeffizient r beträgt 0,540 für die SPD sofern man die jüngste Wahl nicht berücksichtigt. Insofern erscheint aus empirischer Perspektive die von Norpoth und Gschwend (2009) vorgenommene Modifikation durchaus plausibel.

4.1.2 Politisierte Sozialstruktur

Die Analyse des Einflusses sozialstruktureller Variablen auf das Wahlverhalten zur Erklärung langfristig stabiler Muster weist eine bis in die Anfänge der empirischen Wahlsoziologie zurückreichende Tradition auf (vgl. bspw. Müller 1998). Zentral hierfür ist das Postulat einer existenten politisierten Sozialstruktur, worunter Pappi (1986) einen Zusammenhang zwischen gesellschaftlichen Konfliktlinien und Parteienkonstellationen versteht.

In theoretischer Hinsicht wird hier auf verschiedene Ansätze verwiesen, jedoch gilt als klassischer Bezugspunkt wahlsoziologischer Analysen der *cleavage*-Ansatz (Lipset/Rokkan 1967). Diesem Ansatz zufolge bilden sich im Rahmen des Nationenbildungsprozesses starre soziale Konfliktlinien (*cleavages*) heraus, die sich im Laufe der Zeit verfestigen (Abschnitt 4.1.2.1). Die These der Verfestigung der politisierten Sozialstruktur wurde in jüngerer Zeit vielfacher Kritik ausgesetzt und verschiedene modernisierungstheoretische Ansätze postulieren eine zunehmende Auflösung der von Lipset und Rokkan (1967) als starr angenommenen *cleavages*. Abschnitt 4.1.2.2 skizziert diese theoretische Kritik. Zentrale empirische Befunde zum Zusammenhang sozialstruktureller Merkmale mit dem Wählerhalten werden schließlich präsentiert (Abschnitt 4.1.2.3), um daraus Schlüsse für die Verwendung potenzieller Variablen für ein Prognosemodell ziehen zu können (Abschnitt 4.1.2.4).

4.1.2.1 Entstehung und Verfestigung der politisierten Sozialstruktur

Ein Ausgangspunkt der Analyse der sozialstrukturellen Verankerung des Wahlverhaltens ist Lipsets und Rokkans (1967) *cleavage*-Ansatz. Ziel des Ansatzes ist es, die Entstehung von Parteiensystemen zu beschreiben und zu klassifizieren. Im Kern postulieren die Autoren die Herausbildung starrer Konfliktlinien (*cleavages*) entlang tief greifender sozioökonomischer oder wertebasierter Konflikte, welche sich im Zeitverlauf verfestigt haben und seit etwa den 1920er Jahren stabil bleiben (*freezing*-Hypothese).

Ein *cleavage* liegt vor, wenn der gesellschaftliche Konflikt dauerhaft ist und von sozialen Gruppen getragen wird, welche die gegenseitigen Interessen vertreten. Die gegenseitigen Interessen müssen weiterhin von kollektiven Akteuren (etwa Verbänden) artikuliert und vertreten werden und es müssen dauerhafte Koalitionen zwischen diesen kollektiven Akteuren mit politischen Parteien bestehen. Schließlich legitimiert eine Ideologie die Positionen der an den Konflikten beteiligten Akteure (vgl. bspw. Gabriel/Keil 2008).

Die Anzahl der konkurrierenden Parteien wird durch die Anzahl der *cleavages* bestimmt, welche sich im Laufe des Modernisierungsprozesses herausgebildet und verfestigt haben. Nach Lipset und Rokkan (1967) sind vier grundlegende *cleavages* zu unterscheiden, welche kultureller oder sozioökonomischer Art sind und in den einzelnen Staaten unterschiedlich stark ausgeprägt auftreten können. Diese Konfliktlinien sind diejenigen zwischen Staat und Kirche, Zentrum und Peripherie, Stadt und Land sowie Arbeit und Kapital.

Zwar haben Lipset und Rokkan (ebd.) selbst nicht den Anspruch erhoben, einen Ansatz zur Erklärung des Wählerverhaltens vorzulegen, jedoch wurde ihr Beitrag häufig in dieser Weise rezipiert. Zudem beziehen sich Wahlanalysen, welche soziodemographische Merkmale berücksichtigen, in aller Regel auf diesen Ansatz. Im Hinblick auf das Wahlresultat wurde aus dem Ansatz geschlossen, dass zumindest für bestimmte Bevölkerungsgruppen ein homogenes Wahlverhalten zu erwarten ist.

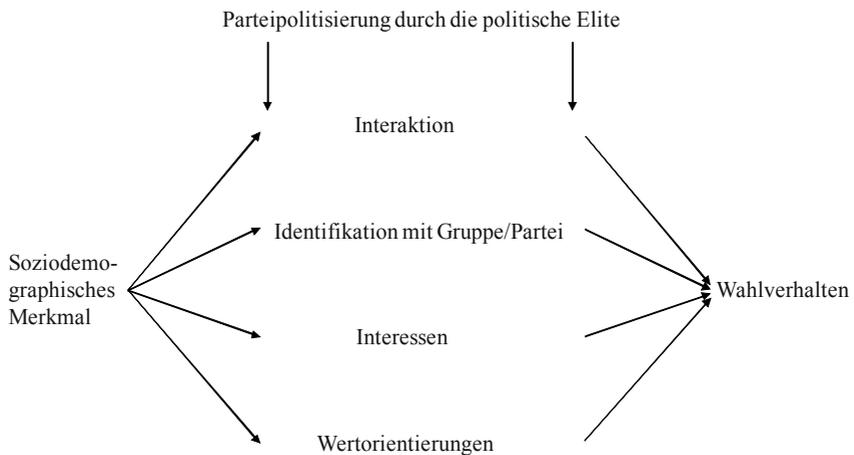
Allerdings weist der Ansatz lediglich auf Beziehungen zwischen Aggregatgrößen hin und liefert keine Erklärung für den Zusammenhang zwischen der Sozialstruktur und dem Wahlergebnis. Entsprechend wurde kritisiert, dass *cleavages* eher als Metapher für die Existenz einer politisierten Sozialstruktur zu sehen sind, denn als brauchbarer Erklärungsansatz (Elff 2006: 53). Für die Erklärung des

Zusammenhangs zwischen sozialstrukturellen Faktoren und dem Wahlverhalten muss auf „Vermittlungsmechanismen“ (Schoen 2005b: 137) verwiesen werden.

Diese können beispielsweise aus dem mikrosoziologischen Erklärungsansatz (Berelson et al. 1986; Lazarsfeld et al. 1955) hergeleitet werden. So ist nach diesem Interaktionsmodell zu erwarten, dass Personen in ähnlichen sozialen Lagen gleichartige politische Interessen entwickeln und entsprechend auch ein ähnliches Wahlverhalten an den Tag legen. Ebenfalls kann erwartet werden, dass – wie auch im sozialpsychologischen Ansatz (Campbell et al. 1964) – die Entwicklung von Wertorientierungen im Sozialisationsprozess mit der sozialen Position variiert. Diese Kongruenz im Hinblick auf verschiedene Werte führt zu homogenem Wahlverhalten. Zu betonen ist, dass in beiden Fällen „individualistisches *issue voting* für die Gruppenkonformität des Wahlverhaltens“ (Schoen 2005b: 151, Hervorhebung im Original) sorgt.

Ein weiterer Vermittlungsmechanismus ist die Weitergabe von Parteipräferenzen über soziale Netzwerke, wonach gleichförmiges Wählen als Nebeneffekt individuellen Handelns erscheint. Eine Übersicht über theoretisch herleitbare Vermittlungsmechanismen ist in Abbildung 4.3 dargestellt.

Abbildung 4.3: Vermittlungsmechanismen zwischen soziodemographischen Merkmalen und dem Wahlverhalten



Anmerkungen: Eigene Darstellung nach Schoen (2005b).

Alternativ zu diesen nachfrageseitigen Vermittlungsmechanismen richtet Elff (2006, 2009) den Blick auf die Rolle politischer Angebote zur Herleitung der Relevanz verschiedener sozialstruktureller Merkmale für das Wählerverhalten. Unabhängig vom gewählten Erklärungsansatz lässt sich mit allen Ansätzen die Relevanz soziodemographischer Merkmale begründen und auf Aggregatebene die nach dem *cleavage*-Ansatz erwarteten Partei-Wähler-Allianzen herleiten.

Gefolgert werden kann aus diesen Überlegungen die Vermutung fester Allianzen zwischen sozialen Gruppen, deren Interessenvertretungen (etwa Verbände) und Parteien und damit auch den Wählern. Postuliert wird folglich ein gruppenkonformes Wahlverhalten, wobei diese Gruppen entweder entlang von Verteilungskonflikten (Arbeit versus Kapital) oder Wertekonflikten (Staat versus Kirche) verlaufen können. Folgt man der *freezing*-These des *cleavage*-Ansatzes ist eine stabile sozialstrukturelle Prägung des Wahlverhaltens im Zeitablauf zu erwarten.

4.1.2.2 Dynamik der politisierten Sozialstruktur

Seit den 1970er Jahren wird die Hypothese der Verfestigung westeuropäischer Parteiensysteme nach Abschluss des Nationenbildungsprozesses (Lipset/Rokkan 1967: 50) verstärkt bezweifelt. Hingegen wird vermehrt die Auflösung der als starr angenommenen Parteienkonstellationen postuliert. In Bezug auf das Wahlverhalten wird einerseits von der Abschwächung von Verteilungskonflikten und damit dem Verlust der Prägekraft des *class voting* ausgegangen. Andererseits wird die Entstehung neuer, wahlrelevanter Allianzen beschrieben.

So diagnostiziert Inglehart (1971, 1977, 1983) einen fundamentalen gesellschaftlichen Wertewandel in den westlichen Demokratien, welcher sich durch die Hinwendung zu postmateriellen Werten äußert. Diese „stille Revolution“ (Inglehart 1977) wird zunächst von der jungen Generation getragen, weitet sich jedoch im Zeitverlauf aus und verlagert sich schließlich auf die Führungselite. Im Prinzip wird der Wertewandel in diesem Ansatz als evolutionäres Ergebnis fortschreitender ökonomischer Entwicklung interpretiert. Aufbauend auf der Maslowschen Bedürfnispyramide (Maslow 1943) postuliert Inglehart die Herausbildung postmaterialistischer Werte in der Nachkriegsgeneration, da deren physiologischen Bedürfnisse befriedigt sind. Unter postmaterialistischen Werten versteht Inglehart etwa die Orientierung an Umweltschutz, Frieden und politischer

Partizipation, während für Materialisten Werte wie Wohlfahrt und Sicherheit handlungsleitend sind. Hieraus, so die These, resultiert eine neue gesellschaftliche Konfliktlinie. Auf Grundlage dieser Konzeption wird insbesondere versucht, die Ausdifferenzierung des linken politischen Lagers sowie die Entstehung grüner Parteien zu interpretieren.

Eng verknüpft mit der Wertewandelthese Ingleharts ist die Anwendung von Lebensstilkonzepten zur Analyse des Wählerverhaltens (z. B. Gluchowski 1987). Lebensstilansätze gehen davon aus, dass der Zusammenhang zwischen der sozioökonomischen Position und der Lebensführung in westlichen Demokratien im Zeitablauf zunehmend loser geworden ist (für eine Übersicht vgl. Burzan 2005: 97ff.). Demnach wird das ästhetisch expressive Verhalten, worunter auch das Wahlverhalten fällt, immer weniger von der sozialen Lage als von Werten und Zielen beeinflusst (Georg 1998: 98). Diese Ansätze versuchen insgesamt, das Wählerverhalten auf ein komplexeres „Zusammenspiel von Sozialstrukturfaktoren, sozialer Integration, Wertvorstellungen und Konsumstilen“ (Gabriel/Keil 2008) zurückzuführen, als es in den klassischen Ansätzen des Wählerverhaltens angelegt ist.

Auch die Individualisierungsthese postuliert ähnliche Entwicklungen. Grundlegend ist hier die Vermutung, dass im Zuge fortschreitender Industrialisierung das Individuum von der Fremd- zur Selbstbestimmung übergeht. Seit den 1950er Jahren, so das Postulat, wird eine qualitativ neue Stufe dieses Prozesses erreicht, der sich im Zwang zur reflexiven Lebensführung und der Pluralisierung von Lebensstilen äußert (Beck 1983; 1986: 121ff.). Zentral hierfür ist, dass die Individuen nun zunehmend selbst verantwortlich für die eigene Identitäts- und Sinnstiftung sind, welche zuvor von sozialen Gruppen als kollektive Orientierungsschemata vorgegeben wurde. Als treibende Kräfte dieser Entwicklung sieht Beck (1986: 122) die Wohlstandsmehrung, den Ausbau wohlfahrtstaatlicher Elemente und vor allem die Bildungsexpansion.

Die Folgerungen aus diesen Überlegungen für das Wahlverhalten, resultieren erstens im Postulat der Abnahme der Erklärungskraft soziodemographischer Variablen im Zeitablauf (Ohr 2005; Schnell/Kohler 1995). Dieser Verlust der Prägungskraft des Wahlverhaltens durch soziodemographische Merkmale – vor allem der sozialen Lage – wird sowohl auf Aggregat- als auch auf Individualebene

gesehen. Zweitens werden unterschiedliche neue *alignments* zwischen sozialen Gruppen und Parteien vorgebacht:⁵⁰ So sei die Kluft zwischen an materialistischen Werten orientierten Wählern auf der einen und an postmaterialistischen Werten orientierten Wählern auf der anderen Seite nun ebenso zu berücksichtigen wie Konflikte zwischen den Geschlechtern, Generationen oder milieuspezifisches Wählerverhalten.

Diese Ansätze sehen sich vielfältiger Kritik ausgesetzt. Ingleharts Wertewandelthese negiert etwa die Existenz von Lebenszykluseffekten und geht von stabilen Kohorteneffekten aus. Die Individualisierungsthese erlaubt keine Ableitung eindeutiger Hypothesen aufgrund zahlreicher widersprüchlicher Aussagen (bspw. Schnell/Kohler 1995). Auf Grundlage von Lebensstilansätzen ist es allenfalls möglich, eine Wählertypologie zu erstellen, Aussagen über Ursache-Wirkungs-Zusammenhänge werden jedoch nicht getroffen. Für die Ableitung wesentlicher Einflussfaktoren sind diese Ansätze daher ungeeignet.⁵¹ In Bezug auf die die Diskussion um die Entstehung neuer *cleavages* wird mittlerweile mehrheitlich die Auffassung vertreten, dass von gesellschaftlichen Konfliktlinien im engen Sinne nicht zu sprechen ist (Schoen 2005b: 150), da die genannten Kriterien von Lipset und Rokkan (1967) von keiner der vorgebrachten neuen Allianzen erfüllt werden.

4.1.2.3 Empirische Befunde

Für Deutschland wird in Anlehnung an Lipset und Rokkan (1967) das sozioökonomische *cleavage* (Arbeit versus Kapital) sowie das konfessionell-religiöse *cleavage* (Katholiken versus Protestanten beziehungsweise Kirche versus Staat) traditionell als relevant angesehen.⁵² Damit sind der Konflikt zwischen Arbeitern

50 Dies wird häufig als Gegensatz zwischen alter und neuer Politik aufgefasst (Hildebrandt/Dalton 1977; Müller 1999: 112ff.).

51 Diese Auffassung, wonach die Anwendung von Lebensstilansätzen im Rahmen der Wahlforschung als wenig fruchtbar anzusehen ist, wird im Übrigen weitgehend geteilt (Oedegaard 2000; Otte 1997).

52 Einschlägige Arbeiten in international vergleichender Perspektive haben unter anderen Bartolini und Mair (2007), Clark und Lipset (2001), Dalton (2005), Evans (1999, 2000), Franklin et al. (1992), Kriesi (1998) sowie Nieuwbeerta (1995) vorgelegt.

und Selbstständigen und der Konflikt zwischen den Konfessionen beziehungsweise Gläubigen und Säkularen angesprochen.

Pappi (1977: 196) argumentiert, dass Deutschland ein asymmetrisches Parteiensystem aufweise, „weil eine Partei der gewerkschaftlichen Wirtschafts Ideologie nicht einer primär ökonomisch konservativen Partei gegenübersteht, sondern einer religiös traditionellen Partei“. Als parteipolitische Repräsentanten des sozioökonomischen *cleavage* ist auf Arbeiterseite die SPD (und neuerdings auch die PDS/Die Linke) und auf Seiten der Selbstständigen die FDP zu nennen. Das religiös-konfessionelle *cleavage* ist durch die Bindung der Katholiken an die CDU/CSU und der Konfessionslosen an Bündnis 90/Die Grünen gekennzeichnet (Pappi/Shikano 2002).

In empirischer Hinsicht ist für den Zeitraum von 1949 bis 2009 ein erheblicher sozialer Wandel in Bezug auf die angesprochene sozialstrukturelle Zusammensetzung der Bevölkerung festzustellen. Dies betrifft die Veränderung der Berufsstruktur ebenso wie die Entwicklung der Konfessionszugehörigkeit oder der Religiosität (z. B. Müller 1999: 98ff.). Umstritten ist jedoch, ob der postulierte Bedeutungsverlust sozialstruktureller Merkmale für das Wahlverhalten empirisch zutrifft.

Besonders kontrovers diskutiert wird dies in Bezug auf die *sozioökonomische Konfliktlinie* (*class voting*) festzustellen ist zunächst eine nach wie vor deutliche Überrepräsentation der traditionellen Wählerklientel der Arbeiter beziehungsweise der Selbstständigen bei SPD und FDP. Ebenfalls ist auf individueller Ebene die Relevanz der sozioökonomischen Stellung anhaltend einflussreich (Arzheimer/Schoen 2007; Emmert/Roth 1995; Kellermann/Rattinger 2005; Müller 1998, 2000; Pappi/Shikano 2002). Gleichwohl sind die empirischen Befunde zum Bedeutungsverlust des *class voting*, wie es in Folge der Individualisierungsthese zu erwarten wäre, nicht eindeutig. Schnell und Kohler (1995, 1997) finden die vermutete abnehmende Prägung der Wahlabsicht durch sozioökonomische Variablen, während Müller (1997, 1998, 2000) in einer Reanalyse diese Tendenz nicht in der von Schnell und Kohler vorgetragenen Prägung findet. Dies ist insbesondere auf unterschiedliche Operationalisierungen der Klassenzugehörigkeit zurückzuführen.

Müller (1998) zeigt außerdem, dass in empirischer Hinsicht wenig Unterstützung für die These der zunehmenden Relevanz von alternativen Konfliktlinien zu finden ist. Vielmehr sind auch die neuen politischen Gruppierungen

stark sozialstrukturell verankert. Insgesamt lässt sich festhalten, dass für Deutschland nach wie vor von einer erheblichen Bedeutung sozioökonomischer Merkmale für das Wahlverhalten auszugehen ist (zusammenfassend Pappi 2002; Roth 1998: 35; Schoen 2005b).

Eindeutiger sind die empirischen Befunde zum *religiös-konfessionellen Konflikt*. Es zeigt sich – auch über die Zeit weitgehend unverändert –, dass Katholiken überproportional häufig die CDU/CSU wählen. Allerdings hat sich in der Zeit eine Bedeutungszunahme der Religiosität herausgestellt. Mittlerweile ist weniger die Konfession an sich als vielmehr die Intensität dieser entscheidend. So besteht insbesondere eine stabile Allianz von praktizierenden Katholiken und der CDU/CSU (Arzheimer/Schoen 2007; Emmert/Roth 1995; Jagodzinski/Quandt 1997; Pappi 1985, 2002; Pappi/Shikano 2002; Schoen 2005b).

Für das weitere Vorgehen wird aus dieser Diskussion gefolgert, dass die für Deutschland traditionell als relevant angesehenen soziodemographischen Merkmale trotz entsprechend postulierter Auflösungstendenzen weiterhin bedeutend sind. Zudem zeigt sich, dass bisher weder überzeugende theoretische Konzepte noch hinreichend empirische Befunde vorliegen, um neue Konfliktlinien berücksichtigen zu können. Dies gilt insbesondere für die Zusammenhänge auf der Makroebene. Deshalb soll die Brauchbarkeit des *class voting* sowie der konfessionellen wie religiösen Orientierung des Elektorats zur Prognose der Stimmenanteile der einzelnen Parteien geprüft werden. Im Einzelnen wird erwartet, dass die Stimmenanteile von CDU/CSU mit dem Anteil der Katholiken sowie vor allem der katholischen Kirchgänger positiv variiert. Der Anteil Konfessionsloser sollte indes positiv den Stimmenanteil von Bündnis 90/Die Grünen beeinflussen. Für die SPD und die PDS/Die Linke ist zu erwarten, dass ihr Stimmenanteil mit sinkendem Arbeiteranteil abnimmt. Der Stimmenanteil der FDP sollte dagegen mit dem Anteil Selbstständiger variieren.

4.1.2.4 Operationalisierung

Die Daten zum Test der vermuteten Zusammenhänge auf Grundlage des Ansatzes zur politisierten Sozialstruktur stützen sich weitgehend auf prozessproduzierte Daten der Bevölkerungsstatistik des Statistischen Bundesamtes. Zunächst wurden Zahlen zur Bevölkerungsentwicklung Deutschlands dem Jahrbuch des Statisti-

schen Bundesamts (2008) entnommen, um die entsprechenden Anteile der einzelnen relevanten Gruppen berechnen zu können. Anzumerken ist hierbei, dass sich die Angaben bis 1990 auf das frühere Bundesgebiet beziehen und seit 1990 auf das gesamte Bundesgebiet. Weiterhin umfassen die Daten die gesamte Bevölkerung und nicht die Wahlberechtigten. Daraus resultieren mögliche Verzerrungen, da die hier enthaltenen ausländischen Bürger sowie die Jugendlichen unter 18 Jahren gerade im Hinblick auf die hier interessierenden sozialstrukturellen Merkmale sich von der Gruppe der Wahlberechtigten unterscheiden.

Für zukünftige Modelle, welche sich ebenfalls auf entsprechende sozialstrukturelle Merkmale stützen – sollte sich dies als fruchtbar erweisen – könnte auf Daten des Mikrozensus zurückgegriffen werden. Diese Möglichkeit wurde hier nicht verfolgt, da der Mikrozensus erst seit 1973 zur Verfügung steht und damit die Bundestagswahlen von 1953, 1957, 1961, 1965, 1969 und 1972 nicht berücksichtigt werden könnten. Angesichts der ohnehin wenigen Datenpunkte erschien dieses Vorgehen nicht sinnvoll.

Alternativ könnte man auf andere Umfragedaten, etwa die deutschen Wahlstudien zurückgreifen. Diesen Weg haben Schnell und Kohler (1995) verfolgt, um im Trend die sozialstrukturelle Zusammensetzung der Bundesrepublik erfassen zu können.⁵³ Gegen dieses Vorgehen spricht, dass die Erhebungsinstrumente im Laufe der Zeit häufiger gewechselt haben, wie Schnell und Kohler (ebd.) betonen und zudem hier die Verwendung von Vorwahlumfragedaten möglichst vermieden werden soll.

Zunächst wird ein enger Zusammenhang zwischen dem Anteil an Katholiken beziehungsweise der religiös aktiven Katholiken und dem Zweitstimmenanteil der CDU/CSU vermutet. Die Deutsche Bischofskonferenz (2009b) stellt Daten zur Mitgliederentwicklung der katholischen Kirche in Deutschland zur Verfügung, auf die hier zurückgegriffen wurde. Diese Angaben wurden mit Berichten des Statistischen Bundesamts verglichen und für die Jahre vor 1965 ergänzt. Hierbei sind nur geringfügige Unterschiede hinsichtlich der absoluten Angaben festzustellen, welche sich nicht auf die Anteilsberechnung auswirken. Weiterhin ist in der kirchlichen Statistik der Deutschen Bischofskonferenz

53 Der Datensatz ist über GESIS – Leibniz-Institut für Sozialwissenschaften zu beziehen (ZA-Nr. 2661). Weder die vorgenannten Personen noch das GESIS tragen irgendeine Verantwortung für die Analyse oder Interpretation der Daten in dieser Arbeit.

(2009a) der durchschnittliche Besuch der Eucharistiefiern ausgewiesen. Diese Angaben werden als Indikator für die Intensität der Religiosität der Katholiken herangezogen. Nähere Angaben zur Erhebungsmethode sind im Bericht der Deutschen Bischofskonferenz nicht enthalten, weshalb eine Beurteilung des Maßes hinsichtlich seiner Validität nicht möglich ist.⁵⁴

Empirisch zeigen sich die erwarteten starken Zusammenhänge zwischen dem Katholikenanteil und dem Anteil von katholischen Gottesdienstbesuchern an der Bevölkerung mit dem Zweitstimmenanteil von CDU/CSU. Der Bravais-Pearson-Korrelationskoeffizient beträgt $r = 0,883$ ($N = 16$) beziehungsweise $r = 0,786$ ($N = 16$). Aufgrund der unklaren Erhebungsmethode der Gottesdienstbesuche wird im Weiteren auf den Katholikenanteil an der Bevölkerung zurückgegriffen.

Weiterhin wird erwartet, dass der Zweitstimmenanteil von Bündnis 90/Die Grünen positiv mit dem Anteil Konfessionsloser variiert. Angaben zu den Konfessionslosen waren über das Statistische Bundesamt nicht zu beziehen. Deshalb wurde auf den von Schnell und Kohler (1995) bereit gestellten Datensatz zurückgegriffen. Dieser enthält hierzu Angaben für die alten Bundesländer bis 1992. Um die Lücke zu schließen, wurden der Anteil Konfessionsloser anhand eigener Berechnungen auf Grundlage der bereits in Abschnitt 3 verwendeten Politbarometer-Daten berechnet. Es liegen hier zumeist getrennte Datensätze für Ost- und Westdeutschland vor. Um den Anteil für Gesamtdeutschland zu schätzen, wurden jeweils die entsprechenden Anteilswerte für die alten beziehungsweise neuen Bundesländer getrennt ermittelt und dann entsprechend der vom Statistischen Bundesamt (2008) ausgewiesenen Bevölkerungsanteile gewichtet, um einen Wert für das gesamte Bundesgebiet zu ermitteln. Da die Politbarometer-Daten bisher nur bis einschließlich 2007 vorliegen, ist eine derartige Bestimmung des Anteils Konfessionsloser für 2008 nicht möglich.

Der Zweitstimmenanteil von Bündnis 90/Die Grünen hängt weniger stark mit dem Anteil Konfessionsloser zusammen als dies für den Zusammenhang zwischen Katholikenanteil und Zweitstimmenanteil von CDU/CSU gilt ($r = 0,494$; $N = 8$).

Die theoretischen Ansätze zur politisierten Sozialstruktur legen weiterhin die Relevanz der sozioökonomischen Konfliktlinie nahe. Dementsprechend ist zu

54 Ein Vergleich mit der von Schnell und Kohler (1995) erstellten Zeitreihe hinsichtlich der Kirchgangshäufigkeit zeigt einen starken linearen Zusammenhang: $r = 0,935$ ($N = 9$).

erwarten, dass die Zweitstimmenanteile der SPD mit dem Rückgang des Arbeiteranteils in der Bevölkerung ebenfalls sinken. Analog dazu sollte ein gleichförmiger Zusammenhang zwischen dem Organisationsgrad der Gewerkschaften und den Zweitstimmenanteilen der SPD vorliegen. Da sich mit Bildung der Linkspartei das Lager der linken – und damit der traditionellen gewerkschafts- und arbeiternahen – Parteien gespalten hat, wird vermutet, dass entsprechende Zusammenhänge auch für die PDS/Die Linke festzustellen sind. Allerdings dürfte dies in verstärktem Maße für die alten Bundesländer gelten, da die WASG wesentlich von ehemaligen SPD-Mitgliedern und Gewerkschaftsnahen getragen wird. Für die neuen Bundesländer ist die Partei insbesondere als Folgepartei der SED zu sehen. Auf der anderen Seite wird erwartet, dass der Selbstständigenanteil mit den Zweitstimmenanteilen der FDP variiert, welche als Klientelpartei der Arbeitgeber identifiziert wurde.

Die Daten hierzu wurden der Erwerbsstatistik des Statistischen Bundesamtes (2009) entnommen und in Beziehung zu den bereits zuvor verwendeten Angaben zur Gesamtbevölkerung gesetzt. Die Erfassung des Arbeiteranteils ist keineswegs unproblematisch. Müller (1998) hat gezeigt, dass mit einer feingliedrigeren Messung der sozialen Lage etwa entlang des Schemas von Erikson und Goldthorpe (1992) die sozioökonomische Prägung des Wahlverhaltens besser erfasst werden kann. In diesem Fall sind die von Schnell und Kohler (1995) festgestellten Abschwächungstendenzen nicht mehr so deutlich zu finden. Das Statistische Bundesamt weist in seiner Erwerbsstatistik nach wie vor eine grobe Untergliederung aus, welche hier zugrunde gelegt wird. Vermutet werden muss daher, dass bei Stützung auf eine angemessenere Erfassung der sozialen Lage die hier berichteten Effekte stärker ausfallen. Neben der Zugehörigkeit zur Arbeiterschicht wird die Gewerkschaftsmitgliedschaft berücksichtigt. Der Deutsche Gewerkschaftsbund (2009) veröffentlicht regelmäßig die aktuellen Mitgliederzahlen der ihm angehörigen Gewerkschaften. Diese Angaben werden in Beziehung zur Gesamtbevölkerung gesetzt und als Indikator für den gewerkschaftlichen Organisationsgrad verwendet.

Die empirischen Befunde ergeben, dass die sozioökonomische Konfliktlinie nur eine geringe Prägestärke auf das Wahlergebnis hat. Der Zusammenhang zwischen dem Arbeiteranteil in der Bevölkerung und den Zweitstimmenanteilen der SPD ist moderat ($r = 0,343$; $N = 15$); die Korrelation mit dem gewerkschaftlichen Organisationsgrad ist geringfügig stärker ausgeprägt ($r = 0,454$; $N = 16$).

Der Zusammenhang der Zweitstimmenanteile der FDP mit dem Selbstständigenanteil ist mit einem Bravais-Pearson-Korrelationskoeffizient von $r = 0,197$ nur schwach ($N = 15$).

Dies entspricht den vorliegenden empirischen Befunden, wonach in Deutschland das sozioökonomische *cleavage* weniger Relevanz besitzt als das religiös-konfessionelle *cleavage*, und auch den theoretischen Ausführungen Lipset und Rokkans (1967), wonach Wertekonflikte eine größere Persistenz aufweisen als Verteilungskonflikte.

4.1.3 *Wirtschaftslage und Regierungstätigkeit*

4.1.3.1 Retrospektives Wählen

Überlegungen zum Einfluss der Regierungstätigkeit auf zukünftige Wahlchancen können auf das räumliche Modell des Parteienwettbewerbs von Downs (1968) zurückgeführt werden. Ausgangspunkt des Ansatzes ist die Annahme, dass Wähler wie Parteien als rationale Akteure aufzufassen sind. Ziel der Wähler ist es, ihren Nutzen aus der Regierungsarbeit zu maximieren; entsprechend werden sie für Parteien stimmen, welche dies am ehesten ermöglichen. Gleichfalls versuchen Parteien die Anzahl der Stimmen bei einer Wahl zu maximieren, um die Regierung stellen zu können. Ziel räumlicher Modelle ist es nun, die Gleichgewichtsbedingungen zwischen Angebot und Nachfrage in Abhängigkeit institutioneller Gegebenheiten zu bestimmen.⁵⁵

Nach dem Grundmodell von Downs (1968: 37) stimmt ein Wähler „für die Partei, von der er glaubt, daß sie ihm während der kommenden Wahlperiode ein höheres Nutzeneinkommen liefern wird als irgendeine andere Partei“. Entscheidend hierbei ist, dass sich der Vergleich der Nutzenströme auf die zukünftig erwartete Arbeit der Partei, sollte sie an die Regierung kommen, bezieht. Diese

55 Die Theoriebildung in Bezug auf räumliche Modelle zur Erklärung des Wählerverhaltens ist weit fortgeschritten. Auf diese ausdifferenzierten Ansätze kann hier nicht eingegangen werden, die Bücher von Enelow und Hinich (1984) sowie Hinich und Munger (1997) können zur Einführung herangezogen werden. Die Anwendung des räumlichen Modells auf Mehrparteiensysteme wird von Shepsle (1991) behandelt.

Annahme erschien bereits Downs (ebd.) wenig plausibel und er nahm folglich an, dass der Wähler die Leistung der amtierenden Regierung bewertet, in Beziehung zur hypothetischen Leistung der Oppositionsparteien setzt und in die Zukunft fortschreibt (Trendfaktor). Fiorina (1981: 155f.) hat in seiner Erweiterung dieses Ansatzes explizit die retrospektive Leistungsbewertung als zentrale Erklärungsvariable aufgenommen.

Bei der empirischen Umsetzung dieser Folgerungen hinsichtlich des Einflusses der Leistungsbewertung der Parteien wird in der Regel auf Umfragedaten zurückgegriffen und die Zufriedenheit der Befragten mit den Parteien in Bezug auf den Umgang mit politischen Sachfragen erhoben. Explizitere Vorhersagen über den Einfluss von Regierungshandeln auf das Wahlergebnis werden im Rahmen von Ansätzen des *economic voting* thematisiert. Im Folgenden wird zunächst die Auswirkung der Beteiligung von Parteien an der Regierung auf zukünftige Wahlchancen angesprochen. In Abschnitt 4.1.5 wird die Relevanz kurzfristiger, tief greifender politischer Ereignisse auf die Leistungsbewertung von Regierungs- beziehungsweise Oppositionsparteien und damit verknüpft deren Wahlergebnis aufgegriffen.

Im Rahmen der *economic-voting*-Studien wird der Einfluss von Indikatoren der Wirtschaftslage im Rahmen von Zeitreihenanalysen auf die Regierungspopularität sowie das Wahlverhalten thematisiert (vgl. Abschnitt 2.1.1). Kennzeichnend für diese Ansätze ist – entgegen der Modelle von Downs (1968) oder Fiorina (1981) –, dass die Beziehung von Makrogrößen in den Blick genommen wird. Ausgangspunkt sind die Überlegungen und Befunde Keys (1960), wonach das Wahlverhalten der Wähler als rationale Antwort auf die Leistung der amtierenden Regierung zu sehen ist. Kramer (1971) zeigt in einer empirischen Studie, dass Wahlergebnisse in den USA erheblich mit objektiven Veränderungen insbesondere der Wirtschaftslage während einer Regierungsamtszeit variieren. Er schließt aus diesen Befunden, dass Keys (1960) Auffassung bestätigt werden kann und das Wahlverhalten unmittelbar auf die Performanz der Regierung in Bezug auf die Wirtschaftspolitik reagiert.

In der Folge wurde versucht, die Zusammenhänge zwischen der Wirtschaftslage und dem Wahlverhalten oder auch der Regierungspopularität (grundlegend hierfür: Mueller 1970) in Anlehnung an makroökonomische Theorien als politische Konjunkturzyklen zu modellieren (Paldam 1997). Aus der Fülle verschiedener theoretischer Vorschläge, welche zumeist auf zeitreihenanalytischen

Modellen beruhen und den empirischen Studien hierzu, lassen sich einige robuste Zusammenhänge ableiten (Nannestad/Paldam 1994).

Als Kernbefund des Forschungsprogramms kann festgehalten werden, dass die Wähler die Regierung für die Wirtschaftslage verantwortlich machen. Regierungsparteien profitieren folglich von einer positiven Wirtschaftsentwicklung, werden bei negativer Wirtschaftslage jedoch bestraft. Für die Opposition gelten die Zusammenhänge analog. Dieser Befund ist jedoch empirisch nur für Zweiparteiensysteme robust. In Mehrparteiensystemen tritt das Problem auf, dass die Zurechenbarkeit von Verantwortung auf die Koalitionsparteien in der Regel nicht eindeutig ist (Paldam 1991). Die für Deutschland vorliegenden Studien zeigen jedoch übereinstimmend, dass die Verantwortungszuweisung überwiegend den größten Parteien zuteil wird. Die größten Effekte finden sich entsprechend jeweils für die CDU/CSU beziehungsweise die SPD, während Effekte der Wirtschaftslage auf die kleinen Parteien weniger stark ausgeprägt sind (vgl. für einen Überblick Bytzek 2007: 39f.). Insofern kann zumindest für Deutschland festgehalten werden, dass eine Anwendung der Ansätze zu politischen Konjunkturzyklen möglich ist.

Als weiterer zentraler Befund der *economic-voting*-Studien kann festgehalten werden, dass die Wirtschaftslage retrospektiv bewertet wird – Effekte der vermuteten wirtschaftlichen Entwicklung sind deutlich geringer. Wähler sind zudem myopisch, das heißt ihr Zeithorizont ist kurz. Den empirischen Befunden zufolge wird die Wirtschaftslage bis ungefähr ein Jahr in die Vergangenheit berücksichtigt (Paldam 1997).

Empirisch bestätigt, wenn auch weniger eindeutig als die bisher genannten Befunde, ist, dass negative wirtschaftliche Entwicklungen sich deutlich stärker auf das Wahlverhalten auswirken als positive Entwicklungen und die Effekte auf die Regierung stärker sind als auf die Opposition.

Schließlich lassen sich empirisch Abnutzungseffekte des Regierens feststellen (Nannestad/Paldam 1994). Es zeigt sich sehr robust, dass Regierungsparteien mit zunehmender Amtszeit Stimmen verlieren – dies gilt im Übrigen für Zweiparteiensysteme ebenso wie für Mehrparteiensysteme. Uneinigkeit besteht jedoch hinsichtlich der theoretischen Erklärung dieses Befundes (Paldam 1997). Zum einen wird argumentiert, dass Wähler schlicht nach gewisser Zeit einen politischen Wechsel herbeisehnen – so auch die Annahme Abramowitz' (1988, 1994, 2008b) in seinem Prognosemodell für die USA (vgl. Abschnitt 2.3.1). Zum

anderen wird argumentiert, dass Oppositionsparteien ausnützen könnten, dass sie immer Mehrheiten gewinnen können, da sie nicht wie die Regierung zu konsistenter Politik verdammt sind und somit mit inkonsistenten Versprechungen verschiedene Wählergruppen an sich binden können (Paldam 1997). Eine formale Modellierung auf Grundlage eines räumlichen Modells des Wählens haben Paldam und Skott (1995) sowie Stevenson (2002) vorgelegt. Hieraus ist ebenfalls zu folgern, dass Regieren jeweils Stimmen kostet und die Stimmenverluste mit der Regierungszeit steigen.

Diesen Überlegungen und empirischen Befunden folgend wird für das Prognosemodell erstens geprüft, ob die wirtschaftliche Entwicklung als Einflussfaktor auf das Wahlverhalten fruchtbar angewendet werden kann. Erwartet wird hierbei, dass die Effekte insbesondere in Bezug auf die amtierende Regierungspartei virulent sind. Zweitens wird auf die Regierungsbeteiligung geprüft und damit der *cost-of-ruling*-Hypothese Rechnung getragen.

4.1.3.2 Operationalisierung

Die Wirtschaftslage wird über die Arbeitslosenquote erfasst. Hintergrund hierzu sind vorliegende empirische Befunde, die konsistent zeigen, dass die Arbeitslosigkeit der einflussreichste Indikator der Wirtschaftslage ist. Weiterhin als relevant wurden die Entwicklung der Inflation sowie das Wirtschaftswachstum identifiziert. Zur Erfassung der Arbeitslosenquote wurde die Arbeitslosenstatistik der Bundesagentur für Arbeit (2009a, 2009b) herangezogen. Entsprechend der diskutierten Befunde, wonach Indikatoren der Wirtschaftslage am stärksten in den letzten Monaten vor der Wahl ausschlaggebend sind, wird jeweils die Arbeitslosenquote zwei Monate vor der Wahl betrachtet.

Die Ergebnisse der Korrelationen zwischen der Arbeitslosigkeitsquote und den Zweitstimmenanteilen ist Tabelle 4.2 zu entnehmen. Die Ergebnisse zeigen für die CDU/CSU einen deutlich negativen Zusammenhang, während für Bündnis 90/Die Grünen und die PDS/Die Linke positive Zusammenhänge zu beobachten sind. Für die SPD zeigt sich ein moderate negative, für die FDP eine nur leicht negative Korrelation.

Tabelle 4.2: Zusammenhang der Arbeitslosenquote mit dem Zweitstimmenanteil, 1953-2009

	Arbeitslosenquote (zwei Monate vor der Wahl)	N
CDU/CSU	-0,715	16
SPD	-0,390	16
FDP	-0,083	16
Bündnis 90/Die Grünen	0,704	9
PDS/Die Linke	0,231	6

Anmerkungen: Angegeben ist jeweils der Bravais-Pearson-Korrelationskoeffizient r . Auf Signifikanztests wird verzichtet, da es sich um die Ergebnisse von Vollerhebungen handelt.

Vermutet wurde, dass diese Zusammenhänge mit der Regierungsbeteiligung variieren. Stellt die CDU/CSU den Kanzler, so ist ein deutlich negativer Zusammenhang zu erwarten, ist sie nicht an der Regierung beteiligt, wäre kein Effekt zu erwarten. Es zeigt sich jedoch, dass die Regierungsbeteiligung keine intervenierende Variable darstellt. Die Korrelation bleibt unverändert negativ und der Bravais-Pearson-Korrelationskoeffizient r verändert sich auch kaum in Bezug auf die Höhe.

Etwas anders stellt sich dies für die SPD dar. Hier zeigt sich ein stärkerer negativer Zusammenhang, sofern die Partei an der Regierung beteiligt ist ($r = -0,660$, $N = 7$), jedoch nur ein geringer, wenn sie Oppositionspartei war ($r = -0,172$, $N = 9$). Dies könnte darauf zurückgeführt werden, dass den Parteien bestimmte Kernkompetenzen zugesprochen werden. Im Fall der SPD gilt gemeinhin die Arbeitsmarktpolitik als ihre Domäne, während der CDU/CSU insbesondere in Bezug auf Wachstum und Inflationsbekämpfung Kompetenzen zugesprochen werden. Entsprechend wird die SPD bestraft, wenn sie in ihrer Kernkompetenz und bei Regierungsbeteiligung versagt.

Für die FDP stellt sich dies im Übrigen ähnlich dar: Im Falle einer Regierungsbeteiligung wirkt sich eine hohe durchschnittliche Arbeitslosigkeit weitaus negativer auf den Zweitstimmenanteil aus als in Oppositionsjahren ($r = -0,463$; $N = 9$ bzw. $r = 0,118$; $N = 7$). Für die anderen Parteien ist die Analyse eines Interaktionseffekts durch die Regierungsbeteiligung auf Aggregatebene nicht

möglich, da Bündnis 90/Die Grünen nur zwei Legislaturperioden an einer Regierung beteiligt waren und die PDS/Die Linke noch nie. Die gefundenen positiven Zusammenhänge überraschen indes weniger, zumal die PDS/ Die Linke häufig als Protestpartei gerade in wirtschaftlich schwierigen Zeiten gilt und Bündnis 90/Die Grünen als Klientelpartei für wirtschaftlich gut abgesicherte Wähler angesehen wird.

Nimmt man nun noch die *cost-of-ruling*-Hypothese näher in den Blick, so kann man zunächst dem Vorschlag Gschwends und Norpoths (2001) folgen, hierfür die Anzahl der Amtsperioden zu zählen. Sie haben gezeigt, dass in Bezug auf den gemeinsamen Stimmenanteil der Regierungsparteien ein deutlicher Effekt zu erkennen ist – und zwar derart, dass mit zunehmender Amtszeit der gemeinsame Stimmenanteil sinkt. Es scheint sich also – zumindest in Bezug auf die gesamte Regierung – der postulierte Trend zum politischen Wechsel mit zunehmender Eintönigkeit auf der Regierungsbank zu bestätigen. Setzt man die Anzahl der Amtsperioden in Beziehung zu den Zweitstimmenanteilen, so zeigen sich für die drei Parteien mit hinreichender Regierungserfahrung die vermuteten negativen Zusammenhänge – allerdings nicht in der erwarteten Stärke. Für die CDU/CSU beträgt der Bravais-Pearson-Korrelationskoeffizient $r = -0,111$ ($N = 10$), für die SPD $r = -0,351$ ($N = 7$) und für die FDP $r = -0,461$ ($N = 9$). Analog stellt sich dies im Übrigen dar, wenn man die Anzahl an Amtsperioden durch die Anzahl an Jahren in der Regierung ersetzt.

4.1.4 Institutionelle Sklerose

4.1.4.1 Die These vom Altern von Demokratien

Einen anderen Zugang zur Ableitung möglicher – längerfristig wirkender – Abnutzungseffekte bietet die Theorie kollektiver Güter von Mancur Olson (1968). Er zeigt, dass die Bereitstellung von kollektiven Gütern durch Gruppen bei Unterstellung rationaler Akteure zumindest schwierig ist, da individuelle Anreize zum Trittbrettfahren bestehen. Olson (ebd.) zufolge üben zwei Variablen entscheidenden Einfluss auf das Bereitstellungspotenzial von Kollektivgütern von Organisationen aus: die Sichtbarkeit des individuellen Beitrags und damit ver-

bunden die soziale Interaktion zwischen den Mitgliedern, und als zweite Variable der Anteil des Individuums am Kollektivgut.

Beide Variablen hängen unmittelbar mit der Gruppengröße zusammen. Die Sichtbarkeit des individuellen Beitrags nimmt mit steigender Gruppengröße ab. Entscheidend ist hier demnach die Wirkung sozialer Kontrolle. Im Fall von großen Gruppen ist die Bereitstellung selektiver Anreize (Sanktionen oder Gratifikationen) zwingend für die Erstellung des Kollektivgutes – dieses lässt sich somit als Nebenprodukt rational handelnder Akteure begreifen (Nebenprodukttheorie der Interessengruppen). Sofern der Anteil des Individuums am Kollektivgut hoch genug ist, bestehen für den Akteur Anreize, die Bereitstellung des Kollektivguts alleine zu gewährleisten (vgl. hierzu Olson 1991: 20ff.).

Wendet man diese Überlegungen auf das Verhalten von Gruppen in Staaten an, so lässt sich argumentieren, dass die Gruppengröße entscheidend dafür ist, ob die Gruppen an der Mehrung sozialer Wohlfahrt partizipieren oder durch die Durchsetzung von Partikularinteressen eine Umverteilung der Wohlfahrt zu ihren Gunsten (*rent seeking*) anstreben:

„Die Organisation kann im Prinzip den Interessen ihrer Mitglieder entweder dadurch dienen, indem sie den Kuchen, den die Gesellschaft produziert, größer macht, so dass ihre Mitglieder selbst dann größere Stücke erhalten, wenn sie die gleichen Anteile haben wie zuvor, oder andererseits dadurch, dass sie größere Anteile oder Stücke des sozialen Kuchen für ihre Mitglieder erlangt“ (Olson 1991: 53).

Folgt man der Argumentation weiter, so ist aus der Anwendung der *Logik des kollektiven Handelns* auf das Gruppenverhalten abzuleiten:

„Die Organisationen für kollektives Handeln in Gesellschaften, die wir betrachten, sind daher eher und in überwältigender Weise auf Kämpfe um die Verteilung von Einkommen und Vermögen ausgerichtet als auf die Produktion weiterer Güter“ (ebd.: 56).

Aus diesen Überlegungen kann man die These institutioneller Sklerose mit zunehmendem Alter von Demokratien ableiten. In einem Staat, in welchem sich Interessen frei herausbilden und organisieren können, werden sich im Laufe der Zeit immer mehr Interessengruppen bilden, welche *rent seeking* betreiben. Der Staat als Adressat dieser auf Umverteilung bedachten Interessengruppen wird zunehmend aufgebläht durch die Notwendigkeit der Bearbeitung dieser Ansinnen

und findet immer weniger Möglichkeiten zur Verwirklichung seiner Aufgabe: der gesamtgesellschaftlichen Wohlstandsmehrung (ebd.: 102).⁵⁶

Überträgt man diesen Ansatz auf Parteien, dann ist auch – zumindest von den kleinen Parteien – Klientelpolitik und damit *rent-seeking*-Verhalten und ein Beitrag zur Beschleunigung institutioneller Sklerose zu erwarten. Neben der wirtschaftlichen Entwicklung als Indikator für diese institutionelle Sklerose ist hierbei auch an die Höhe der Wahlbeteiligung zu denken. Vor dem Hintergrund der Argumentation hinsichtlich des Einflusses der Wirtschaftslage auf das Wahlverhalten ist hieraus zu folgern, dass man einen Alterseffekt nicht nur auf die Wirtschaftslage, sondern auch auf das Wahlverhalten erwarten kann. Allerdings ist von einem umgekehrt u-förmigen Zusammenhang auszugehen: Erst nach einer Konstitutionsphase ist mit einem Abschwung zu rechnen. Ebenfalls zu erwarten ist, dass sich die ausbreitende institutionelle Sklerose negativ auf die politische Partizipation auswirkt.

4.1.4.2 Operationalisierung

Für die Überprüfung des postulierten umgekehrt u-förmigen Zusammenhangs zwischen dem Alter der Bundesrepublik und den Zweitstimmenanteilen der Parteien wurden für die CDU/CSU, die SPD, die FDP sowie Bündnis 90/Die Grünen OLS-Regressionen spezifiziert, deren Ergebnisse in Tabelle 4.3 dargestellt sind. Im Hinblick auf die PDS/Die Linke wurde auf den Einbezug eines quadrierten Altersterm verzichtet, da vermutet wird, dass sich die Partei noch im Konstitutionsprozess befindet und erst in Zukunft mit einem u-förmigen Verlauf der Zweitstimmenanteile gerechnet werden kann. Dieses Vorgehen erscheint insbesondere auch aufgrund der geringen Fallzahl notwendig.

Als Ergebnis lässt sich festhalten, dass mit Ausnahme der FDP der vermutete umgekehrt u-förmige Zusammenhang vorliegt. Ebenfalls ist eine profunde Erklärungskraft der Altersvariable zu konstatieren.

56 Detailliertere Diskussionen sowohl der theoretischen Grundlagen als auch der empirischen Befunde liefern Buchanan et al. (1980), Groß (2003), Keller (1990) sowie Pies und Leschke (1997).

Tabelle 4.3: Zusammenhang des Alters der Bundesrepublik mit dem Zweitstimmenanteil, 1949-2009

	Alter der BRD (in zehn Jahren)	Quadr. Alter der BRD	R ²	N
CDU/CSU	0,066	-0,013	0,670	17
SPD	0,094	-0,016	0,659	17
FDP	-0,023	0,004	0,161	17
Bündnis 90/Die Grünen	0,052	-0,003	0,663	9
PDS/Die Linke	0,044		0,780	6

Anmerkungen: Angegeben sind die unstandardisierten Koeffizienten von OLS-Regressionen. Auf Signifikanztests wird verzichtet, da es sich um die Ergebnisse von Vollerhebungen handelt.

Auch hinsichtlich der politischen Partizipation, gemessen anhand der Höhe der Wahlbeteiligung, bestätigt sich die Altersthese, wenn auch nicht im erwarteten Ausmaß ($r = -0,429$; $N = 17$). Mit zunehmendem Alter der Bundesrepublik sinkt die Wahlbeteiligung. Prüft man die Zusammenhänge zwischen der Höhe der Wahlbeteiligung und den Zweitstimmenanteilen, so ist für die großen Parteien und die FDP festzustellen, dass ihr Stimmenanteil mit zunehmender Wahlbeteiligung steigt, die Stimmenanteile von Bündnis 90/Die Grünen sowie der PDS/Die Linke sinken hingegen bei steigender Wahlbeteiligung (Tabelle 4.4).

Tabelle 4.4: Zusammenhang der Wahlbeteiligung mit dem Zweitstimmenanteil, 1953-2009

	Wahlbeteiligung	N
CDU/CSU	0,716	16
SPD	0,633	16
FDP	-0,117	16
Bündnis 90/Die Grünen	-0,673	9
PDS/Die Linke	-0,727	6

Anmerkungen: Angegeben ist jeweils der Bravais-Pearson-Korrelationskoeffizient r . Auf Signifikanztests wird verzichtet, da es sich um die Ergebnisse von Vollerhebungen handelt.

4.1.5 Wahlbeeinflussende Ereignisse

4.1.5.1 Von Rallye- und Priming-Phänomenen

Neben der Sachfragen- und Kandidatenorientierung werden politische Ereignisse als bedeutsame kurzfristig wirkende Einflussfaktoren angesehen (so genanntes Rallye-Phänomen). Zu denken ist hier beispielsweise an die Auswirkungen des Atomunfalls in Tschernobyl 1986, der Wiedervereinigung oder des Elbehochwassers unmittelbar vor der Bundestagswahl 2002. Dabei wird die Interpretation solcher exogener Ereignisse als wahlentscheidende Einflüsse nicht nur seitens der Medien, sondern auch von Politikwissenschaftlern geteilt (vgl. Bytzek 2007: 15f.).

Systematische Zugänge zur Analyse des Einflusses von Ereignissen auf das Wahlverhalten wurden erstmals im Rahmen der bereits thematisierten politischen Konjunkturzyklen (vgl. Abschnitt 4.1.3.1) durchgeführt. Im Rahmen von Zeitreihenanalysen wurden – zunächst ad hoc – als relevant erachtete Ereignisse als externe Schocks modelliert (Nannestad/Paldam 1994). Eine theoretische Einbindung von Ereignissen zur Ableitung von Hypothesen zu deren Wirkung kann etwa mit Rückgriff auf Erweiterung der räumlichen Theorie des Wählerverhaltens erfolgen. Kernargument ist hier, dass Ereignisse, „die Veränderungen der Popularität betroffener Parteien oder Politiker bewirken ... eine Wirkung in der Wählerschaft insgesamt als Valenz- und nicht als Positionsisues [erzielen]“ (Pappi et al. 2004: 54f.). Auch Fiorina (1981: 155ff.) geht von Effekten von Ereignissen aus. Er denkt hierbei insbesondere an eher politikferne Ereignisse wie Naturkatastrophen, von denen bei entschiedenem Handeln die amtierenden Regierungen profitieren (so genannte *windfall gains*).

Brody und Page (1975) argumentieren, dass die Wirkungsrichtung des Ereignisses auf die Regierungspopularität von zwei Faktoren abhängt: den Werten der Bevölkerung sowie der Rolle der Regierung in Bezug auf das Ereignis. Für die Regierung vorteilhaft sind demnach Ereignisse, welche im Einklang mit den Werten der Bevölkerung stehen und bei denen die Regierung eine führende Rolle spielt. Zu denken ist hier beispielsweise an die Rolle der CDU/CSU-Regierung unter Helmut Kohl im Rahmen des Prozesses der Wiedervereinigung. Ebenfalls belohnt wird entschiedenes Handeln in Krisenzeiten, etwa bei Naturkatastrophen wie der Elbeflut von 2002. Analog gilt, dass „schlechte“ Nachrichten wie Verluste

eigener Soldaten in Kriegen oder auch steigende Arbeitslosigkeit die Regierungspopularität belasten.

Neben der Einbettung von Ereignissen in räumliche Modelle des Wählerverhaltens wird das Rallye-Phänomen im Kontext der Medienwirkungsforschung diskutiert. Zentral hierbei ist die so genannte *Priming*-Hypothese. Diese postuliert, dass die Medien bestimmte Themen gezielt platzieren und hierdurch die Bewertung der Parteien maßgeblich beeinflusst wird (einen Überblick über diesen Forschungszweig liefert Brettschneider 2005a). Unabhängig von der theoretischen Fundierung resultiert aus diesen Überlegungen, dass ein bedeutsamer Einfluss von exogenen Ereignissen auf das Wahlergebnis zu erwarten ist – zumindest dann, wenn diese kurz vor der jeweiligen Wahl eintreten.

4.1.5.2 Operationalisierung

Die Berücksichtigung von wahlbeeinflussenden Ereignissen im Zusammenhang mit Prognosen ist problematisch, da der Einfluss von Ereignissen immer nur post hoc festgestellt werden kann. Allerdings können bei systematischem Einbezug von Ereignissen konditionale Vorhersagen erstellt werden für unterschiedliche Szenarien: Für die Situation mit einflussreichen Ereignissen und ohne. Dies hat den Vorteil, dass kurzfristig wirkende Einflüsse zumindest näherungsweise abgebildet werden können, ohne diese kennen zu müssen. Es lässt sich somit ein Intervall der zu erwartenden Stimmenanteile in Abhängigkeit eines Ereigniseintritts bei der nächsten Wahl angeben.

Hierzu ist jedoch die Identifikation relevanter Ereignisse notwendig. Dies ist keineswegs einfach, da zunächst bestimmt werden muss, was als Ereignis aufgefasst werden soll; dessen Relevanz muss bestimmt werden und schließlich ist auch der Einfluss auf das Wahlverhalten zu prüfen. Für Deutschland hat Bytzek (2007) eine entsprechende Studie zur Wirkung von Ereignissen auf die Regierungspopularität vorgelegt, deren Ergebnisse hier genutzt werden.

Zur Auswahl von Ereignissen stützt sich Bytzek (ebd.: 51ff.) auf drei Kriterien. Das Ereignis muss exogen sein, das heißt nicht durch direktes Regierungshandeln ausgelöst werden. Weiterhin muss es einen Bezug zur Politik aufweisen. Hierbei wird eine Klassifikation verschiedener Ereignisse entwickelt, welche diesbezüglich als relevant eingestuft werden und theoretische Vorhersagen über

die Richtungswirkung auf die Regierungspopularität zulassen. Schließlich wird die Bekanntheit des Ereignisses über die Medienpräsenz des Ereignisses bestimmt – nur hinreichend bekannte und medial präsente Ereignisse können auf die Erwartung der Wähler, ihre Bewertung der Parteien und damit ihr Wahlverhalten wirken. Als hinreichend bekannt wird ein Ereignis dann qualifiziert, wenn es mindestens vier Tage in den Medien vertreten ist. In diesem Kontext wird auch die Intensität der Berichterstattung über die Anzahl der erschienenen Artikel zum Ereignis erfasst. Die empirische Umsetzung dieser Kriterien erfolgte anhand einer Sichtung einschlägiger Chroniken sowie einer Inhaltsanalyse der Frankfurter Allgemeinen Zeitung sowie der Wochenzeitschrift *Der Spiegel*.

Verknüpft mit den theoretischen Überlegungen zur Ereigniswirkung lassen sich für den von Bytzek (2007) untersuchten Zeitraum (1977-2002) die in Tabelle 4.5 dargestellten potenziell einflussreichen Ereignisse und deren Wirkungsrichtung ableiten. Hierbei werden vor dem Hintergrund der nur kurzfristigen Wirkung entsprechender Ereignisse nur diejenigen aufgeführt, welche eine Medienpräsenz in den letzten drei Monaten vor einer Wahl aufweisen.

Um den Datensatz für die Bundestagswahl 2005 zu ergänzen, wurde die Jahreschronik des Deutschen Historischen Museums (2009) nach Ereignissen, welche zunächst als exogen bezeichnet werden können, durchsucht. Eine eingehende inhaltsanalytische Untersuchung war nicht möglich, jedoch wurde nach den ausgewählten Ereignissen im Zeitraum von Juni bis September 2005 eine Stichwortsuche im Online-Archiv der Frankfurter Allgemeinen Zeitung durchgeführt, um zumindest approximativ potenzielle relevante Ereignisse identifizieren zu können. Diese Analyse führte zu dem Ergebnis, dass im entsprechenden Zeitraum kein den Kriterien entsprechendes Ereignis eingetreten ist.⁵⁷ Anders stellt sich dies für die Bundestagswahl 2009 dar. Auch hier wurde für den Zeitraum Juni bis September 2009 eine entsprechende Recherche erstellt und hier zeigte sich, dass die Finanzkrise als ein potenziell einflussreiches Ereignis gewertet werden muss.

57 Unmittelbar in den Sinn kommen mögen einem die Terroranschläge in London (Juli 2005) oder der Hurrikan Katrina Ende August 2005. Beide Ereignisse erfüllen insofern die Kriterien Bytzeks (2007: 52f.) nicht, da internationale Krisen, Naturkatastrophen oder Terroranschläge nur berücksichtigt werden, sofern entweder Deutschland daran unmittelbar beteiligt ist (wie etwa in Afghanistan) oder ein aktives Krisenmanagement durch deutsche Politiker möglich ist (wie etwa bei der Elbeflut 2002).

Für die folgende Analyse wurde ein gerichteter Dummy dem Datensatz zugespielt, der die in Tabelle 4.5 aufgelisteten Effekte widerspiegelt. Die bivariaten Korrelationen weisen auf eher moderate Zusammenhänge hin, welche wiederum parteispezifisch variieren. Erwartet werden könnte, dass die Korrelationen bei den beiden großen Parteien stärker ausgeprägt sind. Dies ist jedoch nicht der Fall. Für die CDU/CSU beträgt $r = 0,366$ und für die SPD $r = 0,587$ bei jeweils neun Beobachtungspunkten. Der stärkste Zusammenhang besteht bei Bündnis 90/Die Grünen mit $r = 0,839$ ($N = 8$) sowie für die PDS/Die Linke ($r = 0,823$; $N = 5$). Für die FDP ist ein moderater Zusammenhang festzustellen ($r = 0,463$; $N = 9$).

Tabelle 4.5: Ereignisse und ihre vermutete Wirkung auf die Zweitstimmenanteile der Parteien, 1980-2009

Ereignis	Bundestagswahl	Effekt auf CDU/CSU	Effekt auf FDP	Effekt auf SPD	Effekt auf Bündnis 90/Die Grünen	Effekt auf PDS/Die Linke
Anschlag Oktoberfest	05.10.1980	++	+	-		
Sandoz-Rheinverschmutzung	25.01.1987	--	-	+	++	
Attentat auf Bundesinnenminister Schäuble	02.12.1990	++	0	0	0	
Wiedervereinigung	02.12.1990	++	++	--	--	
Plutoniumschmuggel	16.10.1994	++	+	0	0	0
Antisemitische Äußerungen Mölleremanns	22.09.2002	0	--	+	+	0
Elbeflut	22.09.2002	-	-	++	+	-
Finanzkrise	27.09.2009	--	+	--	+	+

Anmerkungen: Die Auswahl der Ereignisse beruht auf Bytzek (2007: 170ff.), ebenso die Ableitung der Wirkungsrichtung auf die jeweilige Kanzlerpartei; die Wirkung auf die anderen Parteien geht auf theoretische Überlegungen in Abschnitt 4.1.5.1 zurück; --: negativer Effekt; -: schwach negativer Effekt; 0: kein Effekt; +: schwach positiver Effekt, ++: positiver Effekt.

4.1.6 *Simultanes Wirken vielfältiger Einflüsse*

4.1.6.1 Mostellers Gesetz

Da bisher kein empirisch hinreichend geprüfter und bestätigter einheitlicher Erklärungsansatz des Wählerverhaltens vorliegt, ist zu vermuten, dass die bisher abgeleiteten Einflussfaktoren nur eine unzureichende Prognose von Stimmenanteilen der einzelnen Parteien ermöglichen. Anzunehmen ist, dass zahlreiche weitere, teilweise auch im Vorfeld empirisch nicht zu erfassende Einflüsse relevant sind. Zusammen genommen können diese Einflüsse als zufällig verteilt über die einzelnen Parteien angesehen werden. Trifft dies zu, so kann auf eine statistische Gesetzmäßigkeit rekurriert werden, die angibt, welchen Stimmenanteil die jeweiligen Parteien – bei einer rein zufälligen Aufteilung der Stimmen – erhalten sollten.

Die Frage ist folglich, welche Stimmenanteile sich bei rein zufälliger Aufteilung – bei gegebener Anzahl zur Wahl stehender Parteien – ergeben würden. Die zugrunde liegende Modellierung der Lösung dieses Problems geht auf Mosteller (1965: 63ff.) zurück. Illustriert wird dies anhand eines Beispiels eines zu brechenden Stabes mit Einheitslänge. Im einfachsten Fall wird der Stab einmalig gebrochen. Nimmt man an, dass jeder Punkt als Bruchstelle des Stabes mit gleicher Wahrscheinlichkeit auftreten kann, dann muss die Bruchstelle mit gleicher Wahrscheinlichkeit links oder rechts der Mitte des Stabes liegen. Aus dieser Überlegung folgt, dass der Erwartungswert der Länge des kleineren Bruchstücks $0,5 * 0,5 = 0,25$ beträgt. Das größere Bruchstück hat folglich eine erwartete Länge von $0,75$. Überträgt man diese Argumentation auf den Fall mit zwei Bruchstellen, ergibt sich folgende Aufteilung der Stablängen (vgl. für eine ausführliche Herleitung Gilman 1992; Mosteller 1965: 63ff.):

$$\text{kleinstes Segment : } \frac{1}{3} \left(\frac{1}{3} \right) = \frac{1}{9} = \frac{2}{18}$$

$$\text{mittleres Segment : } \frac{1}{3} \left(\frac{1}{3} + \frac{1}{3-1} \right) = \frac{1}{3} \left(\frac{5}{6} \right) = \frac{5}{18}$$

$$\text{größtes Segment : } \frac{1}{3} \left(\frac{1}{3} + \frac{1}{3-1} + \frac{1}{3-2} \right) = \frac{1}{3} \left(\frac{11}{6} \right) = \frac{11}{18}$$

Allgemein lässt sich die hypothetische Länge x_i^* des i -ten Bruchstücks so ausdrücken:

$$x_i^* = \frac{1}{n} \sum_{j=i}^n \left(\frac{1}{j} \right)$$

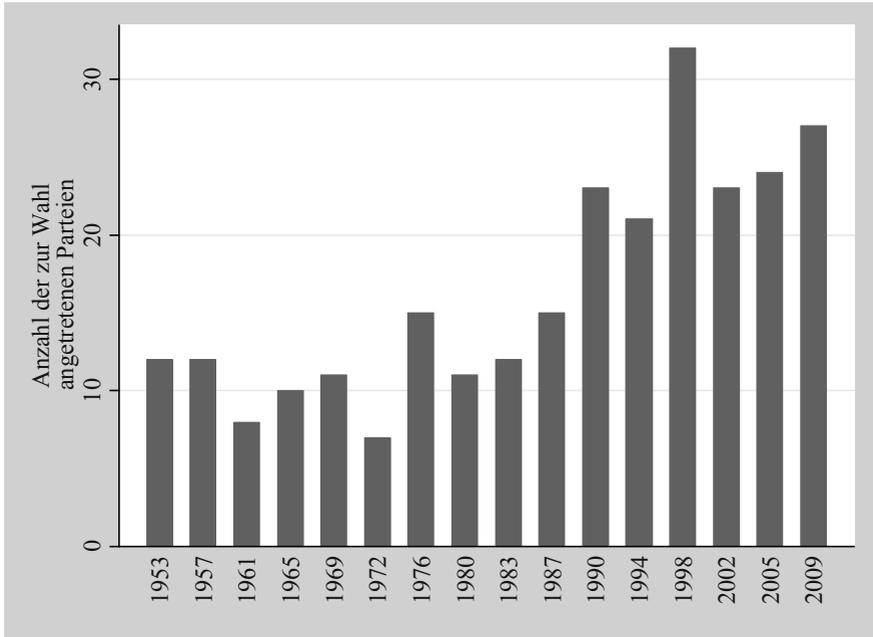
Für alle $i \neq j$ gilt hierbei $x_i^* \neq x_j^*$ sowie $\sum_{i=1}^n x_i = 1$.

Dieses Modell kann nun auf andere Bereiche übertragen werden, etwa zur Bestimmung der bei vollständigem Wettbewerb zu erwartenden Marktanteilen von Unternehmen in spezifischen Segmenten (hierzu Neumann 2000: 23ff.). Die Überlegung ist, dass bei Abwesenheit von Marktbeschränkungen die Zahl der Anbieter die Struktur der Marktanteile bestimmt. Überträgt man dies auf die vorliegende Problemstellung, könnte man bei gegebener Anzahl an Parteien und einer empirisch beobachteten Stimmenanteilsverteilung argumentieren, dass aufgrund mannigfaltiger Wahlmotive diese im Ergebnis als Zufallsprozess aufgefasst werden können und die sich bei einer Wahl ergebenden Stimmenanteile gemäß „Mostellers Zufallsgesetz“ (Neumann 2000: 23) auf die angetretenen Parteien verteilen (für eine ähnliche Argumentation zur Bestimmung erwarteter Marktanteile vgl. Braun/Keuschnigg 2008).

4.1.6.2 Operationalisierung

Im Hinblick auf die Bestimmung der Verteilung der zufällig zu erwartenden Stimmenanteile muss die Anzahl antretender Parteien bekannt sein. Diese ist – anders als beispielsweise bei der Anwendung dieser statistischen Gesetzmäßigkeit auf Märkte – im Fall von Bundestagswahlen leicht zu ermitteln, da sich bis zum 90. Tag vor der Wahl die antretenden Parteien registrieren müssen und durch den Bundeswahlausschuss deren Antritt genehmigt werden muss (Der Bundeswahlleiter 2009b). Die Anzahl der antretenden Parteien für die Bundestagswahlen 1953 bis 2009 ist Abbildung 4.4 zu entnehmen.

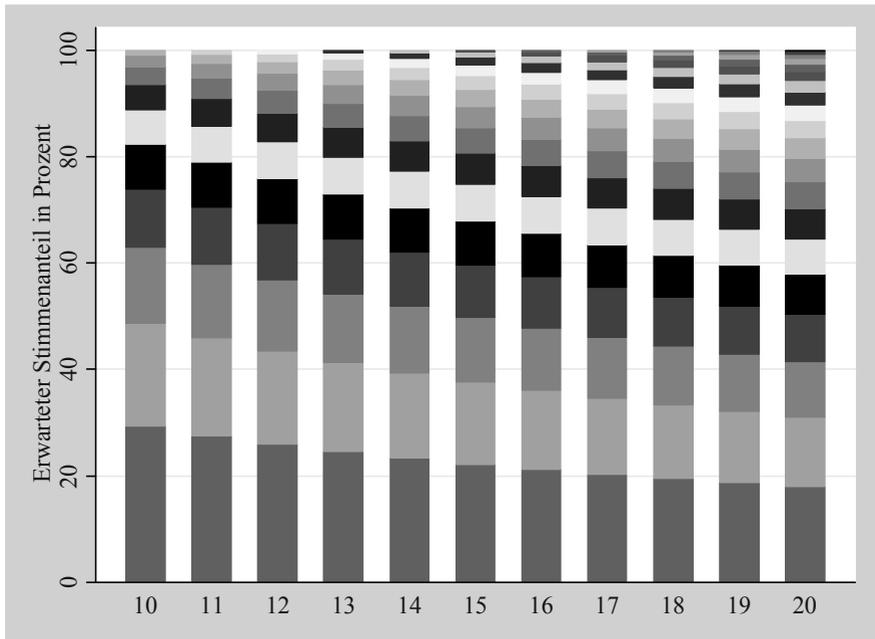
Abbildung 4.4: Anzahl angetretener Parteien bei Bundestagswahlen, 1953-2009



Anmerkungen: Dargestellt ist die zur jeweils angegebenen Bundestagswahl zugelassene Anzahl an Parteien. Die Daten wurden den Datenhandbüchern zur Geschichte des Deutschen Bundestages entnommen (Feldkamp 2005; Schindler 1999) sowie Angaben des Bundeswahlleiters (Gisart 2009).

Für jede Bundestagswahl ist damit die Berechnung der statistisch erwarteten Verteilung der Stimmenanteile über die Parteien möglich. Exemplarisch ist das Ergebnis für eine Anzahl von zehn bis 20 Parteien in Abbildung 4.5 abgebildet. Hieraus ist ersichtlich, dass statistisch keinesfalls eine Gleichverteilung der Stimmenanteile zu erwarten ist, sondern vielmehr selbst bei vielen angetretenden Parteien die fünf größten Parteien einen Großteil der Stimmen unter sich aufteilen.

Abbildung 4.5: Erwartete Verteilung der Stimmenanteile nach Anzahl angetretener Parteien



Anmerkungen: Dargestellt ist die nach Mostellers Zufallsgesetz erwartete Verteilung von Stimmenanteilen in Abhängigkeit der Anzahl zur Wahl stehender Parteien (x-Achse).

Gleichwohl ist zu konstatieren, dass die beobachteten Stimmenanteile nicht zufallsverteilt sind. Berechnet man für jede Wahl den mittleren absoluten Fehler für die nach Mosteller (1965) erwarteten Stimmenanteile, so ergibt sich über alle Wahlen von 1953 bis 2009 ein Wert von 5,9 Prozentpunkten. Aufgrund der Tatsache, dass der MAE mit der Anzahl der berücksichtigten Parteien steigt, könnte man vermuten, dass diese Abweichung nicht besonders groß ausfällt, zumal sich dieser Wert auf bis zu 32 Einzelabweichungen (Bundestagswahl 1998) bezieht, während sich die für die Sonntagsfragen berechneten Abweichungen in der Regel auf fünf Parteien beziehen.

Betrachtet man nur die bei der jeweiligen Wahl in den Bundestag eingezogenen Parteien und berechnet für diese die mittleren absoluten Abweichungen der tatsächlichen Stimmenanteile von den nach Mosteller erwarteten Werten, so

ergibt sich ein MAE von 12,0 Prozentpunkten. Dies ist darauf zurückzuführen, dass die beiden Volksparteien CDU/CSU sowie SPD deutlich größere Stimmenanteile für sich verbuchen können als es für die beiden größten Parteien nach Mostellers Zufallsgesetz zu erwarten ist. Dies verdeutlicht die Notwendigkeit der Berücksichtigung theoretischer Überlegungen zum Wählerverhalten und zeigt zugleich, dass der Parteienwettbewerb in der Bundesrepublik Deutschland auf der Aggregatebene nicht als reiner Wettbewerbsmarkt beschrieben werden kann.

Für die Nutzung von Mostellers Zufallsgesetz im Rahmen eines Prognosemodells ist es notwendig, ein Parteienranking festzulegen. Hierzu wird sich einer einfachen Heuristik bedient: Auf Grundlage des durchschnittlichen Zweitstimmenanteils der jeweiligen Partei bei den letzten drei Bundestagswahlen wird das Größenranking erstellt. Damit orientiert sich das Parteienranking an der langfristigen Parteibindung, wie sie in Abschnitt 4.1.1 abgeleitet wurde.

Es wurde argumentiert, dass mittels Mostellers Zufallsgesetz derjenige Anteil erklärt werden soll, welcher sich nicht auf die theoretisch abgeleiteten Einflussfaktoren zurückführen lässt. Deshalb wird die Differenz des Stimmenanteils bei der letzten Bundestagswahl vom erwarteten Zufallsanteil für das Prognosemodell genutzt. Empirisch zeigen sich in der bivariaten Betrachtung eher moderate Zusammenhänge zwischen den erwarteten Zufallsanteilen und den tatsächlichen Stimmenanteilen (Tabelle 4.6).

Tabelle 4.6: Zusammenhang zwischen Mostellers Zufallsanteilen und den Zweitstimmenanteilen, 1953-2009

	Erwarteter Zweitstimmenanteil nach Mosteller	Differenz des Stimmenanteils bei der letzten Bundestagswahl und des erwarteten Stimmenanteils	N
CDU/CSU	0,692	-0,520	16
SPD	0,380	0,405	16
FDP	0,015	0,178	16
Bündnis 90/Die Grünen	-0,536	0,513	8
PDS/Die Linke	-0,324	0,834	5

Anmerkungen: Angegeben ist jeweils der Bravais-Pearson-Korrelationskoeffizient r . Auf Signifikanztests wird verzichtet, da es sich um die Ergebnisse von Vollerhebungen handelt.

4.1.7 Die Relevanz des Rangplatzes

4.1.7.1 Zipfs Gesetz und Kandidatenorientierung

Die zweite Überlegung zur Integration statistischer Gesetzmäßigkeiten in den Modellvorschlag beruht auf der von Zipf (1965: 56ff.) beschriebenen empirischen Tatsache aus der Linguistik, wonach die Zahl des Auftretens von Wörtern eine Funktion ihres Rangplatzes der entsprechenden Häufigkeitsverteilung ist.⁵⁸

$$n(r) \sim \frac{1}{r^\alpha}$$

Dabei steht n für die Auftrittshäufigkeit und r für den Rangplatz. Zipf hat empirisch gezeigt, dass der Exponent α nahe 1 ist, weshalb Zipfs Gesetz häufig auf eine umgekehrt proportionale Beziehung der Auftrittshäufigkeit eines Elements einer geordneten Menge und dem Kehrwert des Rangplatzes des Elements verkürzt wird (Rapoport 1982). In der dargestellten allgemeineren Form entspricht die Zipfsche Verteilung der Umkehrfunktion der Pareto-Verteilung (Hill 1974):

$$n(r) \sim r^{-a} \Leftrightarrow r(n) \sim n^{\frac{-1}{a}}$$

Neben der Anwendung im Rahmen der Linguistik konnte die Relevanz der Gesetzmäßigkeit in vielen anderen Bereichen gezeigt werden; so etwa im Bereich der Musik im Hinblick auf die Häufigkeit des Auftretens bestimmter Noten (Zanette 2008), in der Malerei bezüglich der Bestimmung der Anteilsverteilung von Farbflächen (Volosin/Orlov 1982), die Häufigkeitsverteilung des Aufrufens von Internetseiten (Adamic/Huberman 2002), dem Kommunikationsverhalten zwischen Tieren (McCowan et al. 2005), in Bezug auf die Größenverteilung von Städten (Gabaix 1999a, 1999b) oder auch der Verteilung von Macht und sozialem Status (Rapoport 1982).

58 Für eine mögliche Erklärung der Gültigkeit von Zipfs Gesetz im Rahmen der Linguistik vgl. Mandelbrot (1953) und Simon (1955). Zur Diskussion des letzt genannten Vorschlags siehe Mandelbrot (1959, 1961) sowie Simon (1960, 1961).

In diesem Rahmen soll Zipfs Gesetzmäßigkeit für die Vorhersage von Stimmenanteilen von Parteien genutzt werden. Eine in den bisherigen theoretischen Überlegungen vernachlässigte Komponente ist die Rolle der Kandidaten. Ein zentrales Ergebnis der empirischen Wahlforschung ist deren Relevanz bei der Wählerentscheidung. So postuliert der sozialpsychologische Ansatz (Campbell et al. 1964: 15ff.), dass die Beurteilung der Kandidaten entlang verschiedener Dimensionen (persönliche Eigenschaften sowie politische Fähigkeiten) als Kurzfristfaktor wesentlich ist neben der Parteiidentifikation und der Sachfragenorientierung.⁵⁹

Um neben politischen Ereignissen weitere kurzfristige Einflüsse einbeziehen zu können, wird auf die Kandidatenorientierung zurückgegriffen. Vermutet wird, dass die Rangfolge der Kandidaten im Urteil der Wähler auf Aggregatebene mittels Zipfs Gesetz den Stimmenanteil der Parteien vorherzusagen ermöglicht. Folglich wird hier von einem Zusammenhang des Zweitstimmenanteils mit dem Kehrwert des Rangplatzes des jeweiligen Spitzenkandidaten ausgegangen.

4.1.7.2 Operationalisierung

Für die Ermittlung des Rangplatzes der Kandidaten muss auf Umfragedaten zurückgegriffen werden. Üblicherweise werden in Vorwahlumfragen Skalometerfragen verwendet, um die Kandidaten zu bewerten. Aus den entsprechenden Datensätzen wurde für jede Partei und Wahl die durchschnittliche Skalometerbewertung herangezogen, um eine Rangliste der Kandidaten zu erstellen.

Hinsichtlich der hierzu verwendeten Daten sind einige Anmerkungen notwendig. Grundsätzlich wurde versucht, zu jeder Wahl Vorwahlumfragen, welche etwa zwei Monate vor dem Wahltermin durchgeführt wurden, heranzuziehen. Dieser Zeitpunkt wurde gewählt, um einen hinreichend großen zeitlichen Abstand zur Wahl zu gewährleisten, um frühzeitige Vorhersagen zu ermöglichen. Dieses Vorgehen wird im Übrigen analog von Gschwend und Norpoth (2001)

59 Die Diskussion über die Relevanz der Kandidatenorientierung wird verstärkt in jüngerer Zeit geführt. Hierbei steht insbesondere die These der Personalisierung der Politik im Mittelpunkt. Dieser zufolge nimmt die Bedeutung der Kandidatenorientierung für die Wählerentscheidung im Zeitablauf zu. Bisherige empirische Ergebnisse sind widersprüchlich, jedoch wird eher davon ausgegangen, dass die Personalisierungsthese nicht zutrifft (Brettschneider 2002; Gabriel/Keil 2008). Vgl. zu dieser Diskussion auch Abschnitt 3.1.

vorgeschlagen, welche die Kanzlerpräferenz in ihr Modell integrieren. Herangezogen werden weiterhin die Bewertungen der jeweiligen Spitzenkandidaten. Die Auswahl dieser ist für die kleineren Parteien nicht unproblematisch, da sowohl die FDP als auch Bündnis 90/Die Grünen erst in jüngerer Zeit offiziell Spitzenkandidaten für Bundestagswahlen küren. Diesem Problem wird insofern pragmatisch begegnet, da die meisten vorliegenden Umfragen für diese Parteien für die betreffenden Wahlen sowieso nur einen Kandidaten durch die Befragten bewerten lassen – diese Bewertung wird dann herangezogen.

Ebenfalls problematisch erscheint die variierende Erfassung der Kandidatenorientierung im Zeitablauf. Für die Bundestagswahlen seit 1980 stellt dies kein Problem dar, da hier auf die Daten des Politbarometers zurückgegriffen werden kann (vgl. zu diesen Daten Abschnitt 3.3.1). Im Rahmen dieser Trenderhebung werden stets dieselben Skalometerfragen eingesetzt. Für frühere Bundestagswahlen stellt sich dies anders dar, hier wurden zum Teil variierende Skalen verwendet oder andere Instrumente zur Kandidatenbewertung eingesetzt, etwa die Bewertung über Likert-Skalen. Unabhängig vom Erhebungsinstrument wird hier jeweils die Rangfolge der Kandidaten im Meinungsbild der Befragten gebildet und dem Datensatz zugespield.⁶⁰

Empirisch zeigen sich für die CDU/CSU ($r = 0,402$; $N = 15$) und die SPD ($r = 0,603$; $N = 15$) Zusammenhänge zwischen dem Zweitstimmenanteil und dem Kehrwert des Rangplatzes der Spitzenkandidaten in fast der erwarteten Stärke. Die Zusammenhänge bei den kleineren Parteien liegen für die FDP bei $r = 0,156$ ($N = 15$), für Bündnis 90/Die Grünen bei $r = 0,631$ ($N = 5$) und für die PDS/Die Linke bei $r = -0,106$ ($N = 4$). Für Bündnis 90/Die Grünen konnten nur fünf Werte berücksichtigt werden, da im Politbarometer vom Juli 2009 keine Bewertung der Spitzenkandidaten der Partei erfolgte (Forschungsgruppe Wahlen 2009b). Ähnliches gilt für die PDS/Die Linke: Hier konnten nur vier Werte ein-

60 Für die Bundestagswahlen von 1953 bis 1976 wurde auf die jeweiligen Wahlstudien zurückgegriffen. Diese Datensätze haben auch Schnell und Kohler (1995) herangezogen, um die sozialstrukturelle Zusammensetzung der Wählerschaft im Zeitablauf zu erfassen (vgl. Abschnitt 1.2.2). Die Datensätze sind über das GESIS – Leibniz-Institut für Sozialwissenschaften zu beziehen. Im Einzelnen wurden folgende Studien verwendet: ZA-Nr. 0145, 3272, 0055, 0525, 0635 und 0823. Weder die vorgenannten Personen noch das GESIS tragen irgendeine Verantwortung für die Analyse oder Interpretation der Daten in dieser Arbeit.

bezogen werden, da die Bewertung der Spitzenkandidaten Gregor Gysi bzw. Oskar Lafontaine nur für die jüngsten Bundestagswahlen erhoben wurde.⁶¹

4.2 Empirische Befunde zur Prognosegüte des Modells

Zur Ableitung von Modellvariablen wurden aufgrund des Mangels an einem Theorieansatz, welcher hinreichend präzise alle wesentlichen Einflussfaktoren auf die Wählerentscheidung spezifiziert, verschiedene theoretische Ansätze herangezogen. Darüber hinaus wurden nur diejenigen Faktoren berücksichtigt, welche als empirisch bestätigt angesehen werden. Da vermutet wird, dass die so hergeleiteten Faktoren keine hinreichend präzisen Vorhersagen von Stimmenanteilen erlauben, wurde weiterhin auf die von Mosteller (1965) begründete Gesetzmäßigkeit rekurriert, um die Verteilung von Stimmenanteilen unter vollständigen Wettbewerbsbedingungen berechnen zu können.

Es wurden einige potenziell denkbaren Modellvariablen präsentiert, welche jedoch nicht alle für ein Prognosemodell genutzt werden können. Da bisher nur 17 Bundestagswahlen stattfanden, stellt schon die Fallzahl eine Restriktion auf wenige zentrale Faktoren dar. Insofern muss eine Auswahl aus den präsentierten Variablen getroffen werden. Dazu werden diese zunächst nach langfristig, mittelfristig und kurzfristig wirkenden Faktoren unterschieden. Diese Unterscheidung schließt an die gängige Klassifikation der Einflussfaktoren in der Wahlforschung an (etwa bei Gschwend/Norpoth 2001).

Die theoretischen Ansätze betonen in der Regel in unterschiedlicher Art und Weise das Zusammenwirken dieser Faktoren. So wird im *cleavage*-Ansatz insbesondere die Relevanz langfristig wirkender sozialstruktureller Variablen postuliert, während in räumlichen Ansätzen des Wählerverhaltens der kurzfristigen retrospektiven Leistungsbewertung eine zentrale Bedeutung zugemessen wird. Im sozialpsychologischen Ansatz wird postuliert, dass die langfristig aufgebaute Parteiidentifikation eine Filterfunktion für die kurzfristig wirkenden Faktoren der Kandidaten- sowie Sachfragenorientierung erfüllt. Allen Ansätzen gemein ist jedoch

61 Bei den Bundestagswahlen 2005 und 2009 trat die PDS/Die Linke mit zwei Spitzenkandidaten an: Gregor Gysi und Oskar Lafontaine. In die Analyse fließt die durchschnittliche Bewertung der beiden Kandidaten ein.

die grundsätzliche Einsicht der Notwendigkeit, sowohl langfristige als auch kurzfristige Faktoren für eine Erklärung des Wahlverhaltens berücksichtigen zu müssen. Deshalb werden die abgeleiteten Variablen in Tabelle 4.7 entlang dieser Klassifikation aufgelistet.

Als langfristige Faktoren werden hier diejenigen bezeichnet, welche unabhängig von Ereignissen in der laufenden Legislaturperiode sind, also die Parteibindung, die sozialstrukturelle Zusammensetzung der Bevölkerung sowie die Indikatoren des Ausmaßes der institutionellen Sklerose. Als kurzfristig wirkend werden hingegen Einflüsse angesehen, welche insbesondere die Wahlentscheidung von unentschiedenen, nicht parteigebundenen Wählern, oder auch Wählern, die *cross pressure* unterliegen, beeinflussen und auch *last minute swings* bewirken können. Im sozialpsychologischen Ansatz werden hier insbesondere die Kandidatenorientierungen benannt, dem Ansatz retrospektiven Wählens sowie der *economic-voting*-Forschung zufolge sind auch exogene Ereignisse hier zu verorten. Als mittelfristig schließlich werden die Indikatoren der Wirtschaftslage sowie der Regierungstätigkeit aufgefasst, da diese in ihrer zeitlichen Wirkung zwischen den beiden anderen Faktoren anzusiedeln sind und von einer Wirkung innerhalb der laufenden Legislaturperiode beziehungsweise bis zu einem Jahr vor der Wahl auszugehen ist.

Tabelle 4.7: Übersicht über die abgeleiteten Variablen für ein Prognosemodell

Langfristfaktoren	Mittelfristfaktoren	Kurzfristfaktoren
Parteibindung	Wirtschaftslage	Politische Ereignisse
Politisierte Sozialstruktur	Regierungstätigkeit	Kandidatenorientierung
Institutionelle Sklerose		

Quelle: Eigene Darstellung.

Diese Klassifikation wird genutzt, um eine Auswahl von Faktoren zu treffen. So soll gewährleistet werden, dass jede Gruppe im Prognosemodell berücksichtigt wird. Für die *langfristigen Faktoren* ist zudem festzuhalten, dass das Alter Deutschlands und die sozialstrukturelle Zusammensetzung miteinander korrelieren. Deshalb wird im Weiteren nur auf letztere zurückgegriffen. Diese Auswahl erfolgt aufgrund der Tatsache, dass für die Modellierung des postulierten und

festgestellten kurvilinearen Zusammenhangs zwischen dem Alter Deutschlands und den Zweitstimmenanteilen zwei Variablen in das Prognosemodell aufgenommen werden müssten.

Für die sozialstrukturelle Zusammensetzung wird der Anteil Katholiken, der Anteil katholischer Kirchgänger, der Arbeiteranteil, der Anteil Selbstständiger sowie der Anteil Konfessionsloser in der (Wahl-)Bevölkerung als relevant für das Abschneiden der einzelnen Parteien angesehen. Hierbei ist jedoch zu berücksichtigen, dass für Bündnis 90/Die Grünen sowie die PDS/Die Linke hinsichtlich der Fallzahlen und damit der möglichen einzuschließenden Parameter noch größere Restriktionen zu beachten sind. Für diese beiden Parteien wird deshalb auf die Berücksichtigung der politisierten Sozialstruktur verzichtet. Im Fall von Bündnis 90/Die Grünen liegen keine prozessproduzierten Daten für den Anteil der Konfessionslosen vor (vgl. Abschnitt 4.1.2.4) und die Allianzen von Großgruppen mit der PDS/Die Linke sind aus theoretischer Perspektive nicht eindeutig.

Weiterhin wird die langfristige Parteibindung berücksichtigt. Die Relevanz der Parteibindung für die individuelle Wahlentscheidung ist nicht nur aus theoretischen Überlegungen heraus als zentral anzusehen. Auch legen die empirischen Befunde, welche eine zentrale Bedeutung der Parteibindung für die individuelle Wahlentscheidung ausweisen, deren Verwendung nahe.

Als *mittelfristig wirkender Faktor* wird dem Vorschlag von Gschwend und Norpoth (2001) gefolgt. Sie nutzen in ihrem Modell zur Vorhersage des gemeinsamen Stimmenanteils der Koalitionspartner bei Bundestagswahlen zur Abbildung von Abnutzungseffekten der Regierungstätigkeit die Anzahl der Perioden, welche die amtierende Regierung im Amt ist. Da die PDS/Die Linke bisher an keiner Regierung beteiligt war, kann dieser Indikator hier nicht für die Vorhersage von Stimmenanteilen der Partei genutzt werden, weshalb hier als mittelfristig wirkender Parameter alternativ auf die Arbeitslosenquote zurückgegriffen wird.

Besonders schwierig ist die Erfassung *kurzfristig wirkender Faktoren* in einem Prognosemodell. Aus theoretischer Perspektive erscheinen Kandidatenorientierungen sowie politische Ereignisse besonders einflussreich. Während jedoch die Relevanz der Kandidatenorientierung für das Wählerverhalten empirisch gut getestet ist, hat bisher nur Bytzek (2007) eine systematische Analyse des Einflusses politischer Ereignisse vorgelegt. Hierbei wird zudem der Einfluss der Ereignisse auf die Regierungspopularität untersucht und nicht auf das tatsächliche Wählerverhalten.

Zudem umfasst die hierzu herangezogene Variable erst Information für die Bundestagswahlen seit 1980. Auf der anderen Seite liegen für die Kandidatenorientierung zwar Daten für alle Bundestagswahlen vor, jedoch nicht für alle Parteien über den gesamten Zeitraum. Schließlich beruht die Konstruktion des aggregierten Rangplatzes der Kandidaten zumindest für Bundestagswahlen vor 1980 auf variierenden Erhebungsinstrumenten.

Für Bündnis 90/Die Grünen sowie die PDS/Die Linke können aufgrund der fehlenden Erfassung von Bewertungen ihrer Spitzenkandidaten in verschiedenen Jahren die Kandidatenorientierung nicht herangezogen werden. Dies wird jedoch als kurzfristig wirkender Faktor für die anderen Parteien genutzt. Für die PDS/Die Linke kann auch die Wirksamkeit von Ereignissen nicht getestet werden, da die Variable keine Varianz für den vorliegenden Untersuchungszeitraum aufweist. Deshalb muss für diese Partei gänzlich auf einen Kurzfristfaktor verzichtet werden. Für Bündnis 90/Die Grünen kann hingegen der gerichtete Dummy als Indikator für politisch wirksame Ereignisse genutzt werden.

Schließlich werden für alle Parteien – mit Ausnahme der PDS/Die Linke aufgrund der geringen Fallzahl – zusätzlich die Differenz des Zweitstimmenanteils der jeweiligen Partei bei der letzten Bundestagswahl und dem erwarteten Stimmenanteil nach Mosteller (1965) benutzt.

Zunächst werden die Befunde zur Güte des aufgestellten Modells präsentiert (Abschnitt 4.2.1). Im zweiten Schritt wird die Prognosegüte anhand von *out-of-sample*-Vorhersagen im Vergleich zu wahlabsichtsfragenbasierten Vorhersagen betrachtet (Abschnitt 4.2.2).

4.2.1 Retrospektive Bewertung des Modells

Aufgrund der geringen Fallzahl sowie fehlender Daten kann nur für die CDU/CSU, die SPD und die FDP ein einheitliches Modell geschätzt werden. Das lineare Regressionsmodell wird folgendermaßen spezifiziert:

$$STIM_i = \beta_0 + \beta_1(PB_i) + \beta_2(ZUF_i) + \beta_3(PS_i) + \beta_4(AMT_i) + \beta_5(KAND_i) + \varepsilon_i$$

wobei:

$STIM_i$ *Zweitstimmenanteil von Partei i*

PB_i *langfristige Parteibindung*

ZUF_i *Differenz des Zweitstimmenanteils bei der letzten Wahl und dem erwarteten Zufallsanteil*

PS_i *Indikatoren der politisierten Sozialstruktur*

AMT_i *Anzahl der Amtsperioden aktueller Regierungsbeteiligung*

$KAND_i$ *Kehrwert des Rangplatzes der aggregierten Bewertung des Spitzenkandidaten*

ε_i *Residuum*

Für Bündnis 90/Die Grünen beziehungsweise PDS/Die Linke werden folgende Modelle geschätzt:

$$STIM_{GR} = \beta_0 + \beta_1(PB_{GR}) + \beta_2(ZUF_{GR}) + \beta_3(AMT_{GR}) + \beta_4(ER_{GR}) + \varepsilon_{GR}$$

$$STIM_{PDS} = \beta_0 + \beta_1(PB_{PDS}) + \beta_2(AL) + \varepsilon_{PDS}$$

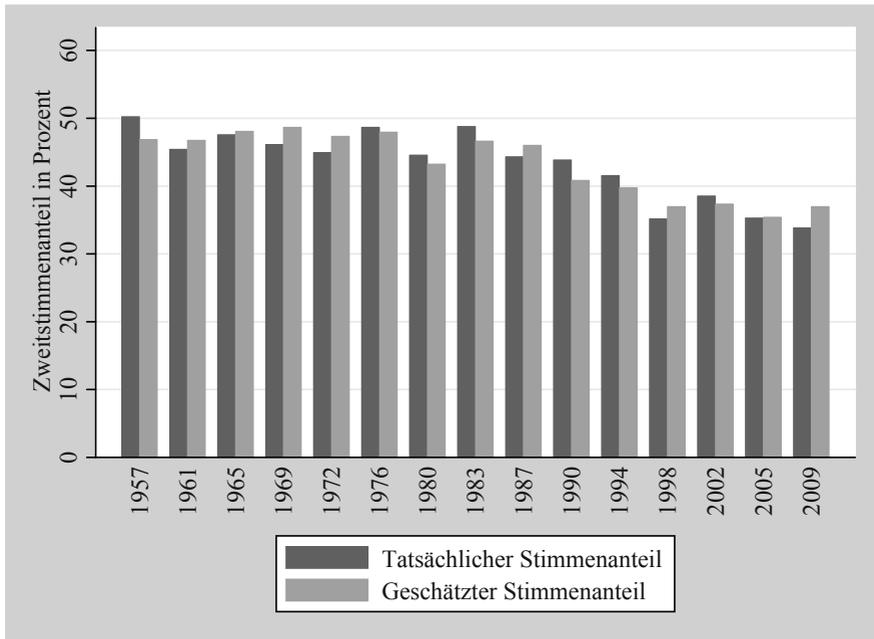
wobei:

ER_{GR} *Gerichteter Dummy für die Wirkung eines politischen Ereignisses*

AL *Arbeitslosenquote zwei Monate vor der Wahl*

Zunächst werden die Ergebnisse der geschätzten Regressionen über den gesamten Zeitraum betrachtet. Dies ermöglicht eine Abschätzung der Anpassung des Gesamtmodells an die Daten. Damit wird eine Antwort auf die Frage möglich, ob sich das Modell überhaupt als Ausgangspunkt etwaiger Prognosen eignet. Die angegebenen Modelle werden auf Grundlage aller vorhandenen Beobachtungen geschätzt und die geschätzten Zweitstimmenanteile aus den OLS-Regressionen mit den tatsächlichen Wahlergebnissen verglichen. Für die CDU/CSU zeigt sich, dass insgesamt eine hohe Anpassungsgüte erreicht wird ($R^2_{adj} = 0,755$; $N = 15$). Auch die linearen Vorhersagen weichen nur geringfügig von den tatsächlich beobachteten Zweitstimmenanteilen ab. Im Durchschnitt beträgt die absolute Abweichung bei der CDU/CSU 1,8 Prozentpunkte. Graphisch ist dies in Abbildung 4.6 veranschaulicht.

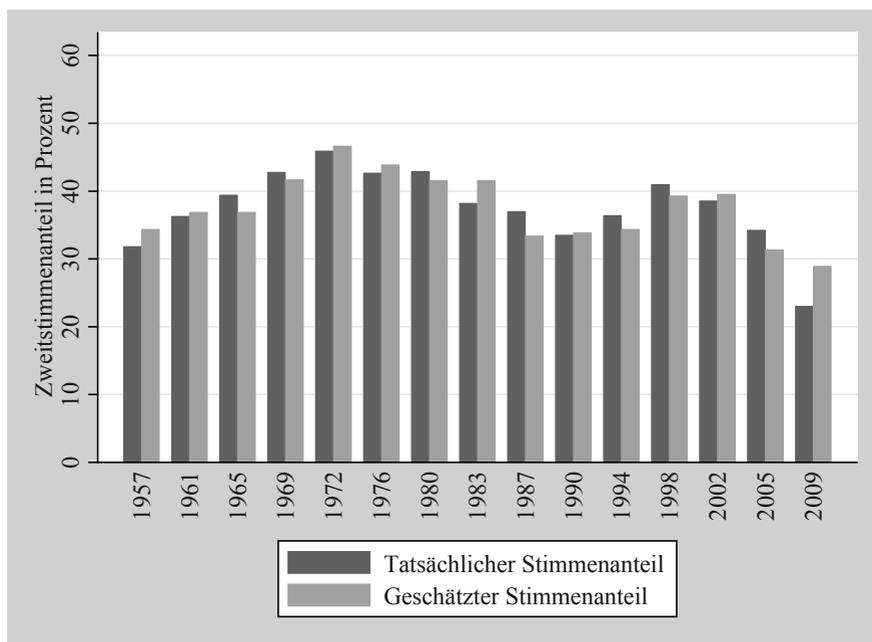
Abbildung 4.6: Tatsächliche und geschätzte Zweitstimmenanteile für die CDU/CSU, 1957-2009



Anmerkungen: Dargestellt sind die tatsächlichen Zweitstimmenanteile sowie die geschätzten Stimmenanteile für die einzelnen Bundestagswahlen auf Basis einer multiplen linearen OLS-Regression.

Auch für die SPD stellt sich dies ähnlich dar (Abbildung 4.7). Die Anpassungsgüte ist vergleichbar mit der für die CDU/CSU ($R^2_{\text{adj}} = 0,786$; $N = 15$). Allerdings ist die durchschnittliche absolute Abweichung der Vorhersagen von den beobachteten Stimmenanteilen mit 2,1 Prozentpunkten höher. Dies ist nicht zuletzt auf die deutliche Überschätzung des Stimmenanteils der SPD bei der Bundestagswahl 2009 zurückzuführen. Dies könnte verbessert werden, indem man – analog zu Norpoth und Gschwend (2009) – für diese Wahl nicht den gleitenden Durchschnitt der Stimmenanteile der letzten drei Wahlen heranzieht, sondern nur denjenigen der Bundestagswahl 2005.

Abbildung 4.7: Tatsächliche und geschätzte Zweitstimmenanteile für die SPD, 1957-2009

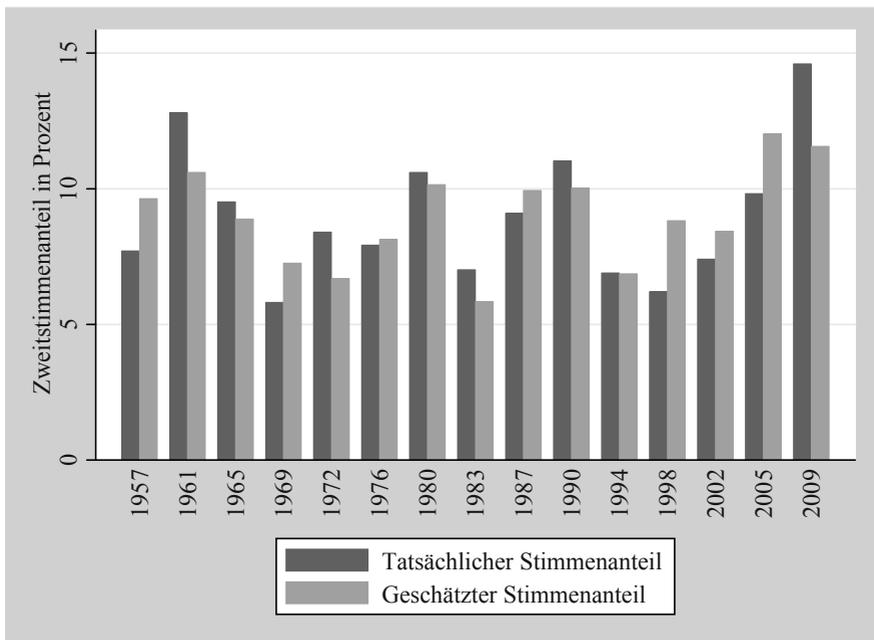


Anmerkungen: Dargestellt sind die tatsächlichen Zweitstimmenanteile sowie die geschätzten Stimmenanteile für die einzelnen Bundestagswahlen auf Basis einer multiplen linearen OLS-Regression.

Für die kleinen Parteien kann aus drei Gründen von einer insgesamt schlechteren Anpassung ausgegangen werden. Erstens befassen sich viele der vorliegenden empirischen Studien der Wahlforschung nur mit den beiden großen Parteien, weswegen die theoretischen Überlegungen allenfalls für die Wahlentscheidung zugunsten von CDU/CSU oder SPD als gut bestätigt angesehen werden können. Zweitens beziehen sich einige der theoretischen Überlegungen insbesondere auf die große Regierungspartei beziehungsweise die größte Oppositionspartei. In diesen Fällen bleibt aus theoretischer Sicht die Spezifizierung der Effekte auf kleinere Parteien zum Teil nebulös. Drittens ist die Datengrundlage für Bündnis 90/Die Grünen und die PDS/Die Linke sehr dünn, weswegen allein aufgrund der wenigen Beobachtungspunkte von großen Unsicherheiten ausgegangen werden muss.

Entsprechend dieser Erwartung weist das Modell für die FDP eine geringe Anpassungsgüte auf ($R^2_{\text{adj}} = 0,285$; $N = 14$). Die Abweichungen der geschätzten von den beobachteten Zweitstimmenanteilen fallen entsprechend deutlicher aus als bei den großen Parteien. Besonders auffällig ist hierbei die merkliche Unterschätzung des Stimmenanteils der FDP auf Grundlage des Modells bei der Bundestagswahl 2009 (Abbildung 4.8). Im Durchschnitt beträgt die absolute Abweichung 1,4 Prozentpunkte. Angesichts des durchschnittlichen Stimmenanteils von 9,0 Prozent der FDP bei den Bundestagswahlen von 1957 bis 2009 erscheint diese Abweichung recht hoch – allerdings hat sich auch bei den wahlabsichtsbasierten Vorhersagen gezeigt, dass insbesondere das Abschneiden der kleinen Parteien fehlerbehaftet ist (vgl. Abschnitt 3.4.2). Der Vergleich der Prognosegüte des Modells mit den Sonntagsfragen wird im folgenden Abschnitt vorgestellt.

Abbildung 4.8: Tatsächliche und geschätzte Zweitstimmenanteile für die FDP, 1957-2009

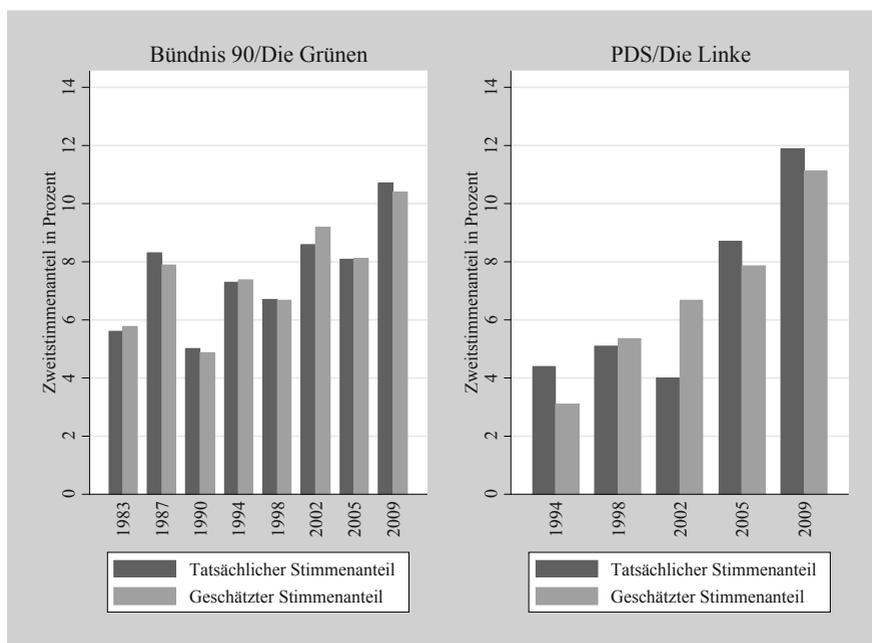


Anmerkungen: Dargestellt sind die tatsächlichen Zweitstimmenanteile sowie die geschätzten Stimmenanteile für die einzelnen Bundestagswahlen auf Basis einer multiplen linearen OLS-Regression.

Vermutet werden kann, dass die Wahlentscheidung für die FDP insbesondere von strategischen Motiven geleitet ist. So zeigt sich beispielsweise, dass der Stimmenanteil der FDP systematisch höher ausfällt als man auf Grundlage der Personen mit Parteibindung für die FDP erwarten würde. Empirisch wurde zudem gezeigt, dass die Partei überproportional stark von Leihstimmeneffekten profitiert (Jesse 1988).

Überraschenderweise ist die mittlere absolute Abweichung bei Bündnis 90/Die Grünen mit 0,2 Prozentpunkten ($R^2_{\text{adj}} = 0,930$; $N = 8$) sehr gering. Bei der PDS/ Die Linke ist die Anpassung erwartungsgemäß etwas schlechter ($R^2_{\text{adj}} = 0,556$; $N = 5$) und auch die mittlere absolute Abweichung fällt mit 1,2 Prozentpunkten höher aus. In Abbildung 4.9 sind die jeweils geschätzten und tatsächlichen Zweitstimmenanteile der beiden Parteien bei Bundestagswahlen einander gegenübergestellt.

Abbildung 4.9: Tatsächliche und geschätzte Zweitstimmenanteile für Bündnis 90/Die Grünen, 1983-2009, und PDS/Die Linke, 1994-2009



Anmerkungen: Dargestellt sind die tatsächlichen Zweitstimmenanteile sowie die geschätzten Stimmenanteile für die einzelnen Bundestagswahlen auf Basis von multiplen linearen OLS-Regressionen.

Diese Betrachtung hat gezeigt, dass mit einfachen Modellen auf der Makroebene bereits eine gute Anpassung an alle der bisher vorliegenden Daten möglich ist. Deutlich wurde hierbei jedoch, dass insbesondere die Wahlentscheidung zugunsten der FDP nicht hinreichend durch klassische Ansätze erklärt werden kann. Das entwickelte Prognosemodell stützt sich im Kern auf die Parteibindung, die Kandidatenorientierung, die Auswirkung der Regierungstätigkeit auf die Stimmenanteile sowie Mostellers Zufallsgesetz. Für die FDP ist dies nicht hinreichend. Für eine genauere FDP-Prognose erscheint vielmehr eine Integration strategischer Überlegungen des Wählers in die bisherigen Ansätze sinnvoll.

Bisher wurden überdies nur post-hoc-Schätzungen der Zweitstimmenanteile betrachtet, womit noch nichts über die Tauglichkeit des hier entwickelten Vorschlags zur Prognose von Bundestagswahlergebnissen ausgesagt ist. Die wird im Folgenden anhand von *out-of-sample*-Prognosen geprüft. Die Vorhersagegüte des Modells wird weiterhin der Güte von Wahlabsichtsumfragen gegenübergestellt.

4.2.2 Vergleich mit wahlabsichtstfragenbasierten Vorhersagen

Aufschlussreicher als die bisherige empirische Betrachtung ist die Evaluation der Güte des Prognosemodells anhand von *out-of-sample*-Prognosen. Auch der Vergleich mit umfragebasierten Vorhersagen erscheint wesentlich zur Abschätzung des empirischen Wertes des Modellvorschlags. Hierzu werden im Folgenden für jede Partei Regressionen unter Ausschluss der interessierenden und aller nachfolgenden Bundestagswahlen berechnet und auf dieser Grundlage der Zweitstimmenanteil für diese Wahl geschätzt. Einschränkend wird hier auf eine Berücksichtigung der PDS/Die Linke aufgrund der geringen Fallzahlen verzichtet. Das sparsame Regressionsmodell umfasst zwar nur zwei Variablen, jedoch lägen für eine Prognose des Stimmenanteils von 2009 nur vier Beobachtungspunkte als Berechnungsgrundlage vor.

Zur Erstellung der *out-of-sample* -Prognose für beispielsweise die Bundestagswahl 2009 werden die in Abschnitt 4.2.1 spezifizierten Modelle unter Ausschluss dieser Wahl geschätzt und die entsprechenden Werte der Modellparameter in die Regressionsgleichung eingesetzt, um so den Stimmenanteil für die Bundestagswahl 2009 zu schätzen. Analog wird für die Bundestagswahlen bis 1994 zurück vorgegangen.

Im zweiten Schritt werden diese aus den Prognosemodellen abgeleiteten Vorhersagen dem tatsächlichen Ergebnis sowie publizierten Sonntagsfragenergebnissen gegenüber gestellt. Um einen fairen Vergleich zu ermöglichen, wird auf Umfragen zurückgegriffen, welche etwa zwei Monate vor der Wahl veröffentlicht wurden. Auf diesen Zeitpunkt beziehen sich auch die in den Modellen verwendeten Informationen etwa zur Bewertung der Spitzenkandidaten.

Als Vergleichsmaßstab wird zum einen die durchschnittliche Performanz aller im betreffenden Zeitraum veröffentlichten Sonntagsfragen herangezogen sowie die jeweilige Projektion des Politbarometers. Der Durchschnittswert stellt insofern keine geeignete Vergleichsgröße dar, da hier bereits mögliche Abweichungen einzelner Institute durch die Mittelung ausgeglichen wurden. Als Maßzahlen wird auf die mittlere absolute Abweichung (MAE), die mittlere absolute prozentuale Abweichung (MAPE) sowie den Vergleichsquotienten ν zurückgegriffen (vgl. Abschnitt 3.3.2).

Zunächst werden für jede Partei out-of-sample-Prognosen für die Bundestagswahlen 1994 bis 2009 erstellt und in der Folge die Abweichungen über alle Parteien für die einzelnen Wahlen berechnet. Die entsprechenden Werte für die mittlere absolute und mittlere absolute prozentuale Abweichung sowie den Vergleichsquotienten ν werden den analog berechneten Werten der Politbarometer-Projektion zwei Monate vor der jeweiligen Wahl und dem Durchschnitt aller publizierten Sonntagsfragenergebnisse zwei bis drei Monate vor der Wahl gegenüber gestellt. Hierbei ist anzumerken, dass für die Bundestagswahlen 1994 und 1998 die PDS/Die Linke aufgrund mangelnder Beobachtungspunkte nicht berücksichtigt wurde.

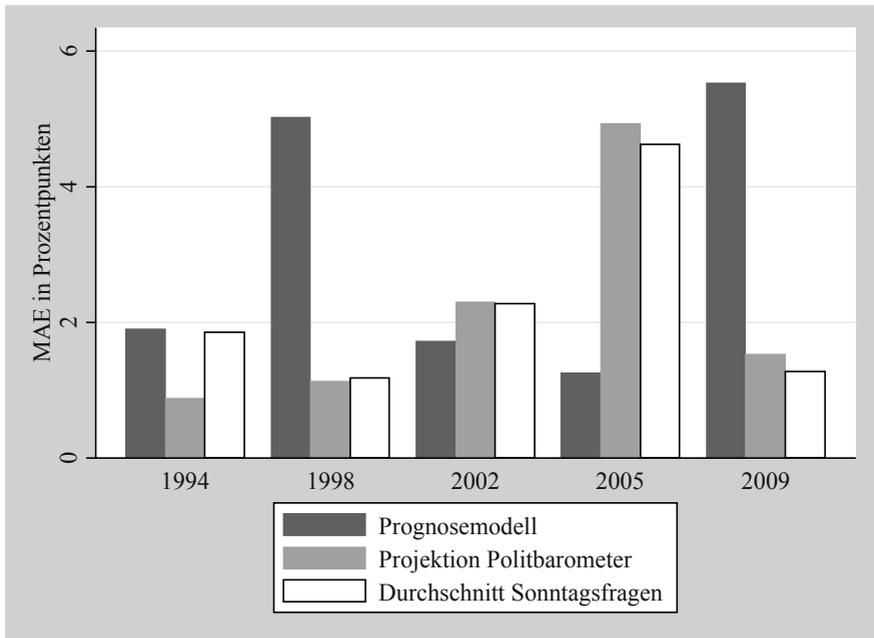
Betrachtet man zunächst die mittlere absolute Abweichung, so zeigt sich, dass die theoriebasierte Vorhersage der Stimmenanteile für die Bundestagswahlen 1998 und 2009 erhebliche absolute Abweichungen aufweisen (Abbildung 4.10). Dies ist 1998 auf das Ausmaß des Fehlers der Prognose des Stimmenanteils der CDU/CSU zurückzuführen. Hier wird der Stimmenanteil um 12,1 Prozentpunkte überschätzt. 2009 konnten mit dem Modell die Stimmenanteile von CDU/CSU, SPD und FDP nur unzureichend vorhergesagt werden.

1998 musste die CDU/CSU starke Stimmenverluste hinnehmen und landete erstmals seit 1949 unter der 40-Prozent-Marke. 2009 erlitt die SPD merkliche Stimmverluste und die FDP gewann überproportional hinzu. Dies kann durch das Prognosemodell insofern nicht erfasst werden, als es sich maßgeblich auf die langfristige Parteibindung stützt. Auch der Erwartungswert nach Mosteller

(1965) resultiert aus dem Stimmenanteilsranking vergangener Wahlen, weshalb etwa für 1998 erwartet wurde, dass die CDU/CSU die stärkste Kraft im Parlament werden würde. Gleichfalls konnte der Abnutzungs- und Kandidateneffekt diesen Stimmenverlust nicht in der ausgefallenen Höhe vorhersagen.

Abgesehen von diesem im Betrachtungszeitraum und für alle Parteien ungewöhnlich schlechten Abschneiden des Modells, erscheint es durchaus konkurrenzfähig. So ist der MAE bei den Bundestagswahlen 2002 und 2005 deutlich niedriger als derjenige der Politbarometer-Projektion, welche im Übrigen im Vergleichszeitraum fast durchgehend besser sind als der Durchschnitt aller Institute. Bei der Bundestagswahl 1994 schneidet das Prognosemodell geringfügig schlechter ab als das Politbarometer, liegt jedoch gleichauf mit den Instituten.

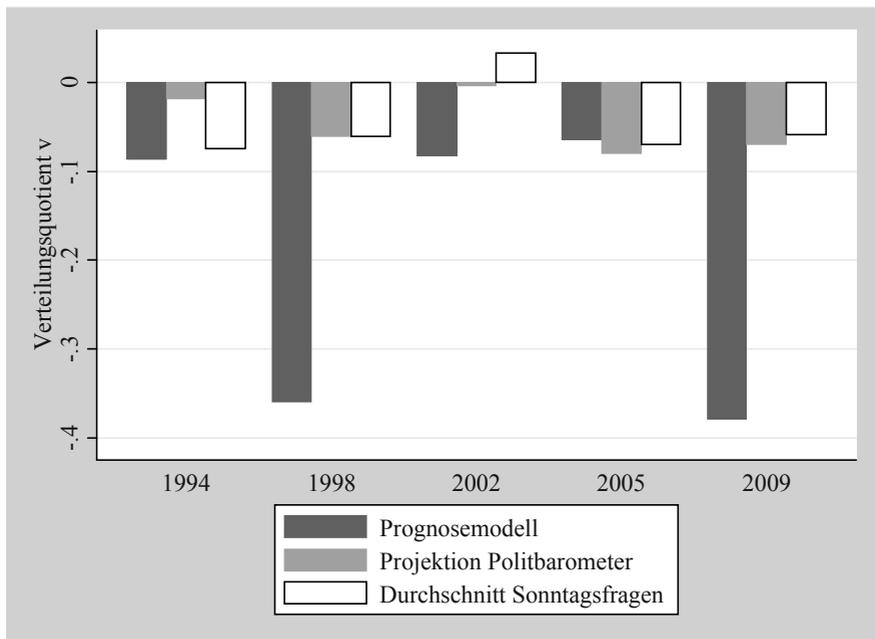
Abbildung 4.10: Vergleich der Prognosegüte des Modells mit wahlabsichtsfragenbasierten Vorhersagen nach Bundestagswahl, mittlerer absoluter Fehler, 1994-2009



Anmerkungen: Eigene Berechnungen und Darstellung. Die Partei PDS/Die Linke wurde nicht berücksichtigt.

Der Vergleichsquotient v auf Basis des Herfindahl-Indexes liefert ähnliche Befunde in Bezug auf die Prognosequalität des Modells (Abbildung 4.11). Die schlechte Performanz des Modells bei den Bundestagswahlen 1998 und 2009 fällt auch hier deutlich aus, da dem Modell zufolge der gemeinsame Stimmenanteil der beiden großen Parteien erheblich überschätzt wird und damit eine deutlich höhere Konzentration vermutet wird. Die Unterschiede sind bei den anderen Wahlen zwischen den einzelnen Vorhersageansätzen deutlich geringer, jedoch ist durchgängig eine Überschätzung des gemeinsamen Stimmenanteils von CDU/CSU und SPD erkennbar. Das Prognosemodell kann bei Betrachtung dieser Maßzahl die Sonntagsfragen nur bei der Bundestagswahl 2005 schlagen, liegt jedoch auch 1994 gleichauf mit dem Durchschnitt aller Sonntagsfragen.

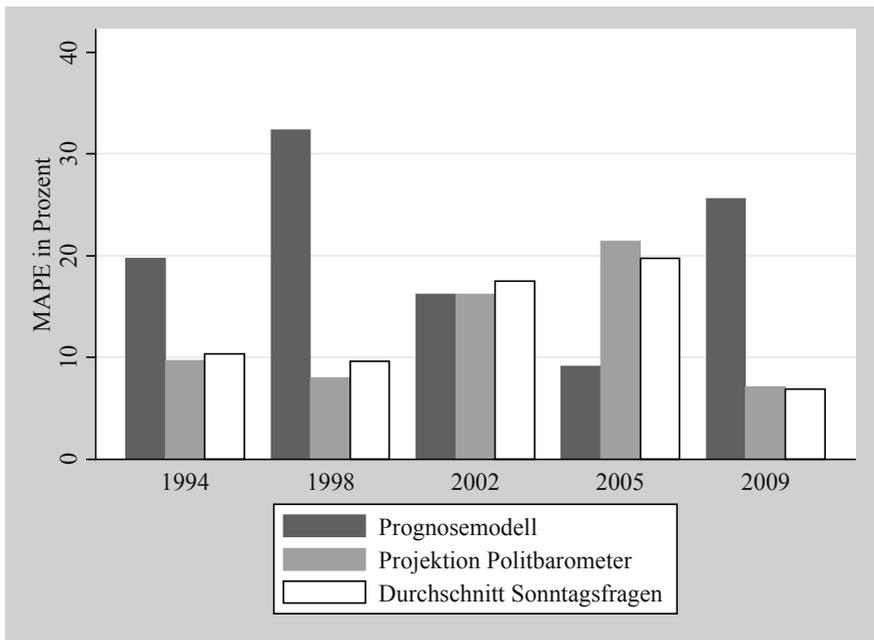
Abbildung 4.11: Vergleich der Prognosegüte des Modells mit wahlabsichtsbasierten Vorhersagen nach Bundestagswahl, Vergleichsquotient v , 1994-2009



Anmerkungen: Eigene Berechnungen und Darstellung. Die Partei PDS/Die Linke wurde nicht berücksichtigt.

Die Betrachtung der mittleren absoluten prozentualen Abweichung zeichnet ebenfalls kein positiveres Bild für das Modell beim Vergleich der Prognosegüte für einzelne Bundestagswahlen (Abbildung 4.12). Hier schlägt das Prognosemodell bei zwei der drei betrachteten Wahlen die Vorhersagen auf Basis der Wahlabsichtsfragen.

Abbildung 4.12: Vergleich der Prognosegüte des Modells mit wahlabsichtsfragebasierten Vorhersagen nach Bundestagswahl, mittlerer absoluter prozentualer Fehler, 1994-2009

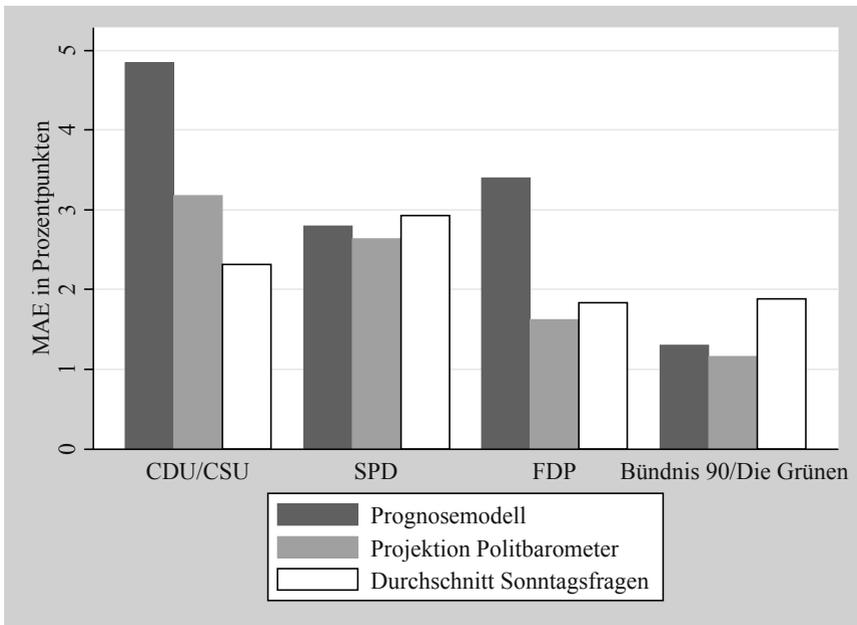


Anmerkungen: Eigene Berechnungen und Darstellung.

In Abbildung 4.13 sind die mittleren absoluten Abweichungen der verschiedenen Vorhersageansätze für die Bundestagswahlen 1994 bis 2009 nach Parteien abgetragen. Auch hier zeigt sich, dass das entwickelte Modell konkurrenzfähig ist. Die größeren Abweichungen bei der CDU/CSU sind auf die Bundestagswahlen 1998 und 2009 zurückzuführen, bei denen das Modell die Stimmenanteile deutlich

überschätzt. Die Stimmenanteile der SPD werden durch das Modell deutlich besser approximiert als durch die Wahlabsichtsumfragen – zumindest wenn man die Bundestagswahl 2009 nicht berücksichtigt. Hier hat das Prognosemodell den Stimmenanteil ebenfalls deutlich überschätzt, wie auch den von CDU/CSU. Dieser wird vom Modell von Norpoth und Gschwend (2009) noch deutlicher überschätzt: So wird bei einer Unterstützung von Angela Merkel von 60 Prozent (Forschungsgruppe Wahlen 2009a) ein Stimmenanteil von 42,8 Prozent für die CDU/CSU vorhergesagt. Aus dem hier vorgeschlagenen Modell resultiert ein Stimmenanteil von 40,6 Prozent. Bei diesem Vergleich muss jedoch berücksichtigt werden, dass es sich bei dem Wert von Norpoth und Gschwend um eine *ex-ante*-Prognose handelt, während die Angabe aus diesem Modell lediglich eine *out-of-sample*-Prognose darstellt.

Abbildung 4.13: Vergleich der Prognosegüte des Modells mit wahlabsichtsbasierten Vorhersagen nach Parteien, mittlerer absoluter Fehler, 1994-2009



Anmerkungen: Eigene Berechnungen und Darstellung.

Die retrospektive Betrachtung hat große Abweichungen für die FDP gezeigt (Abschnitt 4.2.1), diese fallen dementsprechend auch größer aus als die Abweichungen bei den Sonntagfragen. Hervorzuheben sind sicher die guten Vorhersagen für Bündnis 90/Die Grünen – für diese Partei liefert das Modell im Schnitt bessere Vorhersagen der Stimmenanteile als die Sonntagfragen.

Dies gilt gleichermaßen bei Betrachtung der mittleren absoluten prozentualen Abweichung (Abbildung 4.14). Hier zeigt sich, dass trotz der erheblichen Fehlprognose des Modells des Stimmenanteils der CDU/CSU bei der Bundestagswahl 1998 der MAPE nur geringfügig größer ausfällt als bei den Vergleichsgrößen. Allerdings gelingt es auch mit dem Prognosemodell (bisher) nicht, wesentlich bessere Vorhersagen der Stimmenanteile der kleinen Parteien zu liefern. Dies wurde in Abschnitt 3.4.2 aus empirischer Sicht als ein zentrales Manko der Sonntagfragen diagnostiziert. Zwar ist das Modell für Bündnis 90/Die Grünen als wettbewerbsfähig zu bezeichnen, aber insbesondere für die FDP sind die prozentualen Abweichungen noch einmal beträchtlich größer als für die Wahlabsichtsumfragen.

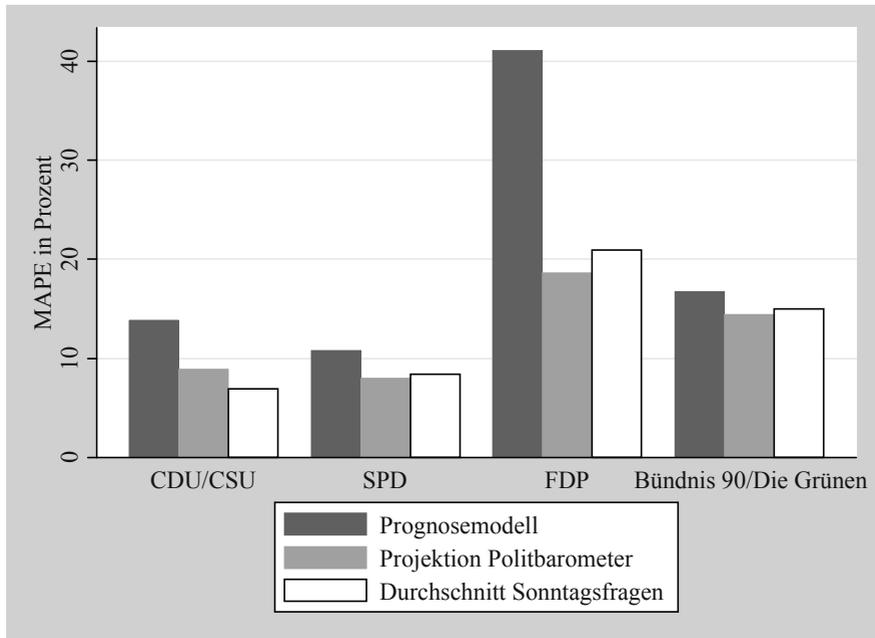
Insgesamt hat diese Betrachtung gezeigt, dass dieser erste Vorschlag für ein Modell zur Prognose von einzelnen Stimmenanteilen auf Grundlage von bisher nur wenigen empirischen Daten als durchaus konkurrenzfähig mit den Wahlabsichtsfragen anzusehen ist. Nichtsdestotrotz offenbaren die Befunde auch zentrale Schwächen des Modells.

Wesentlicher Einflussfaktor ist die langfristige Parteibindung. Damit sind *erstens* große Schwankungen bei den Wahlergebnissen nur unzureichend vorhersehbar – dies ist an der vergleichsweise schlechten Vorhersage der Bundestagswahlen 1998 und 2005 durch das Modell zu sehen. 1998 büßte die CDU/CSU gegenüber der Bundestagswahl 1994 6,3 Prozentpunkte ein, 2009 verlor die SPD 11,2 Prozentpunkte. Diese deutlichen Verluste konnten durch das Modell nicht abgebildet werden – hier müssen sicher weitere kurzfristig wirkende Einflussfaktoren spezifiziert und eingebunden werden, um diese Schwankungen besser erfassen zu können. Dies ist zumindest bisher sicher besser auf Grundlage von umfragebasierten Vorhersagen möglich.

Zweitens vernachlässigt das Modell die Möglichkeit strategischen Wählens. Dies führt vermutlich insbesondere bei der Schätzung der Stimmenanteile der FDP zu deutlichen Abweichungen. Die FDP hat wohl gerade auch im Rahmen der Bundestagswahl 2009 von Leihstimmen von CDU/CSU-Anhängern profitiert

(Weßels 2009). Gerade vor dem Hintergrund, dass strategisches Wählen zunehmend in den Fokus der deutschen Wahlforschung rückt (vgl. bspw. Behnke et al. 2004; Gschwend 2007; Gschwend/Pappi 2004; Hilmer/Schleyer 2000; Pappi/Herrmann 2006; Shikano et al. 2006; Thurner/Pappi 1998), kann man in der Zukunft versuchen, diese Erkenntnisse in ein Prognosemodell zu integrieren.

Abbildung 4.14: Vergleich der Prognosegüte des Modells mit wahlabsichtsfragenbasierten Vorhersagen nach Parteien, mittlerer absoluter prozentualer Fehler, 1994-2009



Anmerkungen: Eigene Berechnungen und Darstellung.

Ein zentrales Manko des Modells ist *drittens* die Ausblendung von möglichen Veränderungen des Parteiensystems. Hierauf ist sicher zum Teil die „Fehlprognose“ der Bundestagswahl 2009 zurückzuführen. Mit der Partei Die Linke ist eine neu entstandene (Fusion der Parteien PDS und WASG) zur Wahl angetreten

und auch die Piratenpartei, welche immerhin 2,0 Prozent der Zweitstimmen erreichten, traten 2009 zum ersten Mal an. Entsprechende Veränderungen sind nicht im Modell abbildbar, sondern lediglich post hoc integrierbar – Norpoth und Gschwend (2009) reagieren auf entsprechende Veränderungen etwa mit ad-hoc-Modifikationen ihres Modells. Ähnliche ad-hoc-Modifikationen des Modells haben bereits zuvor Kritik hervorgerufen (Klein 2005b).

Schließlich ist *viertens* auf die bisher dünne empirische Basis hinzuweisen. Gerade für Bündnis 90/Die Grünen sowie die PDS/Die Linke liegen bisher nur sehr wenige Beobachtungspunkte vor. Insofern ist zu erwarten, dass mit zunehmender Verbreiterung der empirischen Basis entsprechende Prognosemodelle bessere Vorhersagen generieren können.

5 Zusammenfassung und Fazit

In der vorliegenden Arbeit wurden die zentralen Ansätze zur Vorhersage von Wahlergebnissen aus methodologischer und empirischer Sicht analysiert. In diesem abschließenden Teil werden die wesentlichen Befunde der Analysen herausgestellt und der Ertrag der Arbeit bilanziert.

5.1 Zusammenfassung zentraler Ergebnisse

Ausgangspunkt der Arbeit war die Feststellung, dass Vorhersagen von Wahlergebnissen eine große öffentliche Aufmerksamkeit zuteil wird und diese Prognosen häufig als wissenschaftlich klassifiziert werden. Die zentralen Ansätze hierzu stellen umfragebasierte Vorhersagen, Wahlbörsen sowie Prognosemodelle dar.

Ausgehend von *methodologischen Überlegungen* wurden diese Ansätze zunächst darauf hin geprüft, inwieweit sie wissenschaftliche Prognosen liefern und zum Erkenntnisfortschritt zur Erklärung des Wählerverhaltens beitragen können.

Die dominante Form von umfragebasierten Vorhersagen sind politisch gewichtete *Wahlabsichtsfragen* (Sonntagsfragen). Dieser Ansatz erfüllt die abgeleiteten Anforderungen nicht, da die angewendeten Gewichtungsverfahren nicht veröffentlicht werden und somit intersubjektiv nicht nachvollziehbar sind. Damit ist auf dieser Basis kumulative Forschung nicht möglich. Die zweite Form umfragebasierter Prognoseansätze beruht auf Fragen zur Erwartung des Wahlausgangs. Dieser Ansatz wurde in Deutschland bisher kaum genutzt. Aus methodologischer Sicht könnten *Wahlerwartungsfragen* brauchbare Vorhersagen liefern, da Surowieckis (2005) Weisheit der Vielen zufolge bei gegebener Heterogenität der Befragten und der Gewährleistung unabhängiger Entscheidungen der Probanden empirisch angemessene Schätzungen des Wahlergebnisses zu erwarten wären. Aus methodologischer Perspektive bleibt jedoch in diesem Ansatz der kausale Mechanismus, welcher zur optimalen Informationsaggregation führt, nebulös. Zudem ist mit der Anwendung von Wahlerwartungsfragen kein Erkenntnis-

gewinn für die empirische Wahlforschung verbunden; in diesem Kontext können nur Informationsverarbeitungsprozesse untersucht werden. *Wahltagsbefragungen* schließlich sind die von Meinungsforschungsinstituten gerne als Prognose bezeichnete Form der umfragebasierten Ansätze. Die methodologische Betrachtung zeigt jedoch, dass die Vorhersage nach Eintritt des Explanandums erstellt wird und somit nicht von einer wissenschaftlichen Prognose gesprochen werden kann.

Experimentelle Märkte werden seit Ende der 1980er Jahre zur Prognose von Wahlergebnissen herangezogen. Auch in Deutschland wurden bereits mehrfach erfolgreich *Wahlbörsen* zu Landtags- und Bundestagswahlen eingesetzt. Theoretische Basis experimenteller Märkte ist die Hypothese Hayeks (1945), wonach Wettbewerbsmärkte auch unvollständige und asymmetrisch verteilte Informationen optimal aggregieren und somit Vorhersagen zukünftiger Ereignisse möglich sind. Diesem Ansatz zufolge – und das ist der Unterschied zur These Surowieckis (2005) – stellt einzig der Wettbewerbsmechanismus eine effiziente Informationsaggregation sicher. Der Weisheit der Vielen zufolge ist der Aggregationsmechanismus nicht entscheidend. Experimentelle Märkte liefern damit zwar wissenschaftliche Prognosen, leisten jedoch ebenso wenig wie Wahlerwartungsfragen einen Beitrag zur empirischen Wahlforschung.

Die dritte Gruppe von Ansätzen zur Vorhersage von Wahlergebnissen stellen *theoretische Prognosemodelle* dar. Im Rahmen dieser Ansätze wird versucht auf Grundlage einschlägiger theoretischer sowie empirischer Befunde der Wahlforschung Prognosen von Wahlergebnissen zu generieren. Es wurde gezeigt, dass sich in den USA mittlerweile diese Ansätze fest in der Wahlforschung etabliert haben und Anstöße zur Theoriebildung zur Erklärung des Wählerverhaltens liefern. Ebenso wurde deutlich, dass verschiedene Ansätze nebeneinander bestehen: Regressionsbasierte, zeitreihenanalytische und indexbasierte Ansätze. Gerade diese Vielfalt und die entsprechend variierende theoretische Fundierung werden als sehr fruchtbar für die Weiterentwicklung der Disziplin wahrgenommen. Jedoch ist in Bezug auf die methodologischen Anforderungen ein gemischtes Fazit zu ziehen, da insbesondere die zeitreihenanalytischen und indexbasierten Vorschläge zumeist lediglich auf empirischen Befunden und/oder Plausibilitätsüberlegungen beruhen. Auch die regressionsbasierten Prognosemodelle verweisen häufig lediglich auf Korrelate auf der Makroebene und lassen die Spezifizierung von Wirkungsmechanismen vermissen.

Für Deutschland wurde festgestellt, dass zwar die Prognose des Wahlverhaltens durchaus als Ziel der Wahlforschung angesehen wird, jedoch ist bisher nur ein Versuch in diese Richtung unternommen worden. Das Modell von Gschwend und Norpoth (2001) hat zum Ziel, den gemeinsamen Stimmenanteil der amtierenden Regierungsparteien vorherzusagen. Wesentlich für die Zusammensetzung des Deutschen Bundestages und damit verbunden auch die Koalitionsoptionen zur Regierungsbildung ist der Stimmenanteil jeder einzelnen Partei; hierauf richtet sich auch das öffentliche Interesse, weshalb die Entwicklung von Modellen zur Vorhersage der Stimmenanteile einzelner Parteien notwendig erscheint.

Die Prüfung der *empirischen Prognosegüte* der Ansätze hat gezeigt, dass für Deutschland bisher keine befriedigende systematische Analyse von *Wahlabsichtsfragen* vorgelegt wurde. Dies betrifft gleichermaßen mögliche Einflussfaktoren hierauf. Dieser Forschungslücke wurde in dieser Arbeit begegnet, indem zunächst aus theoretischen und methodischen Überlegungen potenzielle Einflussfaktoren auf die Prognosegüte von Sonntagsfragen abgeleitet wurden.

Zentral hierbei erscheint die Berücksichtigung von *house effects*, da die Wahlabsichten institutsspezifisch und nicht intersubjektiv nachvollziehbar politisch gewichtet werden. Weiterhin wurde erwartet, dass die Vorhersagegenauigkeit positiv mit der Stichprobengröße, mit abnehmender Grenzrate positiv mit dem zeitlichen Abstand zur Wahl, positiv mit der Länge des Befragungszeitraums und positiv mit der Wahlbeteiligung variiert. Die Vorhersagen für die neuen Bundesländer sollten den Überlegungen zufolge im Durchschnitt ungenauer sein. Schließlich wurde die Hypothese eines umgekehrt u-förmigen Zusammenhangs zwischen der Prognosegüte und der Zeit hergeleitet.

Zum Test dieser Hypothesen wurde ein Datensatz zusammengestellt, welcher möglichst alle publizierten Sonntagsfragen von 1949 bis 2009 sowie die zusätzlichen Informationen zu den Einflussfaktoren umfasst. Die Hypothesen über den Einfluss methodischer Aspekte der Umfrage konnten in der empirischen Analyse nicht bestätigt werden. Dies betrifft den Stichprobenumfang ebenso wie den Befragungszeitraum. Weitgehende Konformität zwischen den formulierten Hypothesen und den empirischen Befunden besteht hingegen bei den anderen Zusammenhängen. Die Prognosegenauigkeit ist besser bei höherer Wahlbeteiligung, steigt mit zunehmender Nähe zum Wahltermin und weist im Zeitverlauf den postulierten umgekehrt u-förmigen Zusammenhang auf.

Kontraintuitiv ist jedoch der Befund für die neuen Bundesländer. Hier ist die Prognosegenauigkeit höher als für die alten Bundesländer. Schließlich wurde erwartungsgemäß eine institutsspezifische Variation der Prognosegüte festgestellt, welche jedoch insgesamt eher gering ausfällt. In diesem Rahmen überraschte insbesondere, dass das IfD Allensbach trotz der Verwendung von Quotenstichproben nicht systematisch höhere Prognosefehler aufweist.

Die Befunde zum geringen Einfluss methodischer Faktoren werden zum einen auf das erhebliche Ausmaß der politischen Gewichtung zurückgeführt. Exemplarische Analysen mit den Politbarometer-Daten der Forschungsgruppe Wahlen zeigen, dass die publizierten Projektionen bis zu zehn Prozentpunkte von den in den Rohdaten ermittelten Stimmenanteilen abweichen. Der Einfluss der politischen Gewichtung, so die These, überlagert mögliche Effekte des methodischen Designs. Zum anderen zeigt sich, dass methodische Aspekte der Umfragen nach wie vor unzureichend dokumentiert sind. Für nicht einmal die Hälfte der erfassten Sonntagsfragen konnte etwa die Stichprobengröße ermittelt werden. Diese weist zudem eine vergleichsweise geringe Varianz auf. Auch aufgrund dessen überraschen die Befunde der Analyse nicht.

In deskriptiver Hinsicht erscheint wesentlich, dass die Sonntagsfragen eine durchaus passable Prognosegüte aufweisen und im Durchschnitt weitaus bessere Vorhersagen generieren als ihr Ruf nahe legt. Im Schnitt weichen die Vorhersagen um 2,1 Prozentpunkte vom tatsächlichen Stimmenanteil ab. Vor dem Hintergrund, dass hier alle Sonntagsfragen bis zu einem Jahr vor einer Wahl und über den gesamten Untersuchungszeitraum von 1949 bis 2009 enthalten sind, ist dies als eine durchaus beachtliche Prognosequalität zu werten. Weiterhin konnten keine institutsspezifischen systematischen Verzerrungen zugunsten oder zuungunsten bestimmter Parteien festgestellt werden. Die engen Verbindungen einzelner Meinungsforschungsinstitute zu bestimmten Parteien führen demnach nicht zu verzerrten Vorhersagen. Die Prognosequalität variiert wahlspezifisch beträchtlich und so lieferten die Institute bei einzelnen Wahlen, etwa der Bundestagswahl 2005, mit den Sonntagsfragen nur unzureichende Vorhersagen. Gerade im Anschluss an diese Wahlen wurde die Kritik an den Meinungsforschungsinstituten laut – anerkannt werden gelungene Vorhersagen, etwa der Ergebnisse der Bundestagswahlen 1998 oder auch 2009 im Rahmen der Nachwahlberichterstattung hingegen seitens der Medien selten. Schließlich – und das ist ein zentraler Befund dieser Analyse – zeigt sich, dass die Stimmenanteile kleiner Parteien

nur unzureichend vorhergesagt werden können. Die mittlere absolute prozentuale Abweichung beträgt für die kleinen Parteien über den gesamten Zeitraum 29,4 Prozent.

Der Vergleich mit der empirischen Prognosegüte von *Wahlbörsen* zeigt, dass zumindest in Deutschland die Wahlabsichtsfragen keine wesentlich schlechteren Vorhersagen erlauben als die experimentellen Märkte. Deren empirische Überlegenheit ist für die USA gut dokumentiert, wenngleich auch hier systematische Analysen bisher nur vereinzelt vorgelegt wurden. Eine offene Forschungsfrage für Deutschland ist insbesondere der Einfluss von Sonntagsfragen auf die Kursentwicklung von Wahlbörsen. Während häufig die Meinung vertreten wird, die Kurse der Wahlbörsen seien lediglich eine Reaktion auf die Sonntagsfragen, lassen die vorliegenden empirischen Befunde hier noch keinen eindeutigen Schluss zu.

Theoretische *Prognosemodelle* haben sich in den USA als zuverlässige Generatoren von Vorhersagen von Wahlergebnissen erwiesen. Auch das von Gschwend und Norpoth (2003, 2005) für Deutschland entwickelte Modell hat ersten Prüfungen Stand gehalten und den gemeinsamen Stimmenanteil der amtierenden Regierungsparteien exakter vorhergesagt als Umfragen, welche zum gleichen Zeitpunkt erhoben wurden. Deutlich ungenauer als die Umfragen war das Modell jedoch im Rahmen der Bundestagswahl 2009 (Gschwend 2009). Allerdings wurde bisher noch kein Modell vorgelegt, das die Vorhersage von Stimmenanteilen einzelner Parteien ermöglichte.

In dieser Arbeit wurde auf Grundlage einschlägiger theoretischer Überlegungen zur Erklärung des Wählerverhaltens sowie statistischer Gesetzmäßigkeiten ein solcher Vorschlag entwickelt. Das Modell orientiert sich hierbei an den regressionsbasierten Ansätzen der entsprechenden US-amerikanischen Forschung und berücksichtigt als Parameter für die CDU/CSU, die SPD sowie die FDP die langfristige Parteibindung, den erwarteten Stimmenanteil unter vollständigen Wettbewerbsbedingungen, die sozialstrukturelle Zusammensetzung der Bevölkerung, die Anzahl der Amtsperioden der amtierenden Regierung sowie den Kehrwert des Rangplatzes des jeweiligen Spitzenkandidaten in der öffentlichen Bewertung. Für Bündnis 90/Die Grünen sowie die PDS/Die Linke mussten aufgrund der beschränkten Datenlage abgewandelte Modelle entwickelt werden. Für Bündnis 90/Die Grünen wird auf einen Indikator zur sozialstrukturellen Zusammensetzung verzichtet und die Kandidatenorientierung als kurzfristig wirkender Faktor durch den Einfluss politischer Ereignisse ersetzt. Für die

PDS/Die Linke konnte nur die langfristige Parteibindung sowie die Entwicklung der Arbeitslosenquote herangezogen werden.

Empirisch wurde der Vorschlag mittels *out-of-sample*-Prognosen getestet. Hierbei zeigt sich eine passable Prognosequalität, die sich insgesamt als konkurrenzfähig mit den Sonntagsfragen erweist. Bei den Bundestagswahlen 1998 und 2009 resultieren aus dem Modell jedoch deutliche Abweichungen von den tatsächlichen Stimmenanteilen. Besonders hervorzuheben ist, dass die Stimmenanteile von Bündnis 90/Die Grünen auf Grundlage des Modells besser vorhergesagt werden als durch die Sonntagsfragen. Gleichwohl zeigt sich, dass insbesondere für die FDP eine nur unzureichende Prognosequalität erreicht wird.

5.2 Schlussfolgerungen und Ausblick

Für die zukünftige Forschung werden aus der Arbeit verschiedene Schlussfolgerungen gezogen. Die Arbeit hat gezeigt, dass es für die empirische Wahlforschung lohnend ist, sich intensiver als bisher mit Prognosen zu beschäftigen. Dies umfasst sowohl eine Etablierung einer empirischen Dauerbeobachtung existenter Vorhersageansätze als auch die Entwicklung neuer Ansätze zur Prognose von Wahlergebnissen.

Die hier vorgelegten Analysen zur Prognosegüte von Wahlabsichtsfragen haben gezeigt, dass eine systematische Evaluation lohnend erscheint. Erstens genießen Wahlabsichtsfragen eine anhaltend hohe mediale Aufmerksamkeit insbesondere in Wahlkampfzeiten. Dem entgegen steht eine häufig geäußerte Skepsis gegenüber dieser Methode sowie der Meinungsforschung im Allgemeinen. Dies betrifft neben der Negierung der Möglichkeit der Vorhersage von Wahlergebnissen auf Grundlage von Wahlabsichten auch den Vorwurf der Beeinflussung oder gar bewussten Manipulation der Wähler und/oder der politischen Entscheidungsträger.

Dennoch werden die durch kommerzielle Meinungsforschungsinstitute verbreiteten Umfrageergebnisse gemeinhin als wissenschaftlich wahrgenommen. Hierbei bezieht man sich zum einen auf die angewendeten Methoden sowie die häufig enge personelle Verzahnung von Wissenschaft und Meinungsforschung. In diesem Zusammenhang kann die empirische Wahlforschung mit systematischen Analysen der Ergebnisse der Meinungsforschungsinstitute wertvolle Aufklärungsarbeit leisten, die nicht nur der kommerziellen, sondern auch der

akademischen Umfrageforschung nutzen. So zeigt diese Arbeit deutlich, dass häufig kolportierte Hypothesen bezüglich institutsspezifisch verzerrter Sonntagsfragen zugunsten oder zuungunsten bestimmter Parteien nicht haltbar sind. Ebenso wurde belegt, dass die empirische Prognosegüte der Wahlabsichtsfragen mit einer mittleren absoluten Abweichung von 2,1 Prozentpunkten pro Partei vom tatsächlichen Wahlergebnis deutlich besser ist als ihr Ruf. Überwiegend werden die Wahlergebnisse durch die Sonntagsfragen gut approximiert und es sind lediglich wahlspezifische Schwankungen festzustellen, jedoch keine systematischen Veränderungen über die Zeit.

Gleichwohl zeigen die Analysen, dass insbesondere methodische Aspekte bisher unzureichend beantwortet sind. Zwar wird der Disput um das angemessene Auswahlverfahren (Quoten- versus Zufallsstichprobe) nicht mehr geführt, kann jedoch keineswegs als abgeschlossen betrachtet werden. Auch wenn unbestritten ist, dass aus theoretischer Perspektive nur eine Legitimation für Zufallsverfahren existiert und inferenzstatistische Schlüsse nur auf deren Grundlage möglich sind, stellt sich die Frage nach der empirischen Relevanz dieses theoretischen Einwands. So ist die Erhebung von perfekten Zufallsstichproben aufgrund etwa von Non-Response-Problemen nicht realisierbar. Diese Ausfälle variieren – so die übereinstimmenden Befunde der Methodenforschung – systematisch mit Befragtenmerkmalen, weshalb auch aus theoretischer Perspektive die Anwendbarkeit von Inferenzstatistik auf so generierte „Zufallsstichproben“ durchaus bezweifelt werden kann. Zudem zeigt die empirische Analyse in dieser Arbeit, dass das IfD Allensbach, welches Quotenverfahren verwendet, im Schnitt keineswegs schlechtere Vorhersagen liefert als andere Institute, die Zufallsverfahren verwenden. Möglicherweise ist dies auf die politische Gewichtung zurückzuführen. Es ist denkbar, dass das IfD Allensbach aufgrund längerer Erfahrung auf bessere Gewichtungsfaktoren zurückgreifen kann und hier der Effekt des Stichprobenverfahrens überlagert wird. In diesem Zusammenhang ist drittens hervorzuheben, dass die Anwendung von Inferenzstatistik wie etwa die Berechnung von Fehlerintervallen aufgrund der nicht nachvollziehbaren Gewichtungsfaktoren bei der Erstellung der Wahlprognosen nicht mehr zulässig ist. Vor diesem Hintergrund erhält die Klärung der Frage nach der empirischen Relevanz des zugrunde liegenden Stichprobenverfahrens neue Bedeutung.

Auch im Hinblick auf mögliche Einflussfaktoren der Prognosegüte werfen die präsentierten Analysen Fragen auf. In der Arbeit konnten insbesondere die

vermuteten methodischen Effekte nicht bestätigt werden. Auch hier kann eine Überlagerung der Effekte durch die erhebliche politische Gewichtung als Ursache vermutet werden. Jedoch sind auch in vergleichbaren US-amerikanischen Studien etwa hinsichtlich des Effekts der Stichprobengröße widersprüchliche empirische Befunde festzustellen. Weitere systematische Analysen und auch detailliertere methodische Angaben zu den Umfragen wären hier wünschenswert.

Die Zeitreihenanalysen zum Einfluss der politischen Gewichtung haben gezeigt, dass das Ausmaß der Gewichtung mit dem zeitlichen Abstand zur Wahl variiert und insbesondere in wahlfernen Zeiten erheblich ist. Diese Arbeit hatte lediglich zum Ziel, den Effekt der Gewichtung anhand der Politbarometer-Daten exemplarisch zu beleuchten. Die Befunde weisen jedoch auf ein erhebliches Potenzial vertiefender Analysen hin. So könnte man Einflussfaktoren auf die politische Gewichtung im Zeitablauf untersuchen. Hier könnte man durch Betrachtung von Zeitreihen anderer Institute, etwa von Forsa, deren Rohdaten ebenfalls für wissenschaftliche Analysen bereitgestellt werden, mögliche *house effects* beleuchten, oder auch vertiefte Analysen zum Einfluss von einschneidenden Ereignissen beziehungsweise der Berichterstattung hierüber auf das Ausmaß der politischen Gewichtung analysieren.

Die Zeitreihenanalysen haben überdies gezeigt, dass die zugunsten der SPD geäußerten Wahlabsichten im Beobachtungszeitraum 1986 bis 2005 keinen Trend aufweisen. Angesichts der häufig beschriebenen strukturellen Krise der SPD aufgrund des sozialen Wandels, welcher ein Wegbrechen der traditionellen SPD-Wähler bedingt, sowie der neuen Konkurrenz durch die Partei Die Linke, überrascht dieser Befund. Die Analysen in dieser Arbeit schließen nur den Zeitraum bis 2005 ein. Die Linke wurde jedoch erst in diesem Jahr aus der Fusion von PDS/Die Linke und der WASG gegründet. Insofern wäre es interessant, entsprechende Zeitreihenanalysen über den hier betrachteten Zeitraum hinaus auszudehnen und auch Einflussfaktoren auf die Wahlabsicht im Zeitablauf systematisch zu untersuchen. Dies könnte an entsprechende zeitreihenanalytische Arbeiten zu Einflussfaktoren auf Regierungspopularität, wie sie im Kontext der *economic-voting*-Perspektive bereits vorliegen geleistet werden.

Neben der systematischen Analyse existenter Prognoseansätze wie der Wahlabsichtsfragen wurde in dieser Arbeit ein erster Vorschlag für ein Modell zur Prognose der Zweitstimmenanteile von CDU/CSU, SPD, FDP, Bündnis 90/Die Grünen sowie der PDS/Die Linke entwickelt. Die Aufarbeitung des For-

schaftsstands hat eine weitgehende Prognoseskepsis der empirischen Wahlforschung gezeigt. Zwar finden sich zahlreiche Hinweise und Forderungen, die letztlich Prognosen als ein Ziel der Disziplin ansehen, bisher sind jedoch kaum Bemühungen in diese Richtung zu erkennen.

Das vorliegende Modell von Gschwend und Norpoth (2001) zur Vorhersage des gemeinsamen Stimmenanteils der amtierenden Regierungsparteien sowie der im Rahmen dieser Arbeit präsentierte Vorschlag zeigen, dass die Entwicklung von Prognosemodellen für Deutschland möglich ist. In empirischer Hinsicht sind nach den ersten Befunden beide Modelle konkurrenzfähig zu anderen Ansätzen. Hervorzuheben ist, dass das hier präsentierte Modell der erste Vorschlag zur regressionsbasierten Vorhersage von Stimmenanteilen einzelner Parteien darstellt und trotz der bisher dünnen Datenbasis, der Stützung lediglich auf prozessproduzierte Daten und einer sehr einfachen Konstruktionsweise viel versprechend erscheint. Der Vergleich mit Wahlabsichtsfragen anhand von *out-of-sample*-Prognosen offenbart, dass bei den betrachteten Wahlen das Modell im Schnitt mit diesen mithalten kann. Bessere Vorhersagen konnten indes für Bündnis 90/Die Grünen abgeleitet werden. Insofern erscheint auch aus empirischer Perspektive die Intensivierung der Prognosebemühungen lohnend.

Auch in Bezug auf die Theorieentwicklung können Impulse erwartet werden. Diese Vermutung steht in Zusammenhang mit den evidenten Schwächen des präsentierten Modells. So weist die Vorhersage der Stimmenanteile der FDP große Abweichungen auf. Dies ist vermutlich darauf zurückzuführen, dass die FDP übermäßig von Leihstimmen und strategischem Wahlverhalten profitiert, welches in diesem Modellvorschlag nicht berücksichtigt werden konnte. Dies deutet auf eine zentrale Schwäche der Theoriebildung im Rahmen der empirischen Wahlforschung hin. Während die Entscheidung zugunsten oder zuungunsten der beiden großen Parteien mit den klassischen Erklärungsansätzen gut erklärt werden kann, tragen diese nur bedingt zur Erklärung der Wahlentscheidung zugunsten der kleinen Parteien bei. Hierzu ist auch in empirischer Hinsicht zu konstatieren, dass sich die vorliegenden Studien überwiegend auf die Volksparteien beziehen. Insofern ist für die Zukunft der empirischen Wahlforschung eine zentrale Aufgabe darin zu sehen, Überlegungen etwa zum strategischen Wählen in vorliegende theoretische Ansätze zu integrieren, um einen einheitlichen Ansatz zur Erklärung des Wählerverhaltens zu generieren.

Entsprechende Bemühungen könnten beispielsweise an das polit-ökonomische Modell von Merrill III und Grofman (1999) anschließen, das um die Berücksichtigung strategischer Überlegungen von Wählern in Mehrparteiensystemen mit gemischten Wahlsystemen erweitert werden müsste. Für einen umfassenden Test dieses Modells liegen bisher nur unzureichend Daten vor. Allerdings ist in jüngerer Zeit ein verstärktes Interesse an der Berücksichtigung strategischer Überlegungen der Wähler zu konstatieren. Darüber hinaus ist mit der Etablierung einer dauerhaften deutschen nationalen Wahlstudie durch die 2007 gegründete Deutsche Gesellschaft für Wahlforschung die Bereitstellung einer breiteren und kontinuierlichen Datenbasis sichergestellt, welche auch die empirische Überprüfung solcher Modelle erlaubt.

Weiterhin ist in Bezug auf das vorgelegte Prognosemodell zu konstatieren, dass es die zu Beginn der Arbeit aufgestellten Anforderungen an eine wissenschaftliche Prognose nur unzureichend erfüllen kann. Gefordert wurde insbesondere der Verweis auf korrekte kausale Mechanismen. Das vorgelegte Modell verweist jedoch auf Korrelate auf Aggregatebene und mögliche Mechanismen wurden eher implizit benannt. Hieraus können zwei Interpretationen folgen.

Zunächst wäre bei Berücksichtigung der wissenschaftstheoretischen Anforderungen zu konstatieren, dass das vorgelegte Modell unzureichend und nicht als wissenschaftliche Prognose zu klassifizieren ist. Dies trifft damit im Übrigen auf die meisten der bisher im Rahmen der US-amerikanischen Wahlforschung vorgelegten Modelle zu. Aus dieser Perspektive ergibt sich für alle der hier angesprochenen Ansätze, dass allenfalls die Wahlbörsen mit dem Verweis auf den Marktmechanismus als wissenschaftliche Prognosen zu qualifizieren sind. In Bezug auf das hier entwickelte Prognosemodell ist dementsprechend eine stärkere Theoretisierung der angestellten Überlegungen erforderlich.

Andererseits könnte man vor dem Hintergrund der empirisch festgestellten Brauchbarkeit von Prognosemodellen im Allgemeinen und auch der Sonntagsfragen konstatieren, dass die von beispielsweise Bunge (1997, 2004) genannten wissenschaftstheoretischen Anforderungen wenig praktische Relevanz besitzen, da empirisch befriedigende Prognosen offenbar auch ohne Erfüllung aller genannten Anforderungen möglich erscheinen. Aus dieser Perspektive wird die praktische Bedeutung methodologischer Überlegungen hinterfragt.

Hier wird der ersten Interpretation gefolgt, da eine Aufgabe der Orientierung an diesen Kriterien die Abgrenzung von nicht-wissenschaftlichen und

wissenschaftlichen Prognosen zumindest erschweren würde, da klare Bewertungskriterien fehlen. Zudem erscheint auch in Bezug auf Erklärungen die Festlegung auf den skizzierten methodologischen Rahmen zwingend um kumulativen Erkenntnisfortschritt in der Wissenschaft zu ermöglichen. Insofern erscheint die detailliertere Ausarbeitung der theoretischen Bezüge, die als Grundlage des Prognosemodells herangezogen wurden notwendig. An dieser Stelle wird vermutet, dass polit-ökonomische Modelle wie der bereits erwähnte Vorschlag von Merrill III und Grofman (1999) hier als Ausgangspunkt einer integrierten Theorie des Wählerverhaltens mit Verweis auf kausale Mechanismen besonderes Potenzial aufweisen.

Diese Einschätzung folgt zunächst der Überlegung, dass eine zufriedenstellende Theorie zur Erklärung des Wählerverhaltens sowohl das politische Angebot als auch die Nachfrage nach diesen berücksichtigen muss. Nur die ökonomischen Ansätze berücksichtigen die Angebotsseite, während die Überlegungen der Ann-Arbor- sowie der Columbia-Schule lediglich die Nachfrageseite beleuchten. Die Relevanz der Angebotsseite wurde nicht nur in theoretischer Hinsicht bisher zu wenig beleuchtet, auch liegen nur vereinzelte empirische Befunde hierzu vor; diese zeigen jedoch die Relevanz dieser auf. Weiterhin ist darauf hinzuweisen, dass das Modell Erweiterungen zulässt und für zentrale Einflussfaktoren wie etwa der Parteiidentifikation mittlerweile verschiedene Reinterpretationen aus Perspektive des ökonomischen Ansatzes vorliegen. Zu denken ist hier etwa an Fiorinas (1981) Konzept der Parteiidentifikation als Ergebnis wiederholter retrospektiver subjektiver Leistungsbewertungen und Wahlhandlungen in der Vergangenheit und Pappi und Shikanos (2007) Formalisierung dessen mittels Bayesianischem Update. Aufgrund dieser Flexibilität polit-ökonomischer Modelle im Hinblick auf Erweiterungen ist auch eine Integration strategischer Überlegungen denkbar. Schließlich ist der Ansatz von Merrill III und Grofman (1999) bereits explizit als diskretes Entscheidungsmodell formuliert. Damit kann auch in empirischer Hinsicht durch die Anwendung konditionaler Logitmodelle die bisher dominante und unterkomplexe Verwendung von binär logistischen Modellen zur empirischen Überprüfung der theoretischen Ansätze zur Erklärung einer multinomialen Entscheidung auf ein angemessenes statistisches Fundament gestellt werden.

Trotz dieser Einwände und benannten Schwächen des Modells wird insgesamt die Verstärkung von Prognosebemühungen in der Wahlforschung als notwendig erachtet. Zunächst erscheint dies aus professionspolitischer Sicht

lohnend. Prognosen von Wahlergebnissen werden stark nachgefragt und bisher vorrangig von kommerziellen Meinungsforschungsinstituten geliefert. Diese berufen sich dabei mit Bezug auf die Umfrageforschung auf die Wissenschaftlichkeit ihrer Methoden. Durch die immer wieder auftretenden „Fehlprognosen“ durch die Demoskopieinstitute sieht sich in der Folge die gesamte Umfrageforschung Kritik ausgesetzt. Neben den Vorhersagen der Demoskopen gewinnen Wahlbörsen als Prognoseinstrument an Bedeutung. Diese werden insbesondere von Ökonomen favorisiert und bereit gestellt, wodurch diese zunehmend in Kernkompetenzbereiche der Politikwissenschaften und der Soziologie vordringen und die Konkurrenzsituation der Nachbardisziplinen verschärfen.

Vor dem Hintergrund dieser Konstellation kann die Wahlforschung nur durch das Nutzen der eigenen Erkenntnisse zur Entwicklung eigener Prognosemodelle und damit einer offeneren Kommunikation über die Stärken und Schwächen dieser Versuche gewinnen. Von der Intensivierung der Methodenforschung im Hinblick auf die Erhebung politischer Einstellungen würde die gesamte Umfrageforschung profitieren. Die Verstärkung disziplineigener Prognosebemühungen würde zum einen die Erschließung neuerer Ressourcen ermöglichen und zum anderen das Profil der empirischen Wahlforschung in der Öffentlichkeit stärken.

Weiterhin wird aus den ersten empirischen Befunden zum Prognosemodell dieser Arbeit die Zuversicht abgeleitet, dass mittel- bis langfristig die Erstellung konstant sehr guter Prognosen möglich ist. Ein zentrales Problem für die Entwicklung von Prognosemodellen zu Bundestagswahlen stellt die (bisher) geringe Datenbasis dar. Da grundsätzlich davon ausgegangen wird, dass die herangezogenen Theorien nicht nur zur Erklärung des Wählerverhaltens auf Bundesebene Relevanz besitzen, wäre es möglich sich bei zukünftigen Modellen auch auf Landtagswahlen zu stützen. Hier müssten sicherlich teilweise landesspezifische Regelungen im Wahlsystem und andere Parteienkonstellationen berücksichtigt werden. Dennoch ist zu erwarten, dass die spezifizierten Variablen auch bei Landtagswahlen einflussreich sind. Bei Berücksichtigung von Landtagswahlen läge eine deutlich breitere Datenbasis, welche auch umfangreichere empirische Tests von Prognosemodellen als die hier präsentierten ermöglichen würde. Die damit einhergehende Ausdifferenzierung der Prognoseansätze könnte eine Etablierung eines in den USA vergleichbaren Wettbewerbs verschiedener Modelle nach sich ziehen. Hiervon ist insbesondere eine stetige Verbesserung der Prognosekompetenz der empirischen Wahlforschung zu erwarten.

Literaturverzeichnis

- Abramowicz, Michael (2007): Predictocracy. Market Mechanisms for Public and Private Decision Making. New Haven u. a.: Yale University Press.
- Abramowitz, Alan I. (1988): An Improved Model for Predicting Presidential Election Outcomes. In: PS: Political Science and Politics 21: 843-847.
- Abramowitz, Alan I. (1994): Is it Time for Them to Go? Forecasting the 1992 Presidential Election. In: The Political Methodologist 5: 2-3.
- Abramowitz, Alan I. (1996): Bill and Al's Excellent Adventure. Forecasting the 1996 Presidential Election. In: American Politics Quarterly 24: 434-442.
- Abramowitz, Alan I. (2004): When Good Forecasts Go Bad: The Time-for-Change Model and the 2004 Presidential Election. In: PS: Political Science and Politics 37: 745-746.
- Abramowitz, Alan I. (2008a): Forecasting the 2008 Presidential Election with the Time-for-Change Model. In: PS: Political Science & Politics 41: 691-695.
- Abramowitz, Alan I. (2008b): It's About Time: Forecasting the 2008 Presidential Election with the Time-for-Change Model. In: International Journal of Forecasting 24: 209-217.
- Adamic, Lada A./Huberman, Bernardo A. (2002): Zipf's Law and the Internet. In: Glottometrics 3: 143-150.
- Adorno, Theodor W. (1952): Zur gegenwärtigen Stellung der empirischen Sozialforschung in Deutschland. In: Institut zur Förderung öffentlicher Angelegenheiten (Hrsg.): Empirische Sozialforschung. Meinungs- und Marktforschung. Methoden und Probleme. Frankfurt am Main: Institut zur Förderung öffentlicher Angelegenheiten: 27-39.
- Aggarwal, Rajesh K./Wu, Guojun (2003): Stock Market Manipulation – Theory and Evidence. Ann Arbor: University of Michigan. Online: http://www.afajof.org/pdfs/2004program/UPDF/P306_Asset_Pricing.pdf (26. November 2008).
- Albrecht, Richard (1988): Demoskopie als Demagogie. Umfragen aus Allensbach und ihre Konsequenzen. In: Medien-Journal: Zeitschrift für Kommunikationskultur 12: 41-48.
- Aldrich, John H. (1993): Rational Choice and Turnout. In: American Journal of Political Science 37: 246-278.

- Anders, Manfred (1985): Sinkende Ausschöpfungsquoten – und was man dagegen tun kann. In: Kaase, Max/Küchler, Manfred (Hrsg.): Herausforderungen der empirischen Sozialforschung. Beiträge aus Anlaß des zehnjährigen Bestehens des Zentrums für Umfragen, Methoden und Analysen. Mannheim: ZUMA: 75-80.
- Anderson, Christopher J. (2007): The End of Economic Voting? Contingency Dilemmas and the Limits of Democratic Accountability. In: *Annual Review of Political Science* 10: 271-296.
- Ansolabehere, Stephen/Iyengar, Shanto (1994): Of Horseshoes and Horse Races: Experimental Studies of the Impact of Poll Results on Electoral Behavior. In: *Political Communication* 11: 413-430.
- Antholz, Birger (2001): Validität deutscher Wahlprognosen 1949-2000. Hamburg: Universität Hamburg.
- Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e. V. (ADM) (Hrsg.) (1979): Muster-Stichproben-Pläne für Bevölkerungs-Stichproben in der Bundesrepublik Deutschland und West-Berlin. München: Verlag Moderne Industrie.
- Armington, Klaus (1994): Gründe und Folgen geringer Wahlbeteiligung. In: *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 46: 43-64.
- Armstrong, J. Scott/Collopy, Fred (1992): Error Measures for Generalizing About Forecasting Methods: Empirical Comparisons. In: *International Journal of Forecasting* 8: 69-80.
- Armstrong, J. Scott/Graefe, Andreas (2009): Predicting Elections from Biographical Information about Candidates. Philadelphia: Wharton School Online: <http://marketing.wharton.upenn.edu/documents/research/PollyBio25.pdf> (3. Januar 2010).
- Armstrong, J. Scott/Cuzán, Alfred/Jones Jr., Randall J. (2008): Political Forecasting. Philadelphia: University of Pennsylvania. Online: <http://www.forecastingprinciples.com/Political/index.html> (28. August 2008).
- Arzheimer, Kai (2006): ‚Dead men walking?‘ Party Identification in Germany, 1977-2002. In: *Electoral Studies* 25: 791-807.
- Arzheimer, Kai/Falter, Jürgen W. (2005): ‚Goodbye Lenin?‘ Bundes- und Landtagswahlen seit 1990: Eine Ost-West-Perspektive. In: Falter, Jürgen W./Gabriel, Oscar W./Wessels, Bernhard (Hrsg.): *Wahlen und Wähler. Analysen aus Anlaß der Bundestagswahl 2002*. Wiesbaden: VS Verlag für Sozialwissenschaften: 244-283.
- Arzheimer, Kai/Schoen, Harald (2005): Erste Schritte auf kaum erschlossenen Terrain. Zur Stabilität der Parteiidentifikation in Deutschland. In: *Politische Vierteljahresschrift* 46: 629-654.

- Arzheimer, Kai/Schoen, Harald (2007): Mehr als eine Erinnerung an das 19. Jahrhundert? Das sozioökonomische und das religiös-konfessionelle Cleavage und Wahlverhalten 1994-2005. In: Rattinger, Hans/Gabriel, Oscar W./Falter, Jürgen W. (Hrsg.): Der gesamtdeutsche Wähler. Stabilität und Wandel des Wählerverhaltens im wiedervereinigten Deutschland. Baden-Baden: Nomos: 89-112.
- Atteslander, Peter (1980): Demoskopie: Hilfe oder Gefahr für die Politik? In: Politische Studien 31: 135-142.
- Bafumi, Joseph/Shapiro, Robert Y. (2009): A New Partisan Voter. In: The Journal of Politics 71: 1-24.
- Baltagi, Badi H. (1999): Econometrics. 2. Aufl. Berlin u. a.: Springer.
- Barrot, Christian (2007): Prognosegütemaße. In: Albers, Sönke/Klapper, Daniel/Konradt, Udo/Walter, Achim/Wolf, Joachim (Hrsg.): Methodik der empirischen Forschung. 2. Aufl. Wiesbaden: Gabler: 417-430.
- Bartels, Larry M./Zaller, John (2001): Presidential Vote Models: A Recount. In: PS: Political Science & Politics 34: 9-20.
- Bartolini, Stefano/Mair, Peter (2007): Identity, Competition, and Electoral Availability: The Stabilisation of European Electorates, 1885-1985. Colchester: ECPR Press.
- Bassili, John N. (1993): Response Latency versus Certainty as Indexes of the Strength of Voting Intentions in a CATI Survey. In: Public Opinion Quarterly 57: 54-61.
- Bassili, John N. (1995): Response Latency and the Accessibility of Voting Intentions: What Contributes to Accessibility and How it Affects Vote Choice. In: Personality and Social Psychology Bulletin 21: 686-695.
- Bassili, John N./Bors, Douglas A. (1997): Using Response Latency to Increase Lead Time in Election Forecasting. In: Canadian Journal of Behavioural Science 29: 231-238.
- Beck, Ulrich (1983): Jenseits von Klasse und Stand? Soziale Ungleichheit, soziale Individualisierungsprozesse und die Entstehung neuer sozialer Formationen und Identitäten. In: Kreckel, Reinhard (Hrsg.): Soziale Ungleichheiten. Soziale Welt, Sonderband 2. Göttingen: Schwartz: 35-74.
- Beck, Ulrich (1986): Risikogesellschaft. Auf dem Weg in eine andere Moderne. Frankfurt am Main: Suhrkamp.
- Beckmann, Klaus/Werding, Martin (1994): Markets and the Use of Knowledge. Testing the „Hayek Hypothesis“ in Experimental Stock Markets. Passau: Universität Passau. Online: <http://129.3.20.41/eps/exp/papers/9405/9405001.pdf> (2. Mai 2008).
- Beckmann, Klaus/Werding, Martin (1996): ‚Passauer Wahlbörse‘: Information Processing in a Political Market Experiment. In: Kyklos 49: 171-204.

- Beckmann, Klaus/Werding, Martin (1998): Eine Anmerkung zur „Hayek-Hypothese“ in der Experimentellen Ökonomie. In: *Jahrbücher für Nationalökonomie und Statistik* 217: 753-763.
- Behnke, Joachim (2007): *Das Wahlsystem der Bundesrepublik Deutschland. Logik, Technik und Praxis der Verhältniswahl*. Baden-Baden: Nomos.
- Behnke, Joachim/Hergert, Stefani/Bader, Florian (2004): *Stimmensplitting – Kalkuliertes Wahlverhalten unter den Bedingungen der Ignoranz*. Bamberg: Universität Bamberg.
- Bélanger, Éric/Lewis-Beck, Michael S./Nadeau, Richard (2005): A Political Economy Forecast for the 2005 British General Election. In: *British Journal of Politics and International Relations* 7: 191-198.
- Benda, Ernst (1974): *Demoskopie und Recht*. In: Noelle, Elisabeth/Neumann, Erich P. (Hrsg.): *Jahrbuch der öffentlichen Meinung 1968-1973*. Allensbach: Verlag für Demoskopie: XXIX-XL.
- Benda, Ernst (1981): *Meinungsforschung und repräsentative Demokratie*. In: Baier, Horst/Kepplinger, Hans M./Reumann, Kurt (Hrsg.): *Öffentliche Meinung und sozialer Wandel. Public Opinion and Social Change*. Für Elisabeth Noelle-Neumann. Opladen: Westdeutscher Verlag: 96-104.
- Berelson, Bernard R./Lazarsfeld, Paul F./McPhee, William N. (1986): *Voting. A Study of Opinion Formation in a Presidential Campaign*. Chicago u. a.: University of Chicago Press.
- Berg, Joyce E./Forsythe, Robert/Rietz, Thomas A. (1996): *What Makes Markets Predict Well? Evidence from the Iowa Electronic Markets*. In: Albers, Wulf/Güth, Werner/Hammerstein, Peter/Moldovanu, Benny/Van Damme, Eric (Hrsg.): *Understanding Strategic Interaction: Essays in Honor of Reinhard Selten*. New York u. a.: Springer: 444-463.
- Berg, Joyce E./Forsythe, Robert/Nelson, Forrest/Rietz, Thomas (2000): *Results from a Dozen Years of Election Futures Markets Research*. IEM Working Paper. Iowa City: University of Iowa. Online: http://www.biz.uiowa.edu/iem/archive/BFNR_2000.pdf (9. Mai 2008).
- Berg, Joyce E./Nelson, Forrest/Rietz, Thomas (2003): *Accuracy and Forecast Standard Error of Prediction Markets*. IEM Working Paper. Iowa City: University of Iowa. Online: <http://www.biz.uiowa.edu/iem/archive/forecasting.pdf> (3. Oktober 2008).
- Berg, Joyce E./Forsythe, Robert/Nelson, Forrest/Rietz, Thomas A. (2008): *Results from a Dozen Years of Election Futures Markets Research*. In: Plott, Charles R./Smith, Vernon L. (Hrsg.): *Handbook of Experimental Economics Results, Volume 1*. Amsterdam: Horth-Holland: 742-751.
- Berger, Manfred (1973): *Parteiidentifikation in der Bundesrepublik*. In: *Politische Vierteljahresschrift* 14: 215-225.

- Berlemann, Michael (1999): Wahlprognosen: Politische Wahlbörsen versus traditionelle Meinungsforschung. *Dresdner Beiträge zur Volkswirtschaftslehre*. Dresden: Technische Universität Dresden. Online: <http://rcswww.urz.tu-dresden.de/~wpeconom/seiten/pdf/1999/ddpe199901.pdf> (3. Oktober 2008).
- Berlemann, Michael (2000): Wahlbörsen versus traditionelle Meinungsforschung: Die Sachsenwahl '99. In: *List Forum für Wirtschafts- und Finanzpolitik* 26: 51-70.
- Berlemann, Michael (2002): Comment on Gregor Brüggelambert and Arwed Crüger, Forrest Nelson and Reinhard Tietz, and Jörg Bochow, Peter Raupach and Mark Wahrenburg. What Can We Learn From Experimental Asset Markets? In: Bolle, Friedel/Lehmann-Waffenschmidt, Marco (Hrsg.): *Surveys in Experimental Economics. Bargaining, Cooperation and Election Stock Markets*. Heidelberg u. a.: Physica: 251-257.
- Berlemann, Michael/Schmidt, Carsten (2001): Predictive Accuracy of Political Stock Markets. Empirical Evidence from a European Perspective. *Diskussionspapier des Sonderforschungsbereichs 373*. Berlin: Humboldt Universität zu Berlin. Online: <http://www.sfb504.uni-mannheim.de/~cschmidt/paper/dpsfb200157.pdf> (17. März 2009).
- Blais, André (2004): How Many Voters Change Their Minds in the Month Preceding an Election? In: *PS: Political Science & Politics* 37: 801-803.
- Blais, André (2006): What Affects Voter Turnout? In: *Annual Review of Political Science* 9: 111-125.
- Blais, André/Gidengil, Elisabeth/Nevitte, Neil (2006): Do Polls Influence the Vote? In: Brady, Henry E./Johnston, Richard (Hrsg.): *Capturing Campaign Effects*. Ann Arbor: University of Michigan Press: 263-279.
- Blecha, Karl (1969): Wert und Unwert von Wahlprognosen. In: *Angewandte Sozialforschung*: 164-172.
- Bleek, Wilhelm (2001): *Geschichte der Politikwissenschaft in Deutschland*. München: C. H. Beck.
- Blumer, Herbert (1948): Public Opinion and Public Opinion Polling. In: *American Sociological Review* 13: 542-549.
- Bochow, Jörg/Raupach, Peter/Wahrenburg, Mark (2002): What Do Market Makers Achieve? Evidence from a Large Scale Experimental Stock Market. In: Bolle, Friedel/Lehmann-Waffenschmidt, Marco (Hrsg.): *Surveys in Experimental Economics. Bargaining, Cooperation and Election Stock Markets*. Heidelberg u. a.: Physica: 229-250.
- Bohm, Peter/Sonnegard, Joakim (1999): Political Stock Markets and Unreliable Polls. In: *Scandinavian Journal of Economics* 101: 205-222.

- Bolle, Friedel/Lehmann-Waffenschmidt, Marco (Hrsg.) (2002): *Surveys in Experimental Economics. Bargaining, Cooperation and Election Stock Markets*. Heidelberg u. a.: Physica.
- Bourdieu, Pierre (1993): *Soziologische Fragen*. Frankfurt am Main: Suhrkamp.
- Box-Steffensmeier, Janet M./Smith, Renée M. (1996): *The Dynamics of Aggregate Partisanship*. In: *American Political Science Review* 90: 567-580.
- Bradburn, Norman M./Sudman, Seymour (1988): *Polls & Surveys. Understanding What They Tell Us*. San Francisco: Jossey-Bass.
- Braun, Norman (2008): *Theorie in der Soziologie*. In: *Soziale Welt* 59: 373-395.
- Braun, Norman/Keuschnigg, Marc (2008): *Soziale Märkte und ihr Nachweis*. Unveröffentlichtes Manuskript. München. Ludwig-Maximilians-Universität München.
- Bretschneider, Rudolf (1985): *Demoskopie als Politikersatz. Gebrauch und Mißbrauch der politischen Meinungsforschung*. In: Plasser, Fritz/Ulram, Peter A./Welan, Manfred (Hrsg.): *Demokratierituale: zur politischen Kultur der Informationsgesellschaft*. Wien: Böhlau: 271-278.
- Bretschneider, Frank (1991): *Wahlumfragen. Empirische Befunde zur Darstellung in den Medien und dem Einfluß auf das Wahlverhalten in der Bundesrepublik Deutschland und den USA*. München: Minerva.
- Bretschneider, Frank (1992): *Der taktische und rationale Wähler. Über den Einfluss von Wahlumfragen auf das Wählerverhalten bei den Bundestagswahlen 1983 bis 1990*. In: *Politische Vierteljahresschrift* 33: 55-72.
- Bretschneider, Frank (1996): *Wahlumfragen und Medien. Eine empirische Untersuchung der Presseberichterstattung über Meinungsumfragen vor den Bundestagswahlen 1980 bis 1994*. In: *Politische Vierteljahresschrift* 37: 475-493.
- Bretschneider, Frank (2000): *Demoskopie im Wahlkampf – Leitstern oder Irrlicht?* In: Klein, Markus/Jagodzinski, Wolfgang/Mochmann, Ekkehard/Ohr, Dieter (Hrsg.): *50 Jahre Empirische Wahlforschung in Deutschland. Entwicklung, Befunde, Perspektiven, Daten*. Wiesbaden: Westdeutscher Verlag: 477-505.
- Bretschneider, Frank (2002a): *Spitzenkandidaten und Wahlerfolg. Personalisierung, Kompetenz, Parteien*. Wiesbaden: Westdeutscher Verlag.
- Bretschneider, Frank (2002b): *Die Medienwahl 2002. Themenmanagement und Berichterstattung*. In: *Aus Politik und Zeitgeschichte* 49/50: 36-47.
- Bretschneider, Frank (2005a): *Massenmedien und Wählerverhalten*. In: Falter, Jürgen W./Schoen, Harald (Hrsg.): *Handbuch Wahlforschung*. Wiesbaden: VS Verlag für Sozialwissenschaften: 473-500.
- Bretschneider, Frank (2005b): *Bundestagswahlkampf und Medienberichterstattung*. In: *Aus Politik und Zeitgeschichte* 51/52: 19-26.

- Brettschneider, Frank/van Deth, Jan W./Roller, Edeltraud (Hrsg.) (2004): Die Bundestagswahl 2002. Analysen der Wahlergebnisse und des Wahlkampfes. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Brettschneider, Frank/Niedermayer, Oskar/Weßels, Bernhard (Hrsg.) (2007): Die Bundestagswahl 2005. Analysen des Wahlkampfes und der Wahlergebnisse. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Brody, Richard/Page, Benjamin I. (1975): The Impact of Events on Presidential Popularity: The Johnson and Nixon Administrations. In: Wildavsky, Aaron (Hrsg.): Perspectives on the Presidency. Boston: Little, Brown & Co.: 136-148.
- Brody, Richard/Sigelman, Lee (1983): Presidential Popularity and Presidential Elections: An Update and Extension In: Public Opinion Quarterly 47: 325-328.
- Brüggelambert, Gregor (1997): Von Insidern, „marginal traders“ und Glücksrittern: Zur Relevanz von Entscheidungsanomalien in politischen Börsen. In: Jahrbücher für Nationalökonomie und Statistik 216: 45-73.
- Brüggelambert, Gregor (1999): Institutionen als Informationsträger: Erfahrungen mit Wahlbörsen. Marburg: Metropolis.
- Brüggelambert, Gregor/Crüger, Arwed (2002): Election Markets. Experiences from a Complex Market Experiment. In: Bolle, Friedel/Lehmann-Waffenschmidt, Marco (Hrsg.): Surveys in Experimental Economics. Bargaining, Cooperation and Election Stock Markets. Heidelberg u. a.: Physica: 167-191.
- Buchanan, James M./Tollison, Robert D./Tullock, Gordon (Hrsg.) (1980): Toward a Theory of the Rent-Seeking Society. College Station: Texas A & M University Press.
- Buchanan, William (1986): Election Predictions: An Empirical Assessment In: Public Opinion Quarterly 50: 222-227.
- Budge, Ian/Farlie, Dennis (1977): Voting and Party Competition. A Theoretical Critique and Synthesis Applied to Surveys from 10 Democracies. London u. a.: Wiley.
- Bundesagentur für Arbeit (2009a): Arbeitslose für Deutschland (1950-1990) – Monatszahlen. Nürnberg: Bundesagentur für Arbeit. Online: <http://www.pub.arbeitsamt.de/hst/services/statistik/detail/z.html> (12. März 2009).
- Bundesagentur für Arbeit (2009b): Arbeitslose ab Januar 1991 (Insgesamt und Jugendliche) – Monats- und Jahreszahlen. Nürnberg: Bundesagentur für Arbeit. Online: <http://www.pub.arbeitsamt.de/hst/services/statistik/detail/z.html> (9. Januar 2010).
- Bunge, Mario (1959): Causality. The Place of the Causal Principle in Modern Science. Cambridge: Harvard University Press.
- Bunge, Mario (1983): Epistemologie. Aktuelle Fragen der Wissenschaftsfrage. Mannheim u. a.: Bibliographisches Institut.

- Bunge, Mario (1996): *Finding Philosophy in Social Science*. New Haven u. a.: Yale University Press.
- Bunge, Mario (1997): Mechanism and Explanation. In: *Philosophy of the Social Sciences* 27: 410-465.
- Bunge, Mario (2004): How Does It Work? The Search for Explanatory Mechanisms. In: *Philosophy of the Social Sciences* 34: 182-210.
- Burzan, Nicole (2005): *Soziale Ungleichheit. Eine Einführung in die zentralen Theorien*. 2. Aufl. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Bytzek, Evelyn (2007): Ereignisse und ihre Wirkung auf die Popularität von Regierungen. Von der Schleyer-Entführung zur Elbeflut. Baden-Baden: Nomos.
- Bytzek, Evelyn (2008): Der überraschende Erfolg der FDP bei der Bundestagswahl 2005: Leihstimmen oder Koalitionswahl als Ursache? Jahrestagung des Arbeitskreises Wahl- und Einstellungsforschung der Deutschen Vereinigung für politische Wissenschaft „Information – Wahrnehmung – Emotion“. Duisburg.
- Camerer, Colin F. (1998): Can Asset Markets Be Manipulated? A Field Experiment with Racetrack Betting. In: *Journal of Political Economy* 106: 457-482.
- Campbell, Angus/Gurin, Gerald/Miller, Warren E. (1954): *The Voter Decides*. Evanston: Peterson Row.
- Campbell, Angus/Converse, Philip E./Miller, Warren E./Stokes, Donald E. (1964): *The American Voter. An Abridgement*. New York u. a.: John Wiley & Sons.
- Campbell, James E. (1994): Forecasts of the 1992 Presidential Election. In: *The Political Methodologist* 5: 5-9.
- Campbell, James E. (1996): Polls and Votes: The Trial-Heat Presidential Election Forecasting Model, Certainty, and Political Campaigns. In: *American Politics Quarterly* 24: 408-433.
- Campbell, James E. (2001): The Referendum That Didn't Happen: The Forecasts of the 2000 Presidential Election. In: *PS: Political Science & Politics* 34: 33-38.
- Campbell, James E. (2008a): Evaluating U.S. Presidential Election Forecasts and Forecasting Equations. In: *International Journal of Forecasting* 24: 257-269.
- Campbell, James E. (2008b): The Trial-Heat Forecast of the 2008 Presidential Vote: Performance and Value Considerations in an Open-Seat Election. In: *PS: Political Science & Politics* 41: 697-701.
- Campbell, James E. (2009): The 2008 Campaign and the Forecasts Derailed. In: *PS: Political Science & Politics* 42: 19-20.
- Campbell, James E./Lewis-Beck, Michael S. (2008): US Presidential Election Forecasting: An Introduction. In: *International Journal of Forecasting* 24: 189-192.
- Campbell, James E./Wink, Kenneth A. (1990): Trial-Heat Forecasts of the Presidential Vote. In: *American Politics Quarterly* 18: 251-269.

- Cartwright, Dorwin/Zander, Alvin (1968): *Group Dynamics: Research and Theory*. 3. Aufl. New York: Harper & Row.
- Chappell Jr., Henry W./Veiga, Linda G. (2000): *Economics and Elections in Western Europe: 1960-1997*. In: *Electoral Studies* 19: 183-197.
- Clark, Terry N./Lipset, Seymour M. (Hrsg.) (2001): *The Breakdown of Class Politics: A Debate on Post-Industrial Stratification*. Washington: Woodrow Wilson Center Press.
- Converse, Philip E. (1966): *The Concept of a Normal Vote*. In: Campbell, Angus/Converse, Philip E./Miller, Warren E./Stokes, Donald E. (Hrsg.): *Elections and the Political Order*. New York u. a.: John Wiley & Sons: 9-39.
- Converse, Philip E. (1969): *Of Time and Partisan Stability*. In: *Comparative Political Studies* 2: 139-171.
- Converse, Philip E./Traugott, Michael W. (1986): *Assessing the Accuracy of Polls and Surveys*. In: *Science* 234: 1094-1098.
- Couper, Mick P./De Leeuw, Edith D. (2003): *Nonresponse in Cross-Cultural and Cross-National Surveys*. In: Harkness, Janet/van de Vijver, Fons/Mohler, Peter P. (Hrsg.): *Cross-Cultural Survey Methods*. New York u. a.: Wiley: 157-177.
- Crespi, Irving (1988): *Pre-Election Polling. Sources of Accuracy and Error*. New York: Russell Sage.
- Crespi, Leo P. (1952): *America's Interest in German Survey Research*. In: Institut zur Förderung öffentlicher Angelegenheiten (Hrsg.): *Empirische Sozialforschung. Meinungs- und Marktforschung. Methoden und Probleme*. Frankfurt am Main: Institut zur Förderung öffentlicher Angelegenheiten: 215-217.
- Cuzán, Alfred G./Bundrick, Charles M. (1999): *Fiscal Policy as a Forecasting Factor in Presidential Elections*. In: *American Politics Quarterly* 27: 338-353.
- Cuzán, Alfred G./Bundrick, Charles M. (2008): *Forecasting the 2008 Presidential Election: A Challenge for the Fiscal Model*. In: *PS: Political Science & Politics* 41: 717-722.
- Cuzán, Alfred G./Heggen, Richard J. (1984): *A Fiscal Model of Presidential Elections in the United States, 1880-1980*. In: *Presidential Studies Quarterly* 14: 98-108.
- Cuzán, Alfred G./Armstrong, J. Scott/Jones Jr., Randall J. (2005): *Combining Methods to Forecast the 2004 Presidential Election: The Pollyvote*. Annual Meeting of the Southern Political Science Association. New Orleans: All Academic.
- Czerlinski, Jean/Gigerenzer, Gerd/Goldstein, Daniel G. (1999): *How Good are Simple Heuristics?* In: Gigerenzer, Gerd/Todd, Peter M. (Hrsg.): *Simple Heuristics That Make Us Smart*. New York u. a.: Oxford University Press: 97-118.

- D'Agostino, Ralph B./Belanger, Albert/D'Agostino Jr., Ralph B. (1990): A Suggestion for Using Powerful and Informative Tests of Normality. In: *The American Statistician* 44: 316-321.
- Dach, R. Peter (1997): Veröffentlichung von Wahlprognosen: keine verfassungsrechtlichen Bedenken. In: *Zeitschrift für Parlamentsfragen* 28: 229-235.
- Dahlem, Stefan (2001): *Wahlentscheidung in der Mediengesellschaft. Theoretische und empirische Grundlagen der interdisziplinären Wahlforschung*. Freiburg im Breisgau u. a.: Karl Alber.
- Dalton, Russell J. (2005): *Citizen Politics: Public Opinion and Political Parties in Advanced Industrial Democracies*. Washington: CQ Press.
- Dalton, Russell J./Bürklin, Wilhelm (2003): Wähler als Wandervogel: Dealignment and the German Voter. In: *German Politics and Society* 21: 57-75.
- Davis, Douglas D./Holt, Charles A. (1993): *Experimental Economics*. Princeton: Princeton University Press.
- Debus, Marc (2007): Bestimmungsfaktoren des Wahlverhaltens in Deutschland bei den Bundestagswahlen 1987, 1998 und 2002: Eine Anwendung des Modells von Adams, Merrill und Grofman. In: *Politische Vierteljahresschrift* 48: 269-292.
- Debus, Marc (2008): Parteien- und Kandidatensympathie als Einflussfaktoren auf die individuelle Wahlentscheidung. Jahrestagung des Arbeitskreises Wahl- und Einstellungsforschung der Deutschen Vereinigung für politische Wissenschaft „Information – Wahrnehmung – Emotion“. Unveröffentlichtes Manuskript. Duisburg: Universität Duisburg-Essen.
- Decker, Frank/Zadra, Dirk (2001): Politische Meinungsforschung in der Bundesrepublik Deutschland. Eine Bilanz nach fünfzig Jahren. In: *Zeitschrift für Politikwissenschaft* 11: 31-69.
- Der Bundeswahlleiter (2009a): Endgültiges Ergebnis der Bundestagswahl 2009. Online: http://www.bundeswahlleiter.de/de/bundestagswahlen/BTW_BUND_09/ergebnisse/bundesergebnisse/index.html (9. Januar 2010).
- Der Bundeswahlleiter (2009b): Informationen zur Teilnahme an Bundestagswahlen. Wiesbaden: Statistisches Bundesamt. Online: <http://www.bundeswahlleiter.de/de/parteien/wahlteilnahme/bundestagswahl.html> (25. März 2009).
- DeSart, Jay/Holbrook, Thomas (2003): Campaigns, Polls, and the States: Assessing the Accuracy of Statewide Presidential Trial-Heat Polls. In: *Political Research Quarterly* 56: 431-439.
- Deutsche Bischofskonferenz (2009a): *Katholiken und Gottesdienstteilnehmer, 1950-2008*. Bonn: Deutsche Bischofskonferenz. Online: http://www.dbk.de/imperia/md/content/kirchlichestatistik/katholiken_und_gottesdienstteilnehmer-1950-2008.pdf (9. Januar 2010).

- Deutsche Bischofskonferenz (2009b): Bevölkerung und Katholiken in der Bundesrepublik Deutschland, 1965-2008. Bonn: Deutsche Bischofskonferenz. Online: http://www.dbk.de/imperia/md/content/kirchlichestatistik/bevoelkerung_und_katholiken_1965-2008.pdf (9. Januar 2010).
- Deutscher Gewerkschaftsbund (DGB) (2009): Mitglieder in den DGB-Gewerkschaften. Berlin: Deutscher Gewerkschaftsbund. Online: <http://www.dgb.de/dgb/mitgliederzahlen/mitglieder.htm> (9. Januar 2010).
- Deutsches Historisches Museum (2009): 2005: Chronik. Berlin: Deutsches Historisches Museum. Online: <http://www.dhm.de/lemo/html/2005/index.html> (13. März 2009).
- Dickey, David A./Fuller, Wayne A. (1979): Distribution of the Estimators for Autoregressive Time Series with a Unit Root. In: *Journal of the American Statistical Association* 74: 427-431.
- Diekmann, Andreas (2008): Soziologie und Ökonomie: Der Beitrag experimenteller Wirtschaftsforschung zur Sozialtheorie. In: *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 60: 528-550.
- Dodd, Stuart C. (1946): Toward World Surveying. In: *Public Opinion Quarterly* 10: 470-483.
- Donovitz, Frank (1999): Journalismus und Demoskopie. Wahlumfragen in den Medien. Berlin: Vistas.
- Donsbach, Wolfgang (1984): Die Rolle der Demoskopie in der Wahlkampf-Kommunikation. Empirische und normative Aspekte der Hypothese über den Einfluß der Meinungsforschung auf die Wählermeinung. In: *Zeitschrift für Politik* 31: 388-407.
- Dorussen, Han/Taylor, Michael (Hrsg.) (2002): *Economic Voting*. London: Routledge.
- Downs, Anthony (1968): *Ökonomische Theorie der Demokratie*. Tübingen: Mohr.
- Dubois, Eric/Fauvelle-Aymar, Christine (2004): Vote Functions in France and the 2002 Election Forecasts. In: Lewis-Beck, Michael S. (Hrsg.): *The French Voter. Before and after the 2002 Elections*. New York: Palgrave macmillan: 205-230.
- Eberlein, Klaus D. (2001): *Möglichkeiten und Grenzen der Meinungsforschung. Kritische Betrachtungen zu Geschichte, Methoden und Interpretationsmöglichkeiten*. Berlin: Frieling.
- Ehrenfordt, Robert/Maaßen, Stefan (2007): *Systematische Fehler von Wahlprognosen*. München: GRIN. Online: <http://www.grin.com/e-book/84278> (23. Oktober 2008).
- Eith, Ulrich (2008): Wahlforschung, Demoskopie und praktische Politik – Aspekte unterschiedlicher Zugänge und Konsequenzen in vier Thesen. In: *Politische Studien* 57: 18-24.

- Elff, Martin (2006): Politische Ideologien, soziale Konflikte und Wahlverhalten. Die Bedeutung politischer Angebote der Parteien für den Zusammenhang zwischen sozialen Merkmalen und Parteipräferenzen in zehn westeuropäischen Demokratien. Baden-Baden: Nomos.
- Elff, Martin (2009): Social Divisions, Party Positions, and Electoral Behaviour. In: *Electoral Studies* 28: 297-308.
- Elster, Jon (1998): A Plea for Mechanisms. In: Hedström, Peter/Swedberg, Richard (Hrsg.): *Social Mechanisms: An Analytical Approach to Social Theory*. Cambridge u. a.: Cambridge University Press: 45-73.
- Elster, Jon (2007): *Explaining Social Behavior. More Nuts and Bolts for the Social Sciences*. Cambridge u. a.: Cambridge University Press.
- Emmert, Thomas/Roth, Dieter (1995): Zur wahlsoziologischen Bedeutung eines Modells sozialstrukturell verankerter Konfliktlinien im vereinten Deutschland. In: *Historical Social Research* 20: 119-160.
- Enelow, James M./Hinich, Melvin J. (1984): *The Spatial Theory of Voting. An Introduction*. Cambridge u. a.: Cambridge University Press.
- Enelow, James M./Hinich, Melvin J. (Hrsg.) (1990): *Advances in the Spatial Theory of Voting*. Cambridge u. a.: Cambridge University Press.
- Erikson, Robert S./Goldthorpe, John H. (1992): *The Constant Flux. A Study of Class Mobility in Industrial Societies*. Oxford: Clarendon.
- Erikson, Robert S./Sigelman, Lee (1995): The Polls-Review: Poll-Based Forecasts of Midterm Congressional Election Outcomes: Do the Pollsters Get It Right? In: *Public Opinion Quarterly* 59: 589-605.
- Erikson, Robert S./Wlezien, Christopher (2008): Are Political Markets Really Superior to Polls as Election Predictors? In: *Public Opinion Quarterly* 72: 190-215.
- Esser, Hartmut (1975): *Soziale Regelmäßigkeiten des Befragtenverhaltens*. Meisenheim am Glan: Anton Hain.
- Esser, Hartmut (1986): Können Befragte lügen? Zum Konzept des „wahren Wertes“ im Rahmen der handlungstheoretischen Erklärung von Situationsinflüssen bei der Befragung. In: *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 38: 314-336.
- Esser, Hartmut (1990): „Habits“, „Frames“ und „Rational Choice“: Die Reichweite von Theorien der rationale Wahl (am Beispiel der Erklärung des Befragtenverhaltens). In: *Zeitschrift für Soziologie* 19: 231-247.
- Evans, Geoffrey (Hrsg.) (1999): *The End of Class Politics? Class Voting in Comparative Context*. Oxford u. a.: Oxford University Press.
- Evans, Geoffrey (2000): The Continued Significance of Class Voting. In: *Annual Review of Political Science* 3: 401-417.

- Faas, Thorsten/Schmitt-Beck, Rüdiger (2007): Wahrnehmung und Wirkungen politischer Meinungsumfragen. Eine Exploration zur Bundestagswahl 2005. In: Brettschneider, Frank/Niedermayer, Oskar/Weßels, Bernhard (Hrsg.): Die Bundestagswahl 2005. Analysen des Wahlkampfes und der Wahlergebnisse. Wiesbaden: VS Verlag für Sozialwissenschaften: 233-267.
- Fair, Ray C. (1978): The Effect of Economic Events on Votes for President. In: The Review of Economics and Statistics 60: 159-173.
- Fair, Ray C. (1982): The Effect of Economic Events on Votes for President: 1980 Results. In: Review of Economic and Statistics 60: 322-325.
- Fair, Ray C. (1988): The Effect of Economic Events on Votes for President: 1984 Update. In: Political Behavior 10: 168-179.
- Fair, Ray C. (1996): The Effect of Economic Events on Votes for President: 1992 Update. In: Political Behavior 18: 119-139.
- Fair, Ray C. (2002): Predicting Presidential Elections and Other Things. Stanford: Stanford University Press.
- Falter, Jürgen W. (1977): Einmal mehr: Läßt sich das Konzept der Parteiidentifikation auf deutsche Verhältnisse übertragen? In: Politische Vierteljahresschrift 18: 476-500.
- Falter, Jürgen W. (1984): Zur Übertragbarkeit des Konzepts der Parteiidentifikation auf deutsche Verhältnisse. Einige empirische Ergebnisse. In: Holler, Manfred J. (Hrsg.): Wahlanalyse: Hypothesen, Methoden und Ergebnisse. München: tuduv: 13-34.
- Falter, Jürgen W./Rattinger, Hans (1983): Parteien, Kandidaten und politische Streitfragen bei der Bundestagswahl 1980: Möglichkeiten und Grenzen der Normal-Vote-Analyse. In: Kaase, Max/Klingemann, Hans-Dieter (Hrsg.): Wahlen und politisches System. Analysen aus Anlaß der Bundestagswahl 1980. Opladen: Westdeutscher Verlag: 320-421.
- Falter, Jürgen W./Rattinger, Hans (1986): Die Bundestagswahl 1983: Eine Normalwahlanalyse. In: Klingemann, Hans-Dieter/Kaase, Max (Hrsg.): Wahlen und politischer Prozeß. Analysen aus Anlaß der Bundestagswahl 1983. Opladen: Westdeutscher Verlag: 289-337.
- Falter, Jürgen W./Rattinger, Hans (1992): Taugt sie wirklich so wenig, die deutsche Wahlforschung? Eine Replik auf Rainer Olaf-Schultze. In: Zeitschrift für Parlamentsfragen 23: 354-360.
- Falter, Jürgen W./Schoen, Harald (Hrsg.) (2005): Handbuch Wahlforschung. Wiesbaden: VS Verlag für Sozialwissenschaften.

- Falter, Jürgen W./Schoen, Harald/Caballero, Claudio (2000): Dreißig Jahre danach: Zur Validierung des Konzepts ‚Parteidentifikation‘ in der Bundesrepublik. In: Klein, Markus/Jagodzinski, Wolfgang/Mochmann, Ekkehard (Hrsg.): 50 Jahre empirische Wahlforschung in Deutschland. Entwicklung, Befunde, Perspektiven, Daten. Wiesbaden: Westdeutscher Verlag: 235-271.
- Falter, Jürgen W./Gabriel, Oscar W./Weßels, Bernhard (Hrsg.) (2005): Wahlen und Wähler. Analysen aus Anlass der Bundestagswahl 2002. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Faz.Net (2005): Gallup-Institut kritisiert deutsche Wahlforscher. Frankfurt am Main: F.A.Z Electronic Media. Online: <http://www.faz.net/s/RubFC06D389EE76479E9E76425072B196C3/Doc~EDC5D3E7E46454D15BE00F46E95E25A9F~ATpl~Ecommon~Scontent.html> (23. Oktober 2008).
- Fehr, Ernst/Gintis, Herbert (2007): Human Motivation and Social Cooperation: Experimental and Analytical Foundations. In: Annual Review of Sociology 33: 43-64.
- Feld, Lars P./Kirchgässner, Gebhard (2000): Official and Hidden Unemployment and the Popularity of the Government: An Econometric Analysis for the Kohl Government. In: Electoral Studies 19: 333-347.
- Feldkamp, Michael F. (2005): Datenhandbuch zur Geschichte des Deutschen Bundestages 1994 bis 2003. Baden-Baden: Nomos.
- Fessel, Walter (1969): Zur Problematik von Wahlprognosen. In: Angewandte Sozialforschung: 173-184.
- Finkel, Steven E./Schrott, Peter R. (1995): Campaign Effects on Voter Choice in the German Election of 1990. In: British Journal of Political Science 25: 349-377.
- Fiorina, Morris P. (1981): Retrospective Voting in American National Elections. New Haven u. a.: Yale University Press.
- Flores, Benito E./Wichern, Dean W. (2005): Evaluating Forecasts: A Look at Aggregate Bias and Accuracy Measures. In: Journal of Forecasting 24: 433-451.
- Forsa (o.J.): Forsa-Bus 2005. ZA-Archiv-Nummer 4343. Köln: Zentralarchiv für empirische Sozialforschung an der Universität zu Köln.
- Forschungsgruppe Wahlen (2008a): Projektion. Online: <http://www.forschungsgruppe.de/Fachbegriffe/Erklaerungen/#Projektion> (21. Februar 2008).
- Forschungsgruppe Wahlen (2008b): Wie alles begann. Online: http://www.forschungsgruppe.de/Unternehmen/Forschungsgruppe_Wahlen_e.V./Wie_aller_begann/ (21. Februar 2008).
- Forschungsgruppe Wahlen (2009a): Politbarometer Juli I 2009. Online: http://www.forschungsgruppe.de/Umfragen_und_Publikationen/Politbarometer/Archiv/Politbarometer_2009/Juli_I/ (3. Januar 2010).

- Forschungsgruppe Wahlen (2009b): Politbarometer Juli I 2009. Mannheim: Forschungsgruppe Wahlen. Online: http://www.forschungsgruppe.de/Umfragen_und_Publicationen/Politbarometer/Archiv/Politbarometer_2009/Juli_I/ (9. Januar 2010).
- Forschungsgruppe Wahlen (o.J.): Politbarometer 1977-2003 (Partielle Kumulation). Maschinenlesbares Codebuch – ZA Nr. 2391. Köln: Zentralarchiv für empirische Sozialforschung an der Universität zu Köln.
- Forsythe, Robert/Nelson, Forrest/Neumann, George R./Wright, Jack (1992): Anatomy of an Experimental Political Stock Market. In: *American Economic Review* 82: 1142-1161.
- Forsythe, Robert/Nelson, Forrest/Neumann, George R./Wright, Jack (1994): The 1992 Iowa Political Stock Markets: September Forecasts. In: *The Political Methodologist* 5: 15-19.
- Forsythe, Robert/Murray, Frank/Ross, Thomas W. (1998): Markets as Predictors of Election Outcomes: Campaign Events and Judgement Bias in the 1993 UBC Election Stock Market. In: *Canadian Public Policy* 24: 329-351.
- Forsythe, Robert/Rietz, Thomas A./Ross, Thomas W. (1999): Wishes, Expectations and Actions: A Survey on Price Formation in Election Stock Markets. In: *Journal of Economic Behavior and Organization* 39: 83-110.
- Fowler, James H./Baker, Laura A./Dawes, Christopher T. (2008): Genetic Variation in Political Participation. In: *American Political Science Review* 102: 233-248.
- Franklin, Mark/Mackie, Tom/Valen, Henry (Hrsg.) (1992): *Electoral Change. Responses to Evolving Social and Attitudinal Structures in Western Countries*. Cambridge u. a.: Cambridge University Press.
- Fröchling, Henrike (1998): *Demokratie und Wahlen in westlichen Demokratien. Eine vergleichende Rational-Choice-Analyse*. Opladen: Westdeutscher Verlag.
- Fuchs, Dieter/Pfetsch, Barbara (1996): Die Beobachtung der öffentlichen Meinung durch das Regierungssystem. In: van den Daele, Wolfgang/Neidhardt, Friedhelm (Hrsg.): *Kommunikation und Entscheidung. Politische Funktionen öffentlicher Meinungsbildung und diskursiver Verfahren*. Berlin: edition sigma.
- Furmaniak, Karl/Hoschka, Peter/Schunck, Hermann (1975): *Wahlforschung, Demoskopie und Politikberatung: Erwartungen, Möglichkeiten, Mißverständnisse – ein Erfahrungsbericht*. In: *Zeitschrift für Parlamentsfragen* 6: 566-579.
- Gabaix, Xavier (1999a): Zipf's Law for Cities: An Explanation. In: *Quarterly Journal of Economics* 114: 739-767.

- Gabaix, Xavier (1999b): Zipf's Law and the Growth of Cities. In: *American Economic Review* 89: 129-132.
- Gabriel, Oscar W. (1997): Parteiidentifikation, Kandidaten und politische Sachfragen als Bestimmungsfaktoren des Parteienwettbewerbs. In: Gabriel, Oscar W./Niedermayer, Oskar/Stöss, Richard (Hrsg.): *Parteiendemokratie in Deutschland*. Opladen: Westdeutscher Verlag: 233-254.
- Gabriel, Oscar W. (2001): Parteiidentifikation, Kandidaten und politische Sachfragen als Bestimmungsfaktoren des Parteienwettbewerbs. In: Gabriel, Oscar W./Niedermayer, Oskar/Stöss, Richard (Hrsg.): *Parteiendemokratie in Deutschland*. 2. Aufl. Bonn: Bundeszentrale für politische Bildung: 228-249.
- Gabriel, Oscar W./Keil, Silke I. (2008): Theorien des Wählerverhaltens. *Politikon*. Politikwissenschaft online. Osnabrück: Deutsche Vereinigung für Politische Wissenschaft. Online: http://ilias.politikon.org:81/goto.php?target=pg_17959_242&client_id=politikon (2. März 2009).
- Gabriel, Oscar W./Weßels, Bernhard/Falter, Jürgen W. (Hrsg.) (2009): *Wahlen und Wähler. Analysen aus Anlass der Bundestagswahl 2005*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Gallup, George H. (1965): Polls and the Political Process – Past, Present, and Future. In: *Public Opinion Quarterly* 29: 544-549.
- Gallup, George H./Rae, Saul F. (1940): Is There a Bandwagon Vote? In: *Public Opinion Quarterly* 4: 244-249.
- Gallus, Alexander (2002): Demoskopie in Zeiten des Wahlkampfes. „Wirkliche Macht“ oder „Faktor ohne politische Bedeutung“. In: *Aus Politik und Zeitgeschichte* 15/16: 29-36.
- Gallus, Alexander (2003): Wahl als „Demoskopiedemokratie“? Überlegungen zur Meinungsforschung und ihren Wirkungen aus Anlass der Bundestagswahl 2002. In: Jesse, Eckhard (Hrsg.): *Bilanz der Bundestagswahl 2002. Voraussetzungen, Ergebnisse, Folgen*. München: Bayerische Landeszentrale für politische Bildungsarbeit: 124-137.
- Gallus, Alexander/Lühe, Marion (1998): *Öffentliche Meinung und Demoskopie*. Berlin: Landeszentrale für politische Bildungsarbeit.
- Georg, Werner (1998): *Soziale Lage und Lebensstil. Eine Typologie*. Opladen: Leske + Budrich.
- Gibowski, Wolfgang G. (1973): Der Effekt unterschiedlicher Platzierung der Wahlabsichtsfrage im Fragebogen. In: *Politische Vierteljahresschrift* 14: 275-293.
- Gibowski, Wolfgang G. (2000): Wie gut sind Wahlprognosen? Eine Betrachtung am Beispiel der Bundestagswahl 1998. In: Niedermayer, Oskar/Westle, Bettina (Hrsg.): *Demokratie und Partizipation: Festschrift für Max Kaase*. Opladen: Westdeutscher Verlag: 417-422.

- Gigerenzer, Gerd (2007): *Bauchentscheidungen. Die Intelligenz des Unbewussten und die Macht der Intuition*. 2. Aufl. München: C. Bertelsmann.
- Gilman, John J. (1992): Broken Sticks: Why Mergers May Fail to Garner Market Shares. In: *Managerial and Decision Economics* 13: 453-456.
- Gisart, Brigitte (2009): Endgültiges Ergebnis der Wahl zum 17. Deutschen Bundestag am 27. September 2009. Wiesbaden: Statistisches Bundesamt. Online: http://www.bundeswahlleiter.de/de/bundestagswahlen/BTW_BUND_09/veroeffentlichungen/BTW2009_WiSta_2009_11.pdf (9. Januar 2010).
- Gluchowski, Peter (1987): Lebensstile und Wandel der Wählerschaft in der Bundesrepublik Deutschland. In: *Aus Politik und Zeitgeschichte* B12: 18-32.
- Goldberg, Samuel (1968): *Differenzgleichungen und ihre Anwendung in Wirtschaftswissenschaft, Psychologie und Soziologie*. München u. a.: Oldenbourg.
- Goldstein, Daniel G./Gigerenzer, Gerd (2008): The Recognition Heuristic and the Less-Is-More Effect. In: Plott, Charles R./Smith, Vernon L. (Hrsg.): *Handbook of Experimental Economics Results, Volume 1*. Amsterdam: North-Holland: 987-992.
- Graefe, Andreas/Armstrong, J. Scott (2008): Forecasting Elections from Voters' Perceptions of Candidates' Ability to Handle Issues. Philadelphia: Wharton School. Online: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1206683 (3. Januar 2010).
- Graefe, Andreas/Cuzán, Alfred G./Jones Jr., Randall J./Armstrong, J. Scott (2009): Combining Forecasts for U.S. Presidential Elections: The PollyVote. Philadelphia: Wharton School. Online: <http://marketing.wharton.upenn.edu/documents/research/Pollyvote.pdf> (3. Januar 2010).
- Green, Donald/Palmquist, Bradley/Schickler, Eric (1998): Macropartisanship: A Replication and Replique. In: *American Political Science Review* 92: 883-899.
- Greene, Steven H. (1999): Understanding Party Identification: A Social Identity Approach. In: *Political Psychology* 20: 393-403.
- Groß, Jochen (2003): *Der Einfluss von Interessengruppen auf die Wohlfahrt in Staaten der EU. Unveröffentlichte Magisterarbeit*. München: Ludwig-Maximilians-Universität München.
- Gschwend, Thomas (2003): Ticket-Splitting and Strategic Voting in Mixed Electoral Systems. MZES Working Paper. Mannheim: Mannheimer Zentrum für europäische Sozialforschung. Online: <http://www.mzes.uni-mannheim.de/publications/wp/wp-61.pdf> (22. Mai 2009).
- Gschwend, Thomas (2007): Ticket-Splitting and Strategic Voting under Mixed Electoral Rules: Evidence from Germany. In: *European Journal of Political Research* 46: 1-23.

- Gschwend, Thomas (2009): Das Kanzlermodell bei der Wahl 2009: Diesmal kein Volltreffer. Hamburg: Die ZEIT. Online: http://blog.zeit.de/politik-nach-zahlen/2009/10/14/das-kanzlermodell-bei-der-wahl-2009-diesmal-kein-volltreffer_2113 (10. Januar 2010).
- Gschwend, Thomas/Hooghe, Marc (2008): Should I Stay or Should I Go? An Experimental Study on Voter Responses to Pre-Electoral Coalitions. In: *European Journal of Political Research* 47: 556-577.
- Gschwend, Thomas/Norpoth, Helmut (2001): „Wenn am nächsten Sonntag ...“: Ein Prognosemodell für Bundestagswahlen. In: Klingemann, Hans-Dieter/Kaase, Max (Hrsg.): *Wahlen und Wähler. Analysen aus Anlass der Bundestagswahl 1998*. Wiesbaden: Westdeutscher Verlag: 473-499.
- Gschwend, Thomas/Norpoth, Helmut (2003): Knapp aber vorhersehbar: Schröders Wahlsieg. Neues Prognosemodell für Bundestagswahlen. In: *Forum – Forschung*: 10-13.
- Gschwend, Thomas/Norpoth, Helmut (2005): Prognosemodell auf dem Prüfstand. Die Bundestagswahl 2005. In: *Politische Vierteljahresschrift* 46: 682-688.
- Gschwend, Thomas/Pappi, Franz U. (2004): Stimmensplitting und Koalitionswahl. In: Brettschneider, Frank/van Deth, Jan W./Roller, Edeltraud (Hrsg.): *Die Bundestagswahl 2002. Analysen der Wahlergebnisse und des Wahlkampfes*. Wiesbaden: VS Verlag für Sozialwissenschaften:
- Häder, Sabine/Glemser, Axel (2006): Stichprobenziehung für Telefonumfragen in Deutschland. In: Diekmann, Andreas (Hrsg.): *Stichprobenziehung für Telefonumfragen in Deutschland*. Wiesbaden: VS Verlag für Sozialwissenschaften: 148-171.
- Hansen, Jan/Schmidt, Carsten/Strobel, Martin (2004): Manipulation in Political Stock Markets – Preconditions and Evidence. In: *Applied Economics Letters* 11: 459-463.
- Hanson, Robin/Oprea, Ryan (2007): A Manipulator Can Aid Prediction Market Accuracy. George Mason University. Online: <http://hanson.gmu.edu/bias-help.pdf> (5. Juni 2008).
- Hartenstein, Wolfgang (1967): Mit Pognosen leben: Der Einfluß von Wahlvoraussagen auf das Wahlverhalten. In: Böhret, Carl/Grosser, Dieter (Hrsg.): *Interdependenzen von Politik und Wirtschaft. Festschrift für Gert von Eynern*. Berlin: Duncker & Humblot: 285-306.
- Hartenstein, Wolfgang (1969): Gesetzliches Verbot für Wahlprognosen? In: *Zeitschrift für Rechtspolitik*: 201-202.
- Hartleb, Florian/Jesse, Eckhard (2005): Direkte Demokratie in der Bundesrepublik Deutschland: Positionen und Kontroversen. In: Massing, Peter (Hrsg.): *Direkte Demokratie. Eine Einführung*. Schwalbach: Wochenschau: 13-38.

- Hayek, Friedrich A. (1945): The Use of Knowledge in Society. In: American Economic Review 35: 519-530.
- Hedström, Peter (2005): Dissecting the Social. On the Principles of Analytical Sociology. Cambridge u. a.: Cambridge University Press.
- Hedström, Peter/Swedberg, Richard (1998): Social Mechanisms: An Introductory Essay. In: Hedström, Peter/Swedberg, Richard (Hrsg.): Social Mechanisms: An Introductory Essay. Cambridge u. a.: Cambridge University Press: 1-31.
- Hempel, Carl G./Oppenheim, Paul (1948): Studies in the Logic of Explanation. In: Philosophy of Science 15: 135-175.
- Hempel, Carl G. (1965): Aspects of Scientific Explanation and other Essays in the Philosophy of Science. New York u. a.: The Free Press.
- Hempel, Carl G. (1977): Aspekte wissenschaftlicher Erklärung. Berlin u. a.: Walter de Gruyter.
- Hennis, Wilhelm (1957): Meinungsforschung und repräsentative Demokratie. Zur Kritik politischer Umfragen. Tübingen: J. C. B. Mohr (Paul Siebeck).
- Hennis, Wilhelm (1983): Podiumsdiskussion. In: Kaase, Max/Ott, Werner/Scheuch, Erwin K. (Hrsg.): Podiumsdiskussion. Frankfurt am Main u. a.: Campus: 77-80.
- Herfindahl, Orris C. (1950): Concentration in the Steel Industry. Ph.D. Thesis. New York: Columbia University.
- Hibbs, Douglas (2008): Implications of the ‚Bread and Peace‘ Model for the 2008 US Presidential Election. In: Public Choice 137: 1-10.
- Hildebrandt, Kai/Dalton, Russell J. (1977): Die neue Politik. In: Politische Vierteljahresschrift 18: 230-256.
- Hill, Bruce M. (1974): The Rank-Frequency Form of Zipf's Law. In: Journal of the American Statistical Association 69: 1017-1026.
- Hilmer, Richard (2008): Exit Polls – A Lot More than Just a Tool for Election Forecasts. In: Carballo, Marita/Hjelmar, Ulf (Hrsg.): Public Opinion Polling in a Globalized World. Berlin u. a.: Springer: 93-108.
- Hilmer, Richard/Hofrichter, Jürgen (2001): Wahltagsbefragungen in den neunziger Jahren: Überblick und Bilanz. In: Klingemann, Hans-Dieter/Kaase, Max (Hrsg.): Wahlen und Wähler. Analysen aus Anlass der Bundestagswahl 1998. Wiesbaden: Westdeutscher Verlag: 443-472.
- Hilmer, Richard/Hofrichter, Jürgen (2002): Wahlforschung und Wahlberichterstattung: Wie Meinungsforscher Volkes Stimme erfassen. In: planung & analyse 2002: 32-62.

- Hilmer, Richard/Schleyer, Nicolas (2000): Stimmensplitting bei der Bundestagswahl 1998. Strukturen, Trends und Motive. In: van Deth, Jan W./Rattinger, Hans/Roller, Edeltraud (Hrsg.): Die Republik auf dem Weg zur Normalität? Wahlverhalten und politische Einstellungen nach acht Jahren Einheit. Opladen: Leske + Budrich: 173-197.
- Hinich, Melvin J./Munger, Michael C. (1997): *Analytical Politics*. Cambridge u. a.: Cambridge University Press.
- Hinz, Thomas (2005): Mehrebenenanalyse. In: Kühl, Stefan/Strodtholz, Petra/Taffertshofer, Andreas (Hrsg.): *Quantitative Methoden der Organisationsforschung*. Ein Handbuch. Wiesbaden: VS Verlag für Sozialwissenschaften: 361-385.
- Hoffmeyer-Zlotnik, Jürgen H. P. (1997): Random-Route-Stichproben nach ADM. In: Gabler, Siegfried/Hoffmeyer-Zlotnik, Jürgen H. P. (Hrsg.): *Stichproben in der Umfragepraxis*. Opladen: Westdeutscher Verlag: 33-42.
- Hofmann-Göttig, Joachim (2005): Meinungsforschung und Meinungsmache am Beispiel des Bundestagswahlkampfes 2005. Was taugt die „Sonntagsfrage“? Koblenz-Landau: Universität Koblenz-Landau. Online: http://www.mb-wjk.rlp.de/fileadmin/Dateien/Downloads/Ministerium/hofmann-goettig/Meinungsforschung_Meinungsmache.pdf (10. November 2008).
- Hohlfeld, Ralf (2006): Bundestagswahlkampf 2005 in den Hauptnachrichtensendungen. In: *Aus Politik und Zeitgeschichte* 38: 11-17.
- Holbrook, Thomas M. (1996): Reading the Political Tea Leaves. A Forecasting Model of Contemporary Presidential Elections. In: *American Politics Quarterly* 24: 506-519.
- Holbrook, Thomas M. (2007): Forecasting with Mixed Economic Signals: A Cautionary Tale. In: *PS: Political Science & Politics* 34: 39-44.
- Holbrook, Thomas M. (2008): Incumbency, National Conditions, and the 2008 Presidential Election. In: *PS: Political Science & Politics* 41: 709-712.
- Holbrook, Thomas M./DeSart, Jay A. (1999): Using State Polls to Forecast Presidential Election Outcomes in the American States. In: *International Journal of Forecasting* 15: 137-142.
- Horst, Paul/Wallin, Frieda B./Clausen, John A./Reed, Robert/Rosenthal, Erich/Wallin, Paul/Guttman, Louis (1941): *The Prediction of Personal Adjustment: A Survey of Logical Problems and Research Techniques, with Illustrative Application to Problems of Vocational Selection, School Success, Marriage, and Crime*. New York: Social Science Research Council.
- Hox, Joop J. (2002): *Multilevel Analysis. Techniques and Applications*. Mahwah u. a.: Lawrence Erlbaum.
- Huber, Jürgen (2002): *Wahlbörsen: Preisbildung auf politischen Märkten zur Vorhersage von Wahlergebnissen*. Hamburg: Dr. Kovac.

- Hyndman, Rob J./Koehler, Anne B. (2006): Another Look at Measures of Forecast Accuracy. In: *International Journal of Forecasting* 22: 679-688.
- Infratest dimap (2008): Was ist die Sonntagsfrage? Berlin: Infratest dimap. Online: <http://www.infratest-dimap.de/?id=45#faq3> (21. Februar 2008).
- Inglehart, Ronald (1971): The Silent Revolution in Europe: Intergenerational Change in Post-Industrial Societies. In: *American Political Science Review* 65: 991-1017.
- Inglehart, Ronald (1977): *Silent Revolution: Changing Values and Political Styles Among Western Publics*. Princeton: Princeton University Press.
- Inglehart, Ronald (1983): Traditionelle politische Trennungslinien und die Entwicklung der neuen Politik in westlichen Gesellschaften. In: *Politische Vierteljahresschrift* 24: 139-165.
- Institut zur Förderung öffentlicher Angelegenheiten (Hrsg.) (1952): *Empirische Sozialforschung. Meinungs- und Marktforschung. Methoden und Probleme*. Frankfurt am Main: Institut zur Förderung öffentlicher Angelegenheiten.
- Jagodzinski, Wolfgang/Kühnel, Steffen (1990): Zur Schätzung der relativen Effekte von Issueorientierungen, Kandidatenpräferenz und langfristiger Parteibindung auf die Wahlabsicht. In: Schmitt, Karl (Hrsg.): *Wahlen, Parteiliten, politische Einstellungen. Neuere Forschungsergebnisse*. Frankfurt am Main: Peter Lang: 5-63.
- Jagodzinski, Wolfgang/Quandt, Markus (1997): Wahlverhalten und Religion im Lichte der Individualisierungsthese. Anmerkungen zu dem Beitrag von Schnell und Kohler. In: *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 49: 761-782.
- Jérôme, Bruno/Jérôme-Speziari, Veronique (2004): Forecasting the 2002 Elections: Lessons from a Political Economy Model. In: Lewis-Beck, Michael S. (Hrsg.): *The French Voter. Before and After the 2002 Elections*. New York: Palgrave macmillan: 178-204.
- Jesse, Eckhard (1988): Split-Voting in the Federal Republic of Germany: An Analysis of the Federal Elections from 1953 to 1987. In: *Electoral Studies* 7: 109-124.
- Jesse, Eckhard (Hrsg.) (2003): *Bilanz der Bundestagswahl 2002. Voraussetzungen, Ergebnisse, Folgen*. München: Bayerische Landeszentrale für politische Bildungsarbeit.
- Jesse, Eckhard (2008): *Demokratie in Deutschland. Diagnosen und Analysen*. Köln u. a.: Böhlau.

- Jones Jr., Randall J./Armstrong, J. Scott/Cuzán, Alfred G. (2007): Forecasting Elections Using Expert Surveys: An Application to U.S. Presidential Elections. MPRA Paper. München: Universitätsbibliothek der Ludwig-Maximilians-Universität München. Online: <http://mpra.ub.uni-muenchen.de/5301> (3. Oktober 2008).
- Juhász, Zoltán (1993): Wahlabsicht und Rückerinnerung: Zwei Angaben zur aktuellen Bewertung der politischen Parteien? In: Gabriel, Oscar W./Troitzsch, Klaus G. (Hrsg.): Wahlen in Zeiten des Umbruchs. Frankfurt am Main: Peter Lang: 27-50.
- Kaase, Max (1973): Die Bundestagswahl 1972: Probleme und Analysen. In: Politische Vierteljahresschrift 14: 145-190.
- Kaase, Max (Hrsg.) (1977): Wahlsoziologie heute. Analysen aus Anlaß der Bundestagswahl 1976. Opladen: Westdeutscher Verlag.
- Kaase, Max (1999a): Wahlforschung und Demokratie. In: ZUMA Nachrichten 1999: 62-82.
- Kaase, Max (Hrsg.) (1999b): Qualitätskriterien der Umfrageforschung. Denkschrift. Berlin: Akademie-Verlag.
- Kaase, Max (2000): Entwicklung und Stand der Empirischen Wahlforschung in Deutschland. In: Klein, Markus/Jagodzinski, Wolfgang/Mochmann, Ekkehard/Ohr, Dieter (Hrsg.): 50 Jahre Empirische Wahlforschung in Deutschland. Entwicklung, Befunde, Perspektiven, Daten. Wiesbaden: Westdeutscher Verlag: 17-40.
- Kaase, Max/Klingemann, Hans-Dieter (Hrsg.) (1983): Wahlen und politisches System. Analysen aus Anlaß der Bundestagswahl 1980. Opladen: Westdeutscher Verlag.
- Kaase, Max/Ott, Werner/Scheuch, Erwin K. (Hrsg.) (1983): Empirische Sozialforschung in der modernen Gesellschaft. Beiträge und Referate anlässlich und in Zusammenhang mit der gemeinsamen wissenschaftlichen Jahrestagung des Arbeitskreises Deutscher Marktforschungsinstitute (ADM) und der Arbeitsgemeinschaft Sozialwissenschaftlicher Institute (ASI) am 1. und 2. Oktober 1981 in Heidelberg. Frankfurt am Main u. a.: Campus.
- Kagel, John H./Roth, Alvin (Hrsg.) (1995): The Handbook of Experimental Economics. Princeton: Princeton University Press.
- Kaplitza, Gabriele (1982): Die Stichprobe. In: Holm, Kurt (Hrsg.): Die Befragung 1. Der Fragebogen – die Stichprobe. München: A. Francke: 136-177.
- Kastellec, Jonathan P./Gelman, Andrew/Chandler, Jamie P. (2008): The Playing Field Shifts: Predicting the Seats-Votes Curve in the 2008 U.S. House Elections. In: PS: Political Science & Politics 41: 729-732.
- Katz, Daniel (1949): An Analysis of the 1948 Polling Predictions. In: Journal of Applied Psychology 33: 15-28.

- Keller, Berndt (1990): Interessenorganisation und Interessenvermittlung. Die Grenzen eines neoklassischen Institutionalismus in Olsons „Rise and Decline of Nations“. In: Kölner Zeitschrift für Soziologie und Sozialpsychologie 42: 502-524.
- Keller, Felix (2001): Archäologie der Meinungsforschung. Mathematik und die Erzählbarkeit des Politischen. Konstanz: UVK.
- Kellermann, Charlotte (2007): „Trends and Constellations“: Klassische Bestimmungsfaktoren des Wahlverhaltens bei den Bundestagswahlen 1990-2005. In: Rattinger, Hans/Gabriel, Oscar W./Falter, Jürgen W. (Hrsg.): Der gesamtdeutsche Wähler. Stabilität und Wandel des Wählerverhaltens im wiedervereinigten Deutschland. Baden-Baden: Nomos: 297-327.
- Kellermann, Charlotte (2008): Trends and Constellations. Klassische Bestimmungsfaktoren des Wahlverhaltens bei den Bundestagswahlen 1990-2005. Baden-Baden: Nomos.
- Kellermann, Charlotte/Rattinger, Hans (2005): „Round up the usual suspects“: Die Bedeutung klassischer Bestimmungsfaktoren der Wahlentscheidung bei den Bundestagswahlen 1994 bis 2002. In: Falter, Jürgen W./Gabriel, Oscar W./Weßels, Bernhard (Hrsg.): Wahlen und Wähler. Analysen aus Anlass der Bundestagswahl 2002. Wiesbaden: VS Verlag für Sozialwissenschaften: 189-212.
- Kellermann, Charlotte/Rattinger, Hans (2007): Wirtschaftslage, Arbeitslosigkeit und zugeschriebene Regierungsverantwortung als Bestimmungsfaktoren des Wahlverhaltens bei den Bundestagswahlen 2002 und 2005. In: Brettschneider, Frank/Niedermayer, Oskar/Weßels, Bernhard (Hrsg.): Die Bundestagswahl 2005. Analysen des Wahlkampfes und der Wahlergebnisse. Wiesbaden: VS Verlag für Sozialwissenschaften: 369-393.
- Key Jr., Valdimer O. (1955): A Theory of Critical Elections. In: Journal of Politics 17: 3-18.
- Key Jr., Valdimer O. (1960): The Responsible Electorate: Rationality in Presidential Voting, 1936-1960. Cambridge: Belknap Press of Harvard University Press.
- Kirchgässner, Gebhard (1976): Rationales Wählerverhalten und optimales Regierungsverhalten. Ein Beitrag zur Untersuchung des Zusammenhangs zwischen dem wirtschaftlichen und dem politischen Teilsystem moderner demokratischer Gesellschaften. Dissertationsschrift. Konstanz: Universität Konstanz.
- Kirchgässner, Gebhard (1984): Zur Erfassung des Einflusses der Wirtschaftslage auf das Wählerverhalten mit Hilfe von Wahlfunktionen. In: Holler, Manfred J. (Hrsg.): Wahlanalyse: Hypothesen, Methoden und Ergebnisse. München: tuduv: 119-134.

- Kirchgässner, Gebhard (1986): Der Einfluß von Meinungsumfragen auf das Wahlergebnis. In: Klingemann, Hans-Dieter/Kaase, Max (Hrsg.): Wahlen und politischer Prozeß: Analysen aus Anlaß der Bundestagswahl 1983. Op-laden: Westdeutscher Verlag: 232-247.
- Kirchgässner, Gebhard (2009): Die verschwundene Popularitätsfunktion: Sind Arbeitslosigkeit und Inflation für das Wählerverhalten der Deutschen noch von Bedeutung? In: Gabriel, Oscar W./Weßels, Bernhard/Falter, Jürgen W. (Hrsg.): Wahlen und Wähler. Analysen aus Anlass der Bundestagswahl 2005. Wiesbaden: VS Verlag für Sozialwissenschaften: 380-398.
- Kirchgässner, Gebhard/Meyer zu Himmern, Anne (1997): Expected Closeness and Turnout: An Empirical Analysis for the German General Elections, 1983-1994. In: *Public Choice* 91: 3-25.
- Kirchgässner, Gebhard/Wolters, Jürgen (1987): The Influence of Poll Results on Election Outcomes. In: *Mathematical Social Sciences* 13: 165-175.
- Kirchgässner, Gebhard/Wolters, Jürgen (2006): Einführung in die moderne Zeitreihenanalyse. München: Vahlen.
- Kish, Leslie (1965): *Survey Sampling*. New York u. a.: John Wiley & Sons.
- Klärner, Carl (2008): Forecasting the 2008 U.S. House, Senate and Presidential Elections at the District and State Level. In: *PS: Political Science & Politics* 41: 723-728.
- Klein, Markus (2005a): Die TV-Duelle: Events ohne Effekt? In: Güllner, Manfred/Dülmer, Hermann/Klein, Markus/Ohr, Dieter/Quandt, Markus/Rosar, Ulrich/Klingemann, Hans-Dieter (Hrsg.): Die Bundestagswahl 2002. Eine Untersuchung im Zeichen hoher politischer Dynamik. Wiesbaden: VS Verlag für Sozialwissenschaften: 143-159.
- Klein, Markus (2005b): Die „Zauberformel“. Über das erfolgreiche Scheitern des Prognosemodells von Gschwend und Norpoth bei der Bundestagswahl 2005. In: *Politische Vierteljahresschrift* 46: 689-691.
- Klein, Markus/Plötschke, Manuela (2000): Wählen im sozialen Kontext: Mehrebenenanalysen des Wählerverhaltens bei den Bundestagswahlen der Jahre 1969 bis 1998. In: Klein, Markus/Jagodzinski, Wolfgang/Mochmann, Ekkehard/Ohr, Dieter (Hrsg.): 50 Jahre Empirische Wahlforschung in Deutschland. Entwicklung, Befunde, Perspektiven, Daten. Wiesbaden: Westdeutscher Verlag: 182-211.
- Klein, Markus/Rosar, Ulrich (2005a): Die Wähler ziehen Bilanz: Determinanten der Wahlteilnahme und der Wahlentscheidung. In: Güllner, Manfred/Dülmer, Hermann/Klein, Markus/Ohr, Dieter/Quandt, Markus/Rosar, Ulrich/Klingemann, Hans-Dieter (Hrsg.): Die Bundestagswahl 2002. Eine Untersuchung im Zeichen hoher politischer Dynamik. Wiesbaden: VS Verlag für Sozialwissenschaften: 181-190.

- Klein, Markus/Rosar, Ulrich (2005b): Physische Attraktivität und Wahlerfolg. Eine empirische Analyse am Beispiel der Wahlkreiskandidaten bei der Bundestagswahl 2002. In: Politische Vierteljahresschrift 46: 263-287.
- Klein, Markus/Rosar, Ulrich (2007): Wirkungen des TV-Duells im Vorfeld der Bundestagswahl 2005 auf die Wahlentscheidung. Eine empirische Analyse unter besonderer Berücksichtigung von Medieneinflüssen auf die Siegerwahrnehmung und subjektiven Erwartungshaltung an die Debattenperformance der Kandidaten. In: Kölner Zeitschrift für Soziologie und Sozialpsychologie 59: 81-104.
- Klein, Markus/Rosar, Ulrich (2009): Sie, Sie, Sie, oder Er? Die Kanzlerkandidatur Angela Merkels im Spiegel der Daten einer experimentellen Befragung. In: Gabriel, Oscar W./Weßels, Bernhard/Falter, Jürgen W. (Hrsg.): Wahlen und Wähler. Analysen aus Anlass der Bundestagswahl 2005. Wiesbaden: VS Verlag für Sozialwissenschaften: 346-357.
- Klingemann, Hans-Dieter (2002): Politikwissenschaftliche Forschung in Deutschland von 1949 bis Ende der 60er Jahre. In: van Deth, Jan W. (Hrsg.): Von Generation zu Generation. Mannheim: ZUMA: 51-58.
- Klingemann, Hans-Dieter/Kaase, Max (Hrsg.) (2001): Wahlen und Wähler. Analysen aus Anlass der Bundestagswahl 1998. Wiesbaden: Westdeutscher Verlag.
- Kou, Steven G./Sobel, Michael E. (2004): Forecasting the Vote: A Theoretical Comparison of Election Markets and Public Opinion Polls. In: Political Analysis 12: 277-295.
- Kramer, Gerald H. (1971): Short-Term Fluctuations in U.S. Voting Behavior, 1896-1964. In: American Political Science Review 65: 131-143.
- Krauß, Fritz (1994): Gewichtung bei Umfragen zum politischen Meinungsklima. In: Gabler, Siegfried/Hoffmeyer-Zlotnik, Jürgen H. P./Krebs, Dagmar (Hrsg.): Gewichtung in der Umfragepraxis. Opladen: Westdeutscher Verlag: 152-165.
- Kriesi, Hanspeter (1998): The Transformation of Cleavage Politics. The 1997 Stein Rokkan Lecture. In: European Journal of Political Research 36: 165-185.
- Kruke, Anja (2007): Demoskopie in der Bundesrepublik Deutschland. Meinungsforschung, Parteien und Medien 1949-1990. Düsseldorf: Droste.
- Kühn, Arthur (1970): Das Problem der Prognose in der Soziologie. Berlin: Duncker & Humblot.
- Kühnel, Steffen (2001): Kommt es auf die Stimme an? Determinanten von Teilnahme und Nichtteilnahme an politischen Wahlen. In: Koch, Achim/Wasmer, Martina/Schmidt, Peter (Hrsg.): Politische Partizipation in der Bundesrepublik Deutschland. Empirische Befunde und theoretische Erklärungen. Opladen: Leske + Budrich: 11-42.
- Labaw, Patricia (1982): Advanced Questionnaire Design. Cambridge: Abt Books.

- Lau, Richard R. (1994): An Analysis of the Accuracy of „Trial Heat“ Polls During the 1992 Presidential Election. In: *Public Opinion Quarterly* 58: 2-20.
- Lazarsfeld, Paul F./Berelson, Bernard R./Gaudet, Hazel (1955): *The People's Choice. How the Voter Makes Up His Mind in a Presidential Campaign*. 2. Aufl. New York: Columbia University Press.
- Lenk, Hans (1972): *Erklärung, Prognose, Planung. Skizzen zu Brennpunktproblemen der Wissenschaftstheorie*. Freiburg: Rombach.
- Levy, Mark R. (1983): The Methodology and Performance of Election Day Polls. In: *Public Opinion Quarterly* 47: 54-67.
- Lewis-Beck, Michael S. (1985): Election Forecasts in 1984: How Accurate Were They? In: *PS: Political Science and Politics* 18: 53-62.
- Lewis-Beck, Michael S. (2005): Election Forecasting: Principles and Practice. In: *British Journal of Politics and International Relations* 7: 145-164.
- Lewis-Beck, Michael S./Paldam, Martin (2000): Economic Voting: An Introduction. In: *Electoral Studies* 19: 113-121.
- Lewis-Beck, Michael S./Rice, Tom W. (1984): Forecasting Presidential Elections: A Comparison of Naive Models. In: *Political Behavior* 6: 9-21.
- Lewis-Beck, Michael S./Rice, Tom W. (1992): *Forecasting Elections*. Washington: *Congressional Quarterly*.
- Lewis-Beck, Michael S./Skalaban, Andrew (1989): Citizen Forecasting: Can Voters See into the Future? In: *British Journal of Political Science* 19: 146-153.
- Lewis-Beck, Michael S./Stegmaier, Mary (2000): Economic Determinants of Electoral Outcomes. In: *Annual Review of Political Science* 3: 183-219.
- Lewis-Beck, Michael S./Tien, Charles (1996): The Future in Forecasting. Prospective Presidential Models In: *American Politics Quarterly* 24: 468-491.
- Lewis-Beck, Michael S./Tien, Charles (1999): Voters as Forecasters: A Micro-model of Election Prediction. In: *International Journal of Forecasting* 15: 175-184.
- Lewis-Beck, Michael S./Tien, Charles (2000): The Future in Forecasting. Prospective Presidential Models. In: Campbell, James E./Garand, James C. (Hrsg.): *The Future in Forecasting. Prospective Presidential Models*. Thousand Oaks u. a.: Sage: 83-102.
- Lewis-Beck, Michael S./Tien, Charles (2002): Presidential Election Forecasting: the Bush-Gore Draw. In: Dorbatz, Betty A./Buzzell, Timothy/Waldner, Lisa K. (Hrsg.): *Sociological Views on Political Participation in the 21st Century*. Amsterdam: JAI Press: 173-187.
- Lewis-Beck, Michael S./Tien, Charles (2004): Jobs and the Job of President: A Forecast for 2004. In: *PS: Political Science and Politics* 37: 753-758.

- Lewis-Beck, Michael S./Tien, Charles (2008): Forecasting Presidential Elections: When to Change the Model. In: *International Journal of Forecasting* 24: 227-236.
- Lichtman, Allan J. (2008a): *The Keys to the White House: A Surefire Guide to Predicting the Next President*. Lanham: Rowman & Littlefield.
- Lichtman, Allan J. (2008b): *The Keys to the White House: An Index Forecast for 2008*. In: *International Journal of Forecasting* 24: 299-307.
- Lichtman, Allan J./Keilis-Borok, Vladimir I. (1981): Pattern Recognition Applied to Presidential Elections in the United States, 1860-1980: Role of Integral Social, Economic, and Political Traits. In: *Proceedings of the National Academy of Sciences* 78: 7230-7234.
- Liepelt, Klaus (1966): *Wählermeinung und Wahlprognose. Acht Gesichtspunkte zur Lektüre von Wahlvorhersagen*. Bad Godesberg: Institut für angewandte Sozialwissenschaft.
- Liepelt, Klaus (1996): Wie halten wir es mit den Wahlprognosen? Demoskopie als Meß- und Einflußinstrument. In: Oberreuter, Heinrich (Hrsg.): *Parteiensystem am Wendepunkt? Wahlen in der Fernsehdemokratie*. München: Olzog: 94-119.
- Linhart, Eric (2007): Rationales Wählen als Reaktion auf Koalitionssignale am Beispiel der Bundestagswahl 2005. In: *Politische Vierteljahresschrift* 48: 461-484.
- Lipset, Seymour M./Rokkan, Stein (1967): Cleavage Structures, Party Systems, and Voter Alignments: An Introduction. In: Lipset, Seymour M./Rokkan, Stein (Hrsg.): *Party Systems and Voter Alignments: Cross-National Perspectives*. New York u. a.: Free Press, Collier-Macmillan: 1-64.
- Lipsmeier, Gero (1999): Standard oder Fehler? Einige Eigenschaften von Schätzverfahren bei komplexen Stichprobenplänen und aktuelle Lösungsansätze. In: *ZA-Information* 1999: 96-117.
- Loewenstein, Karl (1971): Vorschläge zur Kontrolle der Meinungsforschung. In: *Juristenzeitung* 26: 529-532.
- Lomborg, Bjørn (2001): *The Skeptical Environmentalist. Measuring the Real State of the World*. Cambridge u. a.: Cambridge University Press.
- Lupri, Eugen (1969): Soziologische Bedeutung der Wahlprognose. In: Hartmann, Klaus D. (Hrsg.): *Politische Beeinflussung. Voraussetzungen, Ablauf und Wirkungen*. Frankfurt am Main: Europäische Verlagsanstalt: 99-116.
- Maas, Cora J. M./Hox, Joop J. (2005): Sufficient Sample Sizes for Multilevel Modeling. In: *Methodology* 1: 86-92.
- MacKuen, Michael B./Erikson, Robert S./Stimson, James A. (1989): Macropartisanship. In: *American Political Science Review* 83: 1125-1142.

- Mandelbrot, Benoit (1953): An Information Theory of the Statistical Structure of Language. In: Jackson, Willis (Hrsg.): Communication Theory. New York: Academic Press: 503-512.
- Mandelbrot, Benoit (1959): A Note on a Class of Skew Distribution Functions: Analysis and Critique of a Paper by H. A. Simon. In: Information and Control 2: 90-99.
- Mandelbrot, Benoit (1961): Final Note on a Class of Skew Distribution Functions: Analysis and Critique of a Model by H. A. Simon. In: Information and Control 4: 198-216.
- Markus, Gregory B./Converse, Philip E. (1979): A Dynamic Simultaneous Equation Model of Electoral Choice. In: The American Political Science Review 73: 1055-1070.
- Martin, Elizabeth A./Traugott, Michael W./Kennedy, Courtney (2005): A Review and Proposal for a New Measure of Poll Accuracy. In: Public Opinion Quarterly 69: 342-369.
- Maslow, Abraham (1943): A Theory of Human Motivation. In: Psychological Review 50: 370-396.
- Maurer, Marcus (2008): Warum „Wahlprognosen“ scheitern – vier Thesen zur Rationalität von Wahlentscheidungen in der Mediendemokratie. In: Politische Studien 57: 41-52.
- Mayntz, Renate (2004): Mechanisms in the Analysis of Social Macro-Phenomena. In: Philosophy of the Social Sciences 34: 237-259.
- McAllister, Ian/Studlar, Donley T. (1991): Bandwagon, Underdog, or Projection? Opinion Polls and Electoral Choice in Britain, 1979-1987. In: Journal of Politics 53: 720-741.
- McCowan, Brenda/Doyle, Laurance R./Jenkins, Jon M./Hanser, Sean F. (2005): The Appropriate Use of Zipf's Law in Animal Communication Studies. In: Animal Behavior 69: F1-F7.
- Merkle, Daniel/Langer, Gary/Lammert, David (2008): Methodological Issues in Pre-Election Polling: Lessons from ABC News' 32-Night Tracking Poll. In: Carballo, Marita/Hjelmar, Ulf (Hrsg.): Public Opinion Polling in a Globalized World. Berlin u. a.: Springer: 109-134.
- Merrill III, Samuel/Grofman, Bernard (1999): A Unified Theory of Voting. Directional and Proximity Spatial Models. Cambridge: Cambridge University Press.
- Merritt, Anna J./Merritt, Richard L. (Hrsg.) (1970): Public Opinion in Occupied Germany. The OMGUS Surveys, 1945-1949. Urbana: University of Illinois Press.

- Merritt, Anna J./Merritt, Richard L. (Hrsg.) (1979): *Public Opinion in Semi-sovereign Germany. The HICOG Surveys, 1949-1955*. Urbana: University of Illinois Press.
- Merton, Robert K. (1936): *The Unanticipated Consequences of Purposive Social Action*. In: *American Sociological Review* 1: 894-904.
- Merton, Robert K./Kitt, Alice S. (1950): *Contributions to the Theory of Reference Group Behaviour*. In: Merton, Robert K./Lazarsfeld, Paul F. (Hrsg.): *Continuities in Social Research: Studies in the Scope and Method of The American Soldier*. Glencoe: Free Press: 40-105.
- Miller, Arthur H. (1979): *Normal Vote Analysis. Sensitivity to Change over Time*. In: *American Journal of Political Science* 23: 406-425.
- Miller, Warren E./Shanks, J. Merrill (1996): *The New American Voter*. Cambridge u. a.: Harvard University Press.
- Mitofsky, Warren J. (1998): *The Polls-Review: Was 1996 a Worse Year for Polls than 1948?* In: *Public Opinion Quarterly* 62: 230-249.
- Moeller, Leslie G./Crossley, Archibald M./Gallup, George H./Lazarsfeld, Paul F./Stouffer, Samuel A. (1949): *The SSRC Committee Report*. In: Meier, Norman C./Saunders, Harold W. (Hrsg.): *The Polls and Public Opinion. The Iowa Conference on Attitude and Opinion Research Sponsored by the State University of Iowa, Iowa City*. New York: Henry Holt: 200-214.
- Morgan, Stephen L./Winship, Christopher (2007): *Counterfactuals and Causal Inference. Methods and Principles for Social Research*. Cambridge u. a.: Cambridge University Press.
- Moser, Klaus A./Kalton, Graham (1972): *Survey Methods in Social Investigation*. 2. Aufl. New York: Basic Books.
- Mosteller, Frederick (1965): *Fifty Challenging Problems in Probability with Solutions*. Mineola: Dover.
- Mosteller, Frederick/Hyman, Herbert/Mccarthy, Philip J./Marks, Eli S./Truman, David B. (1949): *The Pre-Election Polls of 1948. Report to the Committee on Analysis of Pre-Election Polls and Forecasts*. New York: Social Science Research Council.
- Mueller, Dennis C. (2003): *Public Choice III*. Cambridge u. a.: Cambridge University Press.
- Mueller, John E. (1970): *Presidential Popularity from Truman to Johnson*. In: *American Political Science Review* 64: 18-34.
- Müller, Walter (1997): *Sozialstruktur und Wahlverhalten. Eine Widerrede gegen die Individualisierungsthese*. In: *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 49: 747-760.

- Müller, Walter (1998): Klassenstruktur und Parteiensystem zum Wandel der Klassenspaltung im Wahlverhalten. In: Kölner Zeitschrift für Soziologie und Sozialpsychologie 50: 3-46.
- Müller, Walter (2000): Klassenspaltung im Wahlverhalten – Eine Reanalyse. In: Kölner Zeitschrift für Soziologie und Sozialpsychologie 52: 790-795.
- Müller, Wolfgang (1999): Sozialstrukturelle Clavages bei Bundestagswahlen in Theorie und Praxis. Persistenz, Realignment oder Dealignment? Frankfurt am Main: Peter Lang.
- Nannestad, Peter/Paldam, Martin (1994): The VP-Function: A Survey of the Literature on Vote and Popularity Functions After 25 Years. In: Public Choice 79: 213-245.
- Nannestad, Peter/Paldam, Martin (2002): The Cost of Ruling. A Foundation Stone for Two Theories. In: Dorussen, Han/Taylor, Michael (Hrsg.): The Cost of Ruling. A Foundation Stone for Two Theories. London: Routledge: 17-44.
- Naschold, Frieder (1971): Wahlprognosen und Wählerverhalten in der BRD. Stuttgart u. a.: W. Kohlhammer.
- Neff, Joachim (1971): Demokratie und Demoskopie. In: Juristenzeitung 26: 16-18.
- Neller, Katja (2005): Kooperation und Verweigerung: Eine Non-Response-Studie. In: ZUMA Nachrichten 29: 9-36.
- Neu, Viola (2008): Die neue Unberechenbarkeit der Wähler. In: Politische Studien 57: 25-33.
- Neumann, Erich P. (1952): Politische und soziale Meinungsforschung in Deutschland. In: Institut zur Förderung öffentlicher Angelegenheiten (Hrsg.): Empirische Sozialforschung. Meinungs- und Marktforschung. Methoden und Probleme. Frankfurt am Main: Institut zur Förderung öffentlicher Angelegenheiten: 44-51.
- Neumann, Manfred (2000): Wettbewerbspolitik. Geschichte, Theorie und Praxis. Wiesbaden: Gabler.
- Nieuwebeerta, Paul (1995): The Democratic Class Struggle in Twenty Countries 1945-1990. Amsterdam: Thesis.
- Niskanen, William A. (1979): Economic and Political Effects on the Popular Vote for the President. In: Rae, Douglas W./Eismeier, Theodore J. (Hrsg.): Public Policy and Public Choice. Beverly Hills u. a.: Sage: 93-117.
- Noelle, Elisabeth (1967): Öffentliche Meinung und soziale Kontrolle. In: Noelle, Elisabeth/Neumann, Erich P. (Hrsg.): Öffentliche Meinung und soziale Kontrolle. Allensbach: Verlag für Demoskopie: XXV-XLV.
- Noelle, Elisabeth/Neumann, Erich P. (Hrsg.) (1956): Jahrbuch der öffentlichen Meinung 1947-1955. Allensbach: Verlag für Demoskopie.
- Noelle, Elisabeth/Neumann, Erich P. (Hrsg.) (1957): Jahrbuch der öffentlichen Meinung 1957. Allensbach: Verlag für Demoskopie.

- Noelle, Elisabeth/Neumann, Erich P. (Hrsg.) (1965): *Jahrbuch der öffentlichen Meinung 1958-1964*. Allensbach: Verlag für Demoskopie.
- Noelle, Elisabeth/Neumann, Erich P. (Hrsg.) (1967): *Jahrbuch der öffentlichen Meinung 1965-1967*. Allensbach: Verlag für Demoskopie.
- Noelle, Elisabeth/Neumann, Erich P. (Hrsg.) (1974): *Jahrbuch der öffentlichen Meinung 1968-1973*. Allensbach: Verlag für Demoskopie.
- Noelle-Neumann, Elisabeth (Hrsg.) (1976): *Allensbacher Jahrbuch der Demoskopie 1974-1976*. Wien u. a.: Fritz Molden.
- Noelle-Neumann, Elisabeth (1983): *Political Opinion Polling in Germany*. In: Worcester, Robert M. (Hrsg.): *Political Opinion Polling. An International Review*. New York: St. Martin's Press: 44-60.
- Noelle-Neumann, Elisabeth (1988): *Die befragte Nation. Möglichkeiten, Wirkungen und Gefahren der Demoskopie*. In: Willms, Bernhard (Hrsg.): *Handbuch zur deutschen Nation. Bd. 3: Moderne Wissenschaft und Zukunftsperspektive*. Tübingen: Hohenrain-Verlag: 241-271.
- Noelle-Neumann, Elisabeth (1992): *Die Entwicklung der empirischen Sozialforschung in Deutschland. Bericht einer Zeitzeugin*. In: Jaufman, Dieter/Kistler, Ernst/Meier, Klaus/Strech, Karl-Heinz (Hrsg.): *Empirische Sozialforschung im vereinten Deutschland: Bestandsaufnahme und Perspektiven*. Frankfurt am Main u. a.: Campus: 47-59.
- Noelle-Neumann, Elisabeth (2000): *Der Beitrag der Wahlforschung zur Methodenentwicklung in der empirischen Sozialforschung*. In: Klein, Markus/Jagodzinski, Wolfgang/Mochmann, Ekkehard/Ohr, Dieter (Hrsg.): *50 Jahre Empirische Wahlforschung in Deutschland. Entwicklung, Befunde, Perspektiven, Daten*. Wiesbaden: Westdeutscher Verlag: 59-81.
- Noelle-Neumann, Elisabeth (2001): *Die Schweigespirale. Öffentliche Meinung – unsere soziale Haut*. 6. Aufl. München: Langen Müller.
- Noelle-Neumann, Elisabeth/Köcher, Renate (Hrsg.) (2002a): *Allensbacher Jahrbuch der Demoskopie 1998-2000*. München: Saur.
- Noelle-Neumann, Elisabeth/Köcher, Renate (Hrsg.) (2002b): *Allensbacher Jahrbuch der Demoskopie 1998-2000. Band 11*. München: Saur.
- Noelle-Neumann, Elisabeth/Petersen, Thomas (2005): *Alle, nicht jeder. Einführung in die Methoden der Demoskopie*. 4. Aufl. Berlin u. a.: Springer.
- Nohlen, Dieter (2007): *Wahlrecht und Parteiensystem. Zur Theorie und Empirie der Wahlsysteme*. 5. Aufl. Opladen u. a.: Barbara Budrich.
- Norpoth, Helmut (1996): *Of Time and Candidates. A Forecast for 1996*. In: *American Politics Quarterly* 24: 443-467.
- Norpoth, Helmut/Gschwend, Thomas (2003): *Against All Odds? The Red-Green Victory*. In: *German Politics and Society* 21: 15-34.

- Norpoth, Helmut/Gschwend, Thomas (2004): Mit Rot-Grün ins Schwarze getroffen: Prognosemodell besteht Feuertaufe. Arbeitspapier Nr. 75. Mannheim: Mannheimer Zentrum für Europäische Sozialforschung. Online: <http://www.mzes.uni-mannheim.de/publications/wp/wp-75.pdf> (10. November 2008).
- Norpoth, Helmut/Gschwend, Thomas (2009): The Chancellor Model: Forecasting German Elections. Mannheim: Center for Doctoral Studies in Social and Behavioral Sciences. Online: http://www.sowi.uni-mannheim.de/gschwend/pdf/publications/BTW09_IntJourn_final.pdf (3. Januar 2010).
- Oedegaard, Ingvill C. (2000): Lebensstile, soziale Milieus und Wahlverhalten in Westdeutschland. In: Klein, Markus/Jagodzinski, Wolfgang/Mochmann, Ekkehard (Hrsg.): 50 Jahre empirische Wahlforschung in Deutschland. Entwicklung, Befunde, Perspektiven, Daten. Wiesbaden: Westdeutscher Verlag: 212-234.
- Ohr, Dieter (2005): Wahlen und Wählerverhalten im Wandel: Der individualisierte Wähler in der Mediendemokratie. In: Güllner, Manfred/Dülmer, Hermann/Klein, Markus/Ohr, Dieter/Quandt, Markus/Rosar, Ulrich/Klingemann, Hans-Dieter (Hrsg.): Die Bundestagswahl 2002. Eine Untersuchung im Zeichen hoher politischer Dynamik. Wiesbaden: VS Verlag für Sozialwissenschaften: 15-30.
- Oliven, Kenneth/Rietz, Thomas A. (2004): Suckers are Born but Markets are Made: Individual Rationality, Arbitrage, and Market Efficiency on an Electronic Futures Market. In: *Management Science* 50: 336-351.
- Olson Jr., Mancur (1968): Die Logik des kollektiven Handelns. Kollektivgüter und die Theorie der Gruppen. Tübingen: J. C. B. Mohr (Paul Siebeck).
- Olson Jr., Mancur (1991): Aufstieg und Niedergang von Nationen. Ökonomisches Wachstum, Stagflation und soziale Starrheit. 2. Aufl. Tübingen: J. C. B. Mohr (Paul Siebeck).
- Opp, Karl-Dieter (2005): Methodologie der Sozialwissenschaften. Einführung in Probleme ihrer Theorienbildung und praktischen Anwendung. 6. Aufl. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Oprea, Ryan/Porter, David/Hibbert, Chris/Hanson, Robin/Tila, Dorina (2007): Can Manipulators Misdlead Market Observers? Online: <http://hanson.gmu.edu/judges.pdf> (5. Juni 2008).
- Otte, Gunnar (1997): Lebensstile versus Klassen – welche Sozialstrukturkonzeption kann die individuelle Parteipräferenz besser erklären? In: Müller, Walter (Hrsg.): Soziale Ungleichheit. Neue Befunde zu Strukturen, Bewußtsein und Politik. Opladen: Leske + Budrich: 303-346.

- Paldam, Martin (1991): How Robust is the Vote Function? In: Norpoth, Helmut/Lewis-Beck, Michael S./Lafay, Jean-Dominique (Hrsg.): How Robust is the Vote Function? Ann Arbor: University of Michigan Press: 9-31.
- Paldam, Martin (1997): Political Business Cycles. In: Mueller, Dennis C. (Hrsg.): Economics and Politics: The Calculus of Support. Cambridge u. a.: Cambridge University Press: 342-370.
- Paldam, Martin/Skott, Peter (1995): A Rational-Voter Explanation of the Cost of Ruling. In: Public Choice 83: 159-172.
- Pappi, Franz U. (1977): Sozialstruktur, gesellschaftliche Wertorientierungen und Wahlabsicht. Ergebnisse eines Zeitvergleichs des deutschen Elektorats 1953 und 1976. In: Kaase, Max (Hrsg.): Wahlsoziologie heute. Analysen aus Anlaß der Bundestagswahl 1976. Opladen: Westdeutscher Verlag: 195-229.
- Pappi, Franz U. (1985): Die konfessionell-religiöse Konfliktlinie in der deutschen Wählerschaft – Entstehung, Stabilität und Wandel. In: Oberndörfer, Dieter/Rattinger, Hans/Schmitt, Karl (Hrsg.): Wirtschaftlicher Wandel, religiöser Wandel und Wertewandel. Folgen für das politische Verhalten in der Bundesrepublik Deutschland. Berlin: Duncker & Humblot: 263-290.
- Pappi, Franz U. (1986): Das Wahlverhalten sozialer Gruppen bei Bundestagswahlen im Zeitvergleich. In: Klingemann, Hans-Dieter/Kaase, Max (Hrsg.): Wahlen und politischer Prozeß. Analysen aus Anlaß der Bundestagswahl 1983. Opladen: Westdeutscher Verlag: 369-384.
- Pappi, Franz U. (2002): Die politisierte Sozialstruktur heute: Historische Reminiscenz oder aktuelles Erklärungspotential? In: Brettschneider, Frank/van Deth, Jan W./Roller, Edeltraud (Hrsg.): Die Das Ende der politisierten Sozialstruktur? Opladen: Leske + Budrich: 25-48.
- Pappi, Franz U./Bytzek, Evelyn (2007): Wirtschaftslage und Popularität der Regierungen Schröder I und II. In: Brettschneider, Frank/Niedermayer, Oskar/Weßels, Bernhard (Hrsg.): Die Bundestagswahl 2005. Analysen des Wahlkampfes und der Wahlergebnisse. Wiesbaden: VS Verlag für Sozialwissenschaften: 349-368.
- Pappi, Franz U./Herrmann, Michael (2006): Strategisches Erststimmenwählen bei deutschen Bundestagswahlen. MZES Working Paper. Mannheim: Mannheimer Zentrum für Europäische Sozialforschung. Online: <http://www.mzes.uni-mannheim.de/publications/wp/wp-92.pdf> (12. Januar 2010).
- Pappi, Franz U./Shikano, Susumu (2001): Personalisierung der Politik in Mehrparteiensystemen am Beispiel deutscher Bundestagswahlen seit 1980. In: Politische Vierteljahresschrift 42: 355-387.

- Pappi, Franz U./Shikano, Susumu (2002): Die politisierte Sozialstruktur als mittelfristig stabile Basis einer deutschen Normalwahl. In: *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 54: 444-475.
- Pappi, Franz U./Shikano, Susumu (2007): *Wahl- und Wählerforschung*. Baden-Baden: Nomos.
- Pappi, Franz U./Thurner, Paul W. (2002): Electoral Behaviour in a Two-Vote System: Incentives for Ticket Splitting in German Bundestag Elections. In: *European Journal of Political Research* 41: 207-232.
- Pappi, Franz U./Shikano, Susumu/Bytzeck, Evelyn (2004): Der Einfluss politischer Ereignisse auf die Popularität von Parteien und Politikern und auf das Parteiensystem. In: *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 56: 51-70.
- Pearl, Judea (2000): *Causality. Models, Reasoning, and Inference*. Cambridge u. a.: Cambridge University Press.
- Peltzman, Sam (1992): Voters as Fiscal Conservatives. In: *Quarterly Journal of Economics* 107: 327-362.
- Pennock, David M./Lawrence, Steve/Giles, C. Lee/Nielsen, Finn A. (2001): *The Power of Play: Efficiency and Forecast Accuracy in Web Market Games*. NEC Research Institute Technical Report. Princeton: NEC Research Institute. Online: <http://artificialmarkets.com/am/pennock-neci-tr-2000-168.pdf> (13. Mai 2008).
- Petersen, Thomas (2008): Public Opinion Polling. *The International Encyclopedia of Communication*. Malden u. a.: Blackwell. Online: http://www.communicationencyclopedia.com/subscriber/tocnode?id=g9781405131995_chunk_g978140513199521_ss126-1 (14. November 2008).
- Petersen, Thomas/Voß, Peter/Sabel, Peter/Grube, Norbert (2004): Der Fragebogen Karls des Großen. In: *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 56: 736-745.
- Pickup, Mark/Johnston, Richard (2008): Campaign Trial Heats as Election Forecasts: Measurement Error and Bias in 2004 Presidential Campaign Polls. In: *International Journal of Forecasting* 24: 270-282.
- Pies, Ingo/Leschke, Martin (Hrsg.) (1997): *Mancur Olsons Logik kollektiven Handelns*. Tübingen: Mohr.
- Plischke, Thomas/Rattinger, Hans (2009): „Zittrige Wählerhand“ oder invalides Messinstrument? Zur Plausibilität von Wahlprojektionen am Beispiel der Bundestagswahl 2005. In: Gabriel, Oscar W./Weßels, Bernhard/Falter, Jürgen W. (Hrsg.): *Wahlen und Wähler. Analysen aus Anlass der Bundestagswahl 2005*. Wiesbaden: VS Verlag für Sozialwissenschaften: 484-509.
- Plott, Charles R./Smith, Vernon L. (Hrsg.) (2008): *Handbook of Experimental Economics Results, Volume 1*. Amsterdam: North-Holland.

- Pokropp, Fritz (1996): *Stichproben: Theorie und Verfahren*. 2. Aufl. München u. a.: Oldenbourg.
- Pool, Ithiel de Sla/Abelson, Robert P./Popkin, Samuel L. (1964): *Candidates, Issues and Strategies. A Computer Simulation of the 1960 Presidential Election*. Cambridge: MIT Press.
- Popkin, Samuel L. (1991): *The Reasoning Voter. Communication and Persuasion in Presidential Campaigns*. Chicago: University of Chicago Press.
- Popper, Karl R. (1980): *Prognose und Prophetie in den Sozialwissenschaften*. In: Gottschlich, Maximilian/Langenbacher, Wolfgang R. (Hrsg.): *Publizistik- und Kommunikationswissenschaft. Ein Textbuch zur Einfhrung*. Wien: Willhelm Braumller: 23-33.
- Popper, Karl R. (1989): *Logik der Forschung*. 9. Aufl. Tbingen: Mohr.
- Porst, Rolf (1996): *Ausschpfungen bei sozialwissenschaftlichen Umfragen. Die Sicht der Institute*. ZUMA-Arbeitsbericht. Mannheim: ZUMA. Online: http://www.social-science-geis.de/Publikationen/Berichte/ZUMA_Arbeitsberichte/96/96_07abs.htm (3. Oktober 2008).
- Posner, Richard (2008): *Prediction Markets and the Election*. The Becker-Posner Blog. A Blog by Gary Becker and Richard Posner. Chicago: Universitt Chicago. Online: <http://www.becker-posner-blog.com/archives/2008/09/> (15. September 2008).
- Rabe-Hesketh, Sophia/Skrondal, Anders (2008): *Multilevel and Longitudinal Modeling Using Stata*. 2. Aufl. College Station: Stata Press.
- Radtke, Gnter D. (1977): *Zur Prognosekompetenz der empirischen Wahlforschung*. In: Kaase, Max (Hrsg.): *Wahlsoziologie heute. Analysen aus Anla der Bundestagswahl 1976*. Opladen: Westdeutscher Verlag: 669-688.
- Rapoport, Anatol (1982): *Zipf's Law Revisited*. In: Guiter, Henri/Arapov, Michail V. (Hrsg.): *Studies on Zipf's Law*. Bochum: Brockmeyer: 1-28.
- Rattinger, Hans (1994): *Parteiidentifikation in Ost- und Westdeutschland nach der Vereinigung*. In: Niedermayer, Oskar/Von Beyme, Klaus (Hrsg.): *Politische Kultur im vereinten Deutschland*. Berlin: Akademie Verlag: 77-104.
- Rattinger, Hans/Faas, Thorsten (2001): *Wahrnehmungen der Wirtschaftslage und Wahlverhalten 1977 bis 1998*. In: Klingemann, Hans-Dieter/Kaase, Max (Hrsg.): *Wahlen und Whler. Analysen aus Anlass der Bundestagswahl 1998*. Wiesbaden: Westdeutscher Verlag: 283-307.
- Rattinger, Hans/Ohr, Dieter (1989): *Wahlprognosen in einer Welt ohne Stichprobenfehler: Analytische berlegungen und empirische Befunde*. In: Falter, Jrgen W./Rattinger, Hans/Troitczsch, Klaus G. (Hrsg.): *Wahlen und politische Einstellungen in der Bundesrepublik Deutschland. Neuere Entwicklungen der Forschung*. Frankfurt am Main: Peter Lang: 282-332.

- Rattinger, Hans/Roßteutscher, Sigrid/Schmitt-Beck, Rüdiger/Weßels, Bernhard (2009): Germal Longitudanal Election Study (GLES). Fragebogen GLE-S1004 (Pre-Release 1.0), Komponente X Online-Tracking IV: Koalitionen. Mannheim: Deutsche Gesellschaft für Wahlforschung. Online: http://www.gesis.org/wahlportal/fileadmin/wahlportal/SecureDownload/fragebogen/GLE_S1004_Prel.0%20-%20Fragebogen.pdf (11. Dezember 2009).
- Raupp, Juliana (2007): Politische Meinungsforschung. Die Verwendung von Umfragen in der politischen Kommunikation. Konstanz: UVK.
- Ray, Russ (2006): Prediction Markets and the Financial „Wisdom of Crowds“. In: *Journal of Behavioral Finance* 7: 2-4.
- Rehmet, Frank (2002): Direkte Demokratie in den deutschen Bundesländern. In: Schiller, Theo/Mittendorf, Volker (Hrsg.): *Direkte Demokratie in den deutschen Bundesländern*. Wiesbaden: Westdeutscher Verlag: 102-114.
- Reinecke, Jost (1991): Interviewer- und Befragtenverhalten. Theoretische Ansätze und methodische Konzepte. Opladen: Westdeutscher Verlag.
- Rhode, Paul W./Strumpf, Koleman S. (2007): *Manipulating Political Stock Markets: A Field Experiment and a Century of Observational Data*. Tucson: University of Arizona. Online: http://www.unc.edu/~cigar/papers/ManipHIT_Jan2007.pdf (26. November 2008).
- Rinne, Horst (2008): *Taschenbuch der Statistik*. 4. Aufl. Frankfurt am Main: Harri.
- Ritter, Gebhard (1992): *Der Streit um die direkte Demokratie in der Bundesrepublik Deutschland*. Dissertationsschrift. Bonn: Rheinische Friedrich-Wilhelms-Universität zu Bonn.
- Roper, Burns W. (1984): Are Polls Accurate? In: *Annals of the American Academy of Political and Social Science* 472: 24-34.
- Roper, Burns W. (1986): Evaluating Polls with Poll Data In: *Public Opinion Quarterly* 50: 10-16.
- Rosenstone, Steven J. (1983): *Forecasting Presidential Elections*. New Haven u. a.: Yale University Press.
- Roth, Dieter (1998): *Empirische Wahlforschung. Ursprung, Theorien, Instrumente und Methoden*. Opladen: Leske + Budrich.
- Rückmann, Kurt (1972): *Demoskopie oder Demagogie? Zur Meinungsforschung in der BRD*. Frankfurt am Main: Marxistische Blätter.
- Russo, Frederica/Williamson, Jon (Hrsg.) (2007): *Causality and Probability in the Sciences*. London: College.
- Sahner, Heinz (1998): Die ASI informiert: Richtlinie für die Veröffentlichung von Ergebnissen der Wahlforschung. In: *Soziale Welt* 49: 97-100.
- Salmon, Wesley C. (1990): *Four Decades of Scientific Explanation*. Minneapolis: University of Minnesota Press.

- Sanders, David (1991): Government Popularity and the Next General Election. In: *Political Quarterly*: 235-261.
- Sanders, David (2005): Popularity Function Forecasts for the 2005 UK General Election. In: *British Journal of Politics and International Relations* 7: 174-190.
- Schaefer, Wolfgang/Miller, Mungo (1998): Schwierigkeiten der Umfrageforschung in den fünfziger Jahren in Deutschland: Erinnerungen und Beobachtungen. In: *ZUMA Nachrichten* 1998: 8-35.
- Schaffer, Lena-Maria/Schneider, Gerald (2005): Die Prognosegüte von Wahlbörsen und Meinungsumfragen zur Bundestagswahl 2005. In: *Politische Vierteljahresschrift* 46: 674-681.
- Scharp, Heinrich (1958): Vom Sinn und Unsinn der politischen Demoskopie. In: *Soziale Welt* 9: 336-344.
- Scheuch, Erwin K. (1956): Die Anwendung von Auswahlverfahren bei Repräsentativ-Befragungen. Unter besonderer Berücksichtigung der Sozialforschung durch akademische Institute. Dissertationsschrift. Köln: Universität zu Köln.
- Scheuch, Erwin K. (1990): Von der Pioniertat zur Institution. Beobachtungen zur Entwicklung der empirischen Sozialforschung. In: Franke, Dieter/Scharioth, Joachim (Hrsg.): *Vierzig Jahre Markt- und Sozialforschung in der Bundesrepublik Deutschland: eine Standortbestimmung*. München u. a.: Oldenbourg: 43-67.
- Scheuch, Erwin K. (2002): Der Aufstieg der empirischen Sozialforschung aus dem Geist des New Deal. In: Sahrer, Heinz (Hrsg.): *Fünfzig Jahre nach Weinheim. Empirische Markt- und Sozialforschung gestern, heute, morgen*. Baden-Baden: Nomos: 51-58.
- Scheuch, Erwin K./Wildenmann, Rudolf (Hrsg.) (1965): *Zur Soziologie der Wahl*. Opladen: Westdeutscher Verlag.
- Schiller, Theo/Mittendorf, Volker (2002): Neue Entwicklungen der direkten Demokratie. In: Schiller, Theo/Mittendorf, Volker (Hrsg.): *Direkte Demokratie. Forschung und Perspektiven*. Wiesbaden: Westdeutscher Verlag: 7-21.
- Schindler, Peter (1999): *Datenhandbuch zur Geschichte des Deutschen Bundestages 1949 bis 1999*. Baden-Baden: Nomos.
- Schlittgen, Rainer/Streitberg, Bernd H. J. (2001): *Zeitreihenanalyse*. 9. Aufl. München: Oldenbourg.
- Schmid, Michael (2006): *Die Logik mechanismischer Erklärungen*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Schmidtchen, Gerhard (1961): *Die repräsentative Quotenauswahl: Bericht über ein Quota-Random-Experiment des Instituts für Demoskopie Allensbach*. Allensbach: Institut für Demoskopie.

- Schmidtchen, Gerhard/Noelle-Neumann, Elisabeth (1963): Die Bedeutung repräsentativer Bevölkerungsumfragen für die offene Gesellschaft. In: Politische Vierteljahresschrift 4: 168-195.
- Schmitt-Beck, Rüdiger (1993): Denn sie wissen nicht, was sie tun ... Zum Verständnis des Verfahrens der Bundestagswahl bei westdeutschen und ostdeutschen Wählern. In: Zeitschrift für Parlamentsfragen 24: 393-415.
- Schmitt-Beck, Rüdiger (1996): Mass Media, the Electorate, and the Bandwagon. A Study of Communication Effects on the Vote Choice in Germany. In: International Journal of Public Opinion Research 8: 266-291.
- Schmitt-Beck, Rüdiger (2001): Ein Sieg der „Kampa“? Politische Symbolik in der Wahlkampagne der SPD und ihre Resonanz in der Wählerschaft. In: Klingemann, Hans-Dieter/Kaase, Max (Hrsg.): Wahlen und Wähler. Analysen aus Anlass der Bundestagswahl 1998. Wiesbaden: Westdeutscher Verlag: 133-162.
- Schmitt-Beck, Rüdiger (2002): Das Nadelöhr am Ende: Die Aufmerksamkeit der Wähler für die Wahlkampfkommunikation als Voraussetzung wirksamer Kampagnen. In: Manchnig, Matthias (Hrsg.): Politik – Medien – Wähler. Wahlkampf im Medienzeitalter. Opladen: Leske + Budrich: 21-48.
- Schmitt-Beck, Rüdiger/Faas, Thorsten (2007): The Campaign and its Dynamics at the 2005 German General Election. In: Clemens, Clayton/Saalfeld, Thomas (Hrsg.): The German Election of 2005. Voters, Parties and Grand Coalition Politics. London u. a.: Routledge: 59-85.
- Schmitt-Beck, Rüdiger/Weick, Stefan/Christoph, Bernhard (2006): Shaky Attachments: Individual-Level Stability and Change of Partisanship among West German Voters, 1984-2001. In: European Journal of Political Research 45: 581-608.
- Schmitt, Hermann (2001): Zur vergleichenden Analyse des Einflusses gesellschaftlicher Faktoren auf das Wahlverhalten. Forschungsfragen, Analysestrategien und einige Ergebnisse. In: Klingemann, Hans-Dieter/Kaase, Max (Hrsg.): Wahlen und Wähler: Analysen aus Anlass der Bundestagswahl 1998. Wiesbaden: Westdeutscher Verlag: 623-645.
- Schnell, Rainer (1997): Nonresponse in Bevölkerungsumfragen. Ausmaß, Entwicklung und Ursachen. Opladen: Leske + Budrich.
- Schnell, Rainer/Kohler, Ulrich (1995): Empirische Untersuchung einer Individualisierungshypothese am Beispiel der Parteipräferenz von 1953-1992. In: Kölner Zeitschrift für Soziologie und Sozialpsychologie 47: 634-657.
- Schnell, Rainer/Kohler, Ulrich (1997): Zur Erklärungskraft sozio-demographischer Variablen im Zeitverlauf. Entgegnung auf Walter Müller sowie auf Wolfgang Jagodzinski und Markus Quandt. In: Kölner Zeitschrift für Soziologie und Sozialpsychologie 49: 783-795.

- Schnell, Rainer/Hill, Paul B./Esser, Elke (1999): Methoden der empirischen Sozialforschung. 6. Aufl. München u. a.: Oldenbourg.
- Schneller, Johannes (1994): Stichprobenbildung nach dem repräsentativen Quoten-Verfahren. In: Gabler, Siegfried/Hoffmeyer-Zlotnik, Jürgen H. P. (Hrsg.): Stichproben in der Umfragepraxis. Opladen: Westdeutscher Verlag: 148-159.
- Schoen, Harald (1999): Mehr oder weniger als fünf Prozent – ist das wirklich die Frage. In: Kölner Zeitschrift für Soziologie und Sozialpsychologie 51: 565-582.
- Schoen, Harald (2002): Wirkungen von Wahlprognosen auf Wahlen. In: Berg, Thomas (Hrsg.): Moderner Wahlkampf. Blick hinter die Kulissen. Opladen: Leske + Budrich: 171-191.
- Schoen, Harald (2004a): Der Kanzler, zwei Sommerthemen und ein Foto-Finish. Priming-Effekte bei der Bundestagswahl 2002. In: Brettschneider, Frank/van Deth, Jan W./Roller, Edeltraud (Hrsg.): Die Bundestagswahl 2002. Analysen der Wahlergebnisse und des Wahlkampfes. Wiesbaden: VS Verlag für Sozialwissenschaften: 23-50.
- Schoen, Harald (2004b): Kandidatenorientierungen im Wahlkampf. Eine Analyse zu den Bundestagswahlkämpfen 1980-1998. In: Politische Vierteljahresschrift 45: 321-345.
- Schoen, Harald (2005a): Wahlkampfforschung. In: Falter, Jürgen W./Schoen, Harald (Hrsg.): Handbuch Wahlforschung. Wiesbaden: VS Verlag für Sozialwissenschaften: 503-542.
- Schoen, Harald (2005b): Soziologische Ansätze in der Wahlforschung. In: Falter, Jürgen W./Schoen, Harald (Hrsg.): Handbuch Wahlforschung. Wiesbaden: VS Verlag für Sozialwissenschaften: 135-185.
- Schoen, Harald/Weins, Cornelia (2005): Der sozialpsychologische Ansatz zur Erklärung von Wahlverhalten. In: Falter, Jürgen W./Schoen, Harald (Hrsg.): Handbuch Wahlforschung. Wiesbaden: VS Verlag für Sozialwissenschaften: 187-242.
- Schultze, Rainer-Olaf (1980): Wahltagbefragung – Chance und Gefahr. In: Zeitschrift für Parlamentsfragen 11. 73-92.
- Schultze, Rainer-Olaf (1991): Außengeleitete Innovation und innengeleiteter Methodenrigorismus – Deutsche Wahlsoziologie auf dem Prüfstand internationalen Vergleichs. In: Zeitschrift für Parlamentsfragen 22. 481-494.
- Servan-Schreiber, Emile/Wolfers, Justin/Pennock, David M./Galebach, Brian (2004): Prediction Markets: Does Money Matter? In: Electronic Markets 14: 243-251.
- Shepsle, Kenneth A. (1991): Models of Multiparty Electoral Competition. Chur: Harwood Academic Publishers.

- Shikano, Susumu/Herrmann, Michael/Thurner, Paul W. (2006): Explaining German Strategic List Voting with a Two-Step Random Utility Approach. Annual Meeting of Midwest Political Science Association.
- Sigelman, Lee (1979): Presidential Popularity and Presidential Elections. In: *Public Opinion Quarterly* 43: 532-534.
- Simon, Herbert A. (1954): Bandwagon and Underdog Effects and the Possibility of Election Predictions In: *Public Opinion Quarterly* 18: 245-253.
- Simon, Herbert A. (1955): On a Class of Skew Distribution Functions. In: *Biometrika* 42: 425-440.
- Simon, Herbert A. (1960): Some Further Notes on a Class of Skew Distribution Functions. In: *Information and Control* 3: 80-88.
- Simon, Herbert A. (1961): Reply to „Final Note” by Benoit Mandelbrot. In: *Information and Control* 4: 217-223.
- Skalaban, Andrew (1988): Do the Polls Affect Elections? Some 1980 Evidence. In: *Political Behavior* 10: 136-150.
- Smith, Tom W. (1978): In Search of House Effects: A Comparison of Responses to Various Questions by Different Survey Organizations. In: *Public Opinion Quarterly* 42: 443-463.
- Smith, Tom W. (1982a): House Effects and the Reproducibility of Survey Measurements: A Comparison of the 1980 GSS and the 1980 American National Election Study. In: *Public Opinion Quarterly* 46: 54-68.
- Smith, Tom W. (1990): The First Straw? A Study of the Origins of Election Polls. In: *Public Opinion Quarterly* 54: 21-36.
- Smith, Vernon L. (1982b): Markets as Economizers of Information: Experimental Examination of the „Hayek Hypothesis”. In: *Economic Inquiry* 20: 165-179.
- Snijders, Tom/Bosker, Roel (1999): *Multilevel Analysis. An Introduction to Basic and Advanced Multilevel Modeling*. London: Sage.
- Snowberg, Erik/Wolfers, Justin (2007): Explaining the Favorite-Longshot Bias: Is it Risk-Love or Misperceptions? Philadelphia: University of Pennsylvania Online: http://bpp.wharton.upenn.edu/jwolfers/Papers/Favorite_Longshot_Bias.pdf (21. Mai 2008).
- Sontheimer, Kurt (1964): Meinungsforschung und Politik. Eine kritische Auseinandersetzung mit den Ansprüchen der Demoskopie. In: *Der Monat* 187: 41-46.
- Sorokin, Pitirim A./Boldyreff, John W. (1932): An Experimental Study of the Influence of Suggestion on the Discrimination and the Valuation of People. In: *American Journal of Sociology* 37: 720-737.
- StataCorp (2007a): *Stata 10: Time Series*. College Station: Stata Press.
- StataCorp (2007b): *Stata Release 10. Reference Q-Z*. College Station: Stata Press.

- Statistisches Bundesamt (2005): Wahlberechtigte, Wähler, Stimmabgabe und Sitzverteilung bei den Bundestagswahlen seit 1949. Wiesbaden: Statistisches Bundesamt. Online: http://www.bundeswahlleiter.de/de/bundestagswahlen/downloads/bundestagswahlergebnisse/btw_ab49_ergebnisse.pdf (7. März 2009).
- Statistisches Bundesamt (2008): Statistisches Jahrbuch 2008 für die Bundesrepublik Deutschland. Wiesbaden: Statistisches Bundesamt. Online: <http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Content/Publikationen/Querschnittsveroeffentlichungen/StatistischesJahrbuch/JahrbuchDownload,templateId=renderPrint.psml> (11. März 2009).
- Statistisches Bundesamt (2009): Erwerbstätige: Deutschland, Jahre, Stellung im Beruf, Geschlecht. Wiesbaden: Statistisches Bundesamt. Online: <https://www-genesis.destatis.de/genesis/online/online.jsessionid=C69323-DD303B45D2A121EEDA4884D6AF.tcgggen1?operation=statistikAbruftabelle&levelindex=0&levelid=1236850392651&index=2> (24. Februar 2009).
- Steeh, Charlotte (2008): Telephone Surveys. In: de Leeuw, Edith D./Hox, Joop J./Dillman, Don A. (Hrsg.): *International Handbook of Survey Methodology*. New York u. a.: Taylor & Francis: 221-238.
- Steeh, Charlotte/Kirgis, Nicole/Cannon, Brian/Dewitt, Jeff (2001): Are They Really as Bad as They Seem? Nonresponse Rates'at the End of the Twentieth Century. In: *Journal of Official Statistics* 17: 227-247.
- Stegmüller, Wolfgang (1983): *Erklärung, Begründung, Kausalität*. Berlin u. a.: Springer.
- Stevenson, Randolph T. (2002): The Cost of Ruling, Cabinet Duration, and the „Median-Gap” Model. In: *Public Choice* 113: 157-178.
- Stögbauer, Andrea (2000): Ausschöpfungsprobleme telefonischer Umfragen. Eine Zwischenbilanz praktischer gesamtdeutscher Erfahrung. In: Hüfken, Volker (Hrsg.): *Methoden in Telefonumfragen*. Wiesbaden: Westdeutscher Verlag: 91-103.
- Stuit, Dewey B./Crossley, Archibald M./Dodd, Stuart C./Bachelder, Joseph E./Gallup, George H./Kroeger, Henry J./Lazarsfeld, Paul F./Meier, Norman C. (1949): Causes of Disparity Between Poll Findings and Election Results and Discussion. In: Meier, Norman C./Saunders, Harold W. (Hrsg.): *The Polls and Public Opinion. The Iowa Conferene on Attitude and Opinion Research Sponsored by the State University of Iowa, Iowa City*. New York: Henry Holt: 159-199.
- Sudman, Seymour (1986): Do Exit Polls Influence Voting Behaviour? In: *Public Opinion Quarterly* 50: 331-339.

- Sunder, Shyam (1995): Experimental Asset Markets: A Survey. In: Kagel, John H./Roth, Alvin E. (Hrsg.): Handbook of Experimental Economics. Princeton: Princeton University Press: 445-500.
- Sunstein, Cass R. (2006): Infotopia. How Many Minds Produce Knowledge. Oxford u. a.: Oxford University Press.
- Surowiecki, James (2005): Die Weisheit der Vielen. Warum Gruppen klüger sind als Einzelne und wie wir das kollektive Wissen für unser wirtschaftliches, soziales und politisches Handeln nutzen können. München: C. Bertelsmann.
- Tajfel, Henri (1981): Human Groups and Social Categories: Studies in Social Psychology. Cambridge u. a.: Cambridge University Press.
- Thiemer, Horst (1973): Demagogie statt Demoskopie? In: Zeitschrift für Markt-, Meinungs- und Zukunftsforschung 16: 3111-3116.
- Thurner, Paul W./Pappi, Franz U. (1998): Measuring and Explaining Strategic Voting in the German Electoral System. Arbeitspapier Nr. 21. Mannheim: Mannheimer Zentrum für Europäische Sozialforschung. Online: <http://www.mzes.uni-mannheim.de/publications/wp/wp2-21.pdf> (10. November 2008).
- TNS Emnid (2009): Chronik – Emnid 1945 bis heute. Bielefeld: TNS Emnid. Online: http://www.tns-emnid.com/das_unternehmen/chronik.asp (28. Januar 2009).
- Todd, Peter M./Rieskamp, Jörg/Gigerenzer, Gerd (2008): Social Heuristics. In: Plott, Charles R./Smith, Vernon L. (Hrsg.): Social Heuristics. Amsterdam: North-Holland: 1035-1046.
- Tortora, Robert D. (2004): Response Trends in a National Random Digital Dialing Survey. In: Metodoloski zvezki 1: 21-32.
- Toulmin, Stephen E. (1968): Voraussicht und Verstehen. Ein Versuch über die Ziele der Wissenschaft. Frankfurt am Main: Suhrkamp.
- Traugott, Michael W. (2001): Assessing Poll Performance in the 2000 Campaign. In: Public Opinion Quarterly 65: 389-419.
- Traugott, Michael W. (2005): The Accuracy of the National Preelection Polls in the 2004 Presidential Election. In: Public Opinion Quarterly 69: 642-654.
- Tufte, Edward R. (1978): Political Control of the Economy. Princeton: Princeton University Press.
- Tziralis, Georgios/Tatsiopoulos, Ilias (2008): Prediction Markets: An Extended Literature Review. Athen: National Technical University of Athens. Online: http://gtziralis.googlepages.com/PredictionMarkets_AnExtendedLiterature-Review_TziralisTatsiopoulos.pdf (10. November 2008).
- Ulmer, Fritz (1987a): Wahlprognosen und Meinungsumfragen. Der Orakelspruch mit dem repräsentativen Querschnitt. In: Zum Nachdenken 15: 3-17.
- Ulmer, Fritz (1987b): Wahlprognosen und Meinungsumfragen und der Absatzhandel mit Prozentzahlen. In: Zeitschrift für Markt-, Meinungs- und Zukunftsforschung 30: 6729-6875.

- Ulmer, Fritz (o.J.): Grabmahl für die unbekannte Fehlprognose. Wuppertal. Online: <http://www.wahlprognosen-info.de/index2.htm?inhalt.htm> (13. September 2006).
- Urban, Dieter/Mayerl, Jochen (2007): Antwortlatenzzeiten in der survey-basierenden Verhaltensforschung. In: *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 59: 692-713.
- van Der Eijk, Cees (2005): Election Forecasting: A Sceptical View. In: *British Journal of Politics and International Relations* 7: 210-214.
- Verbeek, Marno (2008): *A Guide to Modern Econometrics*. 3. Aufl. Chichester u. a.: Wiley.
- Volosin, Boris A./Orlov, Jurij K. (1982): Das verallgemeinerte Zipf-Mandelbrotsche Gesetz und die Verteilung der Anteile von Farbflächen in der Malerei. In: Orlov, Jurij K./Boroda, Moisei G./Nadarejvili, Isabella S. (Hrsg.): *Sprache, Text, Kunst. Quantitative Analysen*. Bochum: Brockmeyer.
- von der Heyde, Christian (2002): Probleme und Entwicklungstrends der Umfrageforschung. Stichprobenverfahren in der Umfrageforschung. In: Sahner, Heinz (Hrsg.): *Fünfzig Jahre nach Weinheim. Empirische Markt- und Sozialforschung gestern, heute, morgen*. Baden-Baden: Nomos: 141-148.
- von Friedeburg, Ludwig (1961): Zum politischen Potential der Umfrageforschung. In: *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 13: 201-216.
- von Knorring, Ekkehard (1972): Probleme der theoretischen Deduktion von Prognosen. Eine Replik. In: *Zeitschrift für die gesamte Staatswissenschaft* 128: 145-148.
- Voss, D. Stephen/Gelman, Andrew/King, Gary (1995): The Polls, a Review: Pre-election Survey Methodology: Details from Eight Polling Organizations, 1988 and 1992. In: *Public Opinion Quarterly* 59: 98-132.
- Wendt, Friedrich (1960): Wann wird das Quotenverfahren begraben? In: *Allgemeines statistisches Archiv* 44: 35-40.
- Weßels, Bernhard (2009): Splitting sichert den Wechsel. Mit taktischem Wahlverhalten verhindert der Bürger Schwarz-Rot. *WZB-Mitteilungen* 126. Berlin: Wissenschaftszentrum Berlin. Online: <http://www.wzb.eu/publikation/pdf/wm126/33-37.pdf> (13. Januar 2010)
- West, Darrell M. (1991): Polling Effects in Election Campaigns. In: *Political Behavior* 13: 151-164.
- Whiteley, Paul F. (1979): Electoral Forecasting from Poll Data: the British Case. In: *British Journal of Political Science* 9: 219-236.
- Whiteley, Paul F. (2005): Forecasting Seats from Votes in British General Elections. In: *British Journal of Politics and International Relations* 7: 165-173.
- Wild, Jürgen (1970): Probleme der theoretischen Deduktion von Prognosen. In: *Zeitschrift für die gesamte Staatswissenschaft* 126: 553-576.

- Wildenmann, Rudolf (1992): *Wahlforschung*. Mannheim u. a.: B.I.-Taschenbuchverlag.
- Witterauf, Peter (2008): Steigende Unsicherheit bei Wahlprognosen – wie damit umgehen? In: *Politische Studien* 59: 14-17.
- Wlezien, Christopher/Erikson, Robert S. (1996): Temporal Horizons and Presidential Election Forecasts. In: *American Politics Quarterly* 24: 492-505.
- Wolf, Klaus (2007): *Meinungsforschungsinstitute in der Wahlumfrageberichterstattung. Eine Analyse überregionaler Qualitätszeitungen vor den Bundestagswahlen 1980-2005*. München: GRIN.
- Wolfers, Justin/Leigh, Andrew (2002): Three Tools for Forecasting Federal Elections: Lessons from 2001. In: *Australian Journal of Political Science* 37: 223-240.
- Wolfers, Justin/Zitzewitz, Eric (2004): Prediction Markets. In: *Journal of Economic Perspectives* 18: 107-126.
- Wolfers, Justin/Zitzewitz, Eric (2006): *Prediction Markets in Theory and Practice*. IZA Discussion Paper. Bonn: Forschungsinstitut zur Zukunft der Arbeit. Online: <http://ftp.iza.org/dp1991.pdf> (3. Oktober 2008).
- Woodward, James (2003): *Making Things Happen. A Theory of Causal Explanation*. Oxford u. a.: Oxford University Press.
- Worcester, Robert M. (1983): *Political Opinion Polling. An International Review*. New York: St. Martin's Press.
- Woyke, Wichard (2005): *Stichwort: Wahlen. Ein Ratgeber für Wähler, Wahlhelfer und Kandidaten*. 11. Aufl. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Wüst, Andreas M. (1998): Wahl- und Umfragesforschung im Bundestagswahljahr 1998 – mehr als kühne Prognosen? In: *Gegenwartskunde: Zeitschrift für Gesellschaft, Wirtschaft, Politik und Bildung* 47: 323-333.
- Wüst, Andreas M. (2003): Stimmung, Projektion, Prognose? In: Wüst, Andreas M. (Hrsg.): *Politbarometer*. Opladen: Leske + Budrich: 83-108.
- Yu, Philip L. H./Lam, Eddy K. F. (1997): How to Predict Election Winners from a Poll. In: *Journal of Applied Statistics* 24: 11-23.
- Zanette, Damián (2008): Zipf's Law and the Creation of Musical Context. Río Negro: Instituto Balseiro. Online: http://arxiv.org/PS_cache/cs/pdf/0406/0406015v1.pdf (23. Februar 2009).
- Zipf, George K. (1965): *Human Behavior and the Principle of Least Effort. An Introduction to Human Ecology*. Nachdruck. New York u. a.: Hafner.
- Zukin, Cliff (2004): *Sources of Variation in Published Election Polling: A Primer*. New Brunswick: Rutgers University. Online: <http://www.princeton.edu/~ccameron/KoreaIIE/IIE337/Zukin.Primers.2006.pdf> (10. November 2008).