

Kapitel 1

Dieter Ohr und Hans Rattinger: Zur Beziehung zwischen in der Vorwahlzeit erhobenen Wahlabsichten und Wahlergebnissen

1.1 Einleitung

Berichte über Ergebnisse der sogenannten „Sonntagsfrage“ („Wenn am nächsten Sonntag Bundestagswahl wäre, welche Partei würden Sie wählen?“)¹ sind seit Jahren fester Bestandteil des öffentlichen Vorlaufs von Wahlen. Politische Parteien, Presse, Fernsehen und Kommentatoren widmen ihnen große Aufmerksamkeit. Dabei gedenkt längere Zeit vor einer Wahl kaum jemand, diese Wahl durch die Sonntagsfrage ernsthaft zu „prognostizieren“. Trotzdem wird sie regelmäßig gestellt, auch dann, wenn die nächste Bundestagswahl noch Jahre entfernt ist. Bei einem größeren zeitlichen Abstand ist dann oft vom „Meinungsklima“, der „aktuellen politischen Stimmung“ oder einer „Wahlprojektion“ die Rede. Man beabsichtigt also lange vor einer Wahl — bestenfalls — nur anzugeben, wie gewählt würde, *wenn* zu diesem Zeitpunkt gewählt würde, was aber in Wirklichkeit nicht der Fall ist, so daß ein objektiver Maßstab für die Genauigkeit solcher Aussagen fehlt.

Da die Wahlabsicht bei Stichproben von in der Regel etwa 1000 bis 2000 Wahlberechtigten erhoben wird, besteht natürlich die Möglichkeit zufälliger Schwankungen. Dabei ist die Beziehung zwischen Stichprobenwerten und den Werten für die gesamte Bevölkerung einfach: Man ermittelt die klassischen Konfidenzintervalle mit den üblichen Fehlerwahrscheinlichkeiten. Zusätzlich zu den Punktschätzungen für die Stimmenanteile der Parteien werden oft diese

¹Neben der „Sonntagsfrage“ wird die Wahlabsicht auch oft über eine Wahlsimulation erhoben, bei der die Befragten sich durch Ankreuzen auf einem Stimmzettel für eine Partei entscheiden.

Konfidenzintervalle angegeben.² Die „Populationswerte“ sind aber gar nicht existent, da ja zum Befragungszeitpunkt gerade *nicht* gewählt wird. Man könnte natürlich einwenden, so würde die Bevölkerung *derzeit* (innerhalb eines bestimmten Vertrauensintervalls) wählen. Nur, überprüfen läßt sich diese Vermutung nicht einmal prinzipiell, weil die Wahl eben nicht zum Befragungszeitpunkt, sondern erst später stattfindet. So gesehen könnte die Berechnung von Konfidenzintervallen als unverbindlich erscheinen, weil kein objektiver *Referenzmaßstab* vorhanden ist. Der Informationsgehalt eines Vertrauensintervalls um die aus der Stichprobe geschätzte „aktuelle Stimmung“ könnte als nur sehr begrenzt angesehen werden. Dem läßt sich entgegen, daß derartige Vertrauensintervalle sehr wohl auch relevante Information enthalten, etwa ob eine Partei in der aktuellen Stimmung wirklich signifikant vor einer anderen oder über oder unter der Sperrklausel liegt.

Wesentlich interessanter, nicht zuletzt für die Parteien selbst, könnte dagegen der Grad der Annäherung zwischen der jeweiligen „aktuellen Stimmungslage“ und dem späteren Wahlergebnis sein. So läßt sich vermuten, daß diese Annäherung immer stärker wird, je näher die Wahl rückt. Eine Reihe von in diesem Aufsatz noch zu diskutierenden Faktoren spricht dafür, daß die Übereinstimmung von projiziertem und tatsächlichem Wahlergebnis maßgeblich von der *zeitlichen Distanz zum Wahltermin* abhängt. Nimmt man also den einzig verfügbaren objektiven Referenzmaßstab, nämlich das Ergebnis der folgenden Wahl, so kann man, wie noch zu zeigen sein wird, die zeitliche Dimension der Beziehung zwischen Wahlabsichten und Wahlergebnis als Grundlage für ein alternatives, *zeitabhängiges Vertrauensintervall* heranziehen.

Solche „Projektionen“ aufgrund der Sonntagsfrage sind meist *Gewichtungen* der ursprünglichen Resultate einer Meinungsumfrage. Bringen diese Gewichtungen die Verteilung der Wahlabsichten zuverlässig näher an das tatsächliche Wahlergebnis? Um diese Frage beantworten zu können, muß zuerst untersucht werden, wie die Abweichung der gewichteten Wahlabsichten vom späteren Wahlergebnis sich mit dem zeitlichen Abstand vom Wahltag verändert. Dann ist diese zeitliche Assoziation zu vergleichen mit den entsprechenden Abweichungen der ungewichteten Wahlabsichten vom Endergebnis.

Wenn wir in diesem Aufsatz das Ergebnis der folgenden Wahl als Referenz für Wahlabsichten betrachten, unabhängig davon, ob von den jeweiligen Urhebern irgendein prognostischer Anspruch erhoben wird oder nicht, so deshalb, weil dadurch die Analyse einer theoretisch, empirisch sowie für die Praxis re-

²Auf die Problematik des Zufallsfehlers für die Erstellung kategorialer „Prognosen“ (Kommt Partei A über die 5 %-Hürde, schafft Partei B die absolute Mehrheit?) weist Ulmer (1987: 96-97) hin. Hat etwa eine Partei in der Grundgesamtheit aller Wahlberechtigten einen Anteil von 4,5 Prozent und beträgt der Stichprobenumfang 2000, so werden ca. 15 Prozent aller denkbaren Stichproben (fälschlicherweise) ein Überschreiten der 5 %-Hürde anzeigen.

levanten Fragestellung möglich wird: Mit welcher Zuverlässigkeit kann eine zu einem bestimmten Zeitpunkt erstellte „Projektion“ aufgrund der Sonntagsfrage Aufschlüsse über das spätere Wahlergebnis liefern? Wie groß kann unser Vertrauen sein, daß eine solche „Projektion“ beispielsweise fünf Monate vor dem späteren Wahltermin das Wahlergebnis mit einer bestimmten Genauigkeit annähert? Dabei soll nochmals betont werden, daß lange vor einer Wahl Wahlabsichtserhebungen nach der Intention ihrer Urheber den späteren Wahlausgang *nicht* vorhersagen sollen (und wohl auch nicht können).

Andererseits spricht unseres Erachtens vieles dafür, nicht nur solche Erhebungen mit dem Wahlergebnis zu vergleichen, die wenige Tage vor dem Wahltermin durchgeführt wurden. Wir haben uns entschieden, die Grenze bei einem zeitlichen Abstand vom Wahltermin von einem Jahr zu ziehen. Nicht zuletzt aufgrund der im Vorwahljahr stetig wachsenden Berichterstattung in den Medien rückt die Wahl dann immer stärker in das öffentliche Interesse. Ebenso werden die Wahlabsichten der Bevölkerung im Vorwahljahr mit steigender Frequenz erhoben und in der Presse immer mehr im Hinblick auf die nahende Wahl interpretiert. „Projektionen“ und „Meinungsklima“ werden damit zunehmend nicht nur lediglich aktuelle Stimmungen reflektieren, sondern immer näher an das tatsächliche spätere Wahlverhalten der Bevölkerung heranrücken. Um die zeitliche Dimension dieses Prozesses geht es in diesem Aufsatz.

Um Fehlinterpretationen zu vermeiden: Diese Problemstellung ist nicht notwendigerweise diejenige der Institute, die regelmäßig Wahlabsichten erheben. Nach eigenem Verständnis erarbeiten sie damit Befunde über Trends aktueller „Stimmungslagen“. Wenn *wir* diese Befunde in Beziehung setzen zu späteren Wahlergebnissen und dabei finden, daß die Abweichungen mit dem zeitlichen Abstand zunehmen, darf dies nicht als Kritik mißverstanden werden an der Leistungsfähigkeit solcher Institute oder ihrer regelmäßigen Erhebung von Wahlabsichten. Wer (sinnvollerweise) nicht vorgibt, mit dem Stimmungsreport auch eine „Wahlprognose“ abzuliefern, kann kaum wegen grober Prognosefehler attackiert oder wegen genauer Vorhersagen gefeiert werden. Der (sinnvolle) Verzicht auf prognostischen Anspruch, wenn eine Wahl gar nicht ansteht, macht auch die regelmäßige Abfrage von Wahlabsichten keinesfalls irrelevant oder überflüssig — im Gegenteil: Die Meinungsforschungsinstitute erfüllen damit eine wichtige Dienstleistung für die Öffentlichkeit und die Wissenschaft. Die analytischen Problemstellungen dieses Aufsatzes etwa wären ohne das derart kontinuierlich erhobene Datenmaterial überhaupt nicht bearbeitbar.

Zwei Fragestellungen werden also in diesem Aufsatz untersucht: Was läßt sich auf der Grundlage verfügbarer Daten über die *zeitliche Dimension der Beziehung zwischen Wahlabsichten und -ergebnissen* sagen? Wichtig ist hierbei sowohl die Stärke wie auch die funktionale Form dieser Assoziation. Unsere zweite Fragestellung baut auf der Klärung der ersten Frage auf: Was ergibt

sich aus der Kenntnis dieser zeitlichen Abhängigkeit für das Vertrauen in die Annäherung von späteren Wahlergebnissen durch Wahlabsichten aus Umfragedaten? Zunächst wird jedoch im folgenden Abschnitt kurz die Praxis der Erarbeitung derartiger „Wahlprojektionen“ etc. dargestellt.

1.2 Zum Verfahren der Gewichtung von in Umfragen erhobenen Wahlabsichten

Datengrundlage zur Erstellung von „Wahlprojektionen“ ist in der Regel ein demographisch gewichteter Umfragedatensatz³, wobei die demographische Gewichtung meist zumindest mit den Merkmalen Alter und Geschlecht, teilweise auch mit der Ortsgröße und dem Bundesland durchgeführt wird. Eine Schwierigkeit ergibt sich durch die Antwortverweigerung von Befragten und durch „Unentschiedene“, die sich bei der Befragung nicht für eine Partei entscheiden. Erfahrungsgemäß fallen so bis zu 20 Prozent der Befragten weg. Man behilft sich dabei oft mit der Annahme, daß sich „Unentschiedene“ und „Verweigerer“ wie die Befragten mit Angabe einer Wahlabsicht auf die Parteien verteilen, rechnet also einfach mit den angegebenen Wahlabsichten weiter. Diese Annahme kann natürlich falsch sein und stellt in jedem Fall eine potentielle Fehlerquelle dar (s. dazu Rattinger/Ohr 1989). Alternativen zu dieser Behandlung der Unentschiedenen und Verweigerer liegen in der Ersetzung ihrer Wahlabsichten durch die angegebene Parteiidentifikation oder in der Zuordnung dieser Gruppe zu den Parteien anhand verschiedener Merkmale wie der Parteisymphathie, der „liebsten“ Partei, Positionen und Kompetenzeinschätzungen bei Sachfragen oder der sozialstrukturellen Verortung des Befragten.⁴

Die auf die einzelnen Parteien entfallenden Prozentanteile werden dann oft einer sogenannten *politischen Gewichtung* unterzogen. Nur „politisch gewichtete“ Ergebnisse werden aus guten Gründen in der Regel veröffentlicht. In der Praxis haben sich mit den beiden Varianten der sogenannten *Rückeringungsgewichtung*, dem *Quotientenverfahren* und dem *Differenzenverfahren*, zwei Formen der politischen Gewichtung durchgesetzt. Des öfteren werden auch Mischverfahren (etwa Anwendung des Differenzenverfahrens für die kleineren Parteien, Anwendung des Quotientenverfahrens für die großen Parteien)

³Die zu befragenden Wahlberechtigten werden in der letzten Stufe eines grundsätzlich mehrstufigen Auswahlverfahrens (zuerst Auswahl der sample points, zuletzt Auswahl der Befragten) in der Praxis vor allem über den sogenannten *Random Walk* oder das *Quotenverfahren* ermittelt. Zunehmend werden auch *Telefoninterviews* eingesetzt (zu einem Überblick über diese Verfahren vgl. Falter/Schumann 1989: 5).

⁴Das Vorgehen bei der Zuordnung von Wahlberechtigten ohne Wahlabsicht zu den Parteien mittels der *Parteisymphathie* ist in Rattinger/Ohr (1989: 289) dargestellt, zu den weiteren Zuordnungsmethoden siehe Hofmann (1976: 6-7).

eingesetzt. Manchmal werden auch die auf diese Weise politisch gewichteten Resultate noch nicht publiziert, sondern es erfolgt eine nochmalige Korrektur, die sich im wesentlichen auf Erfahrung und Fingerspitzengefühl des Analytikers gründet.⁵ Solche Mischverfahren oder „Feinkorrekturen“ bleiben im folgenden unberücksichtigt.

Die den Verfahren der Rückerinnerungsgewichtung zugrundeliegende Logik wurde an anderer Stelle ausführlich dargestellt, ebenso Möglichkeiten zur Korrektur von Schwächen einiger in der Praxis eingesetzter Rückerinnerungsgewichtungen (Rattinger/Ohr 1989). Wir können uns deshalb hier kurz fassen: Zuerst wird in der konkreten Stichprobe die Erinnerung an das Wahlverhalten bei der letzten Wahl ermittelt. Dieses erinnerte Wahlergebnis wird dann dem tatsächlichen Ergebnis dieser Wahl gegenübergestellt. Die in der Umfrage ermittelten Wahlabsichtsanteile für die Parteien werden nun entweder multipliziert mit dem Verhältnis aus amtlichem Vorwahlergebnis und Rückerinnerungswahlergebnis (Quotientenverfahren), oder man addiert die Differenz der beiden Größen zu den Wahlabsichtsanteilen der Umfrage (Differenzenverfahren). Der Verfahrensweise der Rückerinnerungsgewichtung liegt also die Überlegung zugrunde, daß eine retrospektive Überschätzung (Unterschätzung) einer bestimmten Partei einhergeht mit einer Überschätzung (Unterschätzung) dieser Partei in der Stichprobe im Hinblick auf die aktuelle Stimmung. Genau dies soll durch die Rückerinnerungsgewichtung behoben werden. Tab. 1.1 stellt anhand einer exemplarisch ausgewählten Umfrage diese *einfachsten* Methoden der politischen Gewichtung dar und zeigt ferner mit den parteiweisen Fehlern sowie dem mittleren absoluten Fehler einige einfache Maße für die Übereinstimmung zwischen den ungewichteten beziehungsweise gewichteten Wahlabsichten und dem späteren Wahlergebnis.

Wie auch das Beispiel in Tab. 1.1 zeigt, spricht die bislang vorliegende Evidenz sehr überwältigend für den Einsatz von Rückerinnerungsgewichtungen (Rattinger/Ohr 1989). In der Praxis existieren viele komplexere, institutspezifische Gewichtungsformen.⁶ Einige einfache Varianten der in der Praxis

⁵Die nochmalige Korrektur der schon politisch gewichteten Ergebnisse berücksichtigt zum Beispiel die Popularität der Spitzenkandidaten oder die Kompetenzeinschätzungen der Parteien bei verschiedenen Sachfragen in der Umfrage. Erscheint ein Parteienanteil auch nach der Rückerinnerungsgewichtung vor dem Hintergrund dieser weiteren Befunde als „zu hoch“ oder „zu niedrig“, erfolgt eine weitere Anpassung.

⁶Zwei weitere Verfahren der politischen Gewichtung nehmen die absoluten Häufigkeiten einer Kreuztabellierung von Rückerinnerungsfrage und Wahlabsichtsfrage zum Ausgangspunkt. Das erste Verfahren wendet die Gewichtungsfaktoren des Quotientenverfahrens getrennt für jede Partei der Rückerinnerungsfrage auf die absoluten Häufigkeiten der Kreuztabelle an und ermittelt dann mit den so korrigierten Häufigkeiten die Anteile für die „Projektion“. Im zweiten Verfahren werden die aus dem Vergleich von amtlichem Vorwahlergebnis und Rückerinnerungsfrage gewonnenen Gewichtungsfaktoren durch *Übergangswahrscheinlichkeiten* ersetzt, die sich aus der Kreuztabellierung von Rückerinnerungsfrage und Wahl-

Tabelle 1.1: Quotienten- und Differenzenverfahren der Rückerinnerungsgewichtung am Beispiel der ersten Welle einer Panel-Studie zur Bundestagswahl 1987 (September 1986, n=1954, ZA-Nr.1537)

a) Ablauf der Berechnung

	Demographisch gewichtete Wahlabsicht	Wahlabsicht ohne Verweigerer/Unentsch.	Rückerinnerung 1983	Rückerinnerung 1983 ohne Verweigerer/Nichtwähler	Amtliches Wahlergebnis 1983 (Zweitst.)
	(1)	(2)	(3)	(4)	(5)
CDU/CSU	38,1	43,0	40,0	45,6	48,8
SPD	38,7	43,7	38,0	43,3	38,2
FDP	3,8	4,3	3,7	4,3	7,0
GRÜNE	7,0	7,9	5,0	5,7	5,6
Sonstige	0,9	1,1	1,0	1,1	0,4
Verw./Unentsch.	11,4	—	Verw./NW 12,2	—	—
Summe	100,0	100,0	100,0	100,0	100,0

	Quotient (5/4)	Differenz (5-4)
CDU/CSU	1,070	3,2
SPD	0,882	-5,1
FDP	1,628	2,7
GRÜNE	0,982	-0,1
Sonstige	0,364	-0,7
Summe	—	0,0

	Wahlabsicht(2) *Quotient	Wahlabsicht(2) *Quotient Summe = 100	Wahlabsicht(2) +Differenz	Wahlergebnis 1987
CDU/CSU	46,01	46,1	46,2	44,3
SPD	38,54	38,7	38,6	37,0
FDP	7,00	7,0	7,0	9,1
GRÜNE	7,76	7,8	7,8	8,3
Sonstige	0,40	0,4	0,4	1,3
Summe	99,71	100,0	100,0	100,0

b) Fehler

	Ungewichtete Wahlabsichten	Quotientenverfahren	Differenzenverfahren
CDU/CSU	-1,3	1,8	1,9
SPD	6,7	1,7	1,6
FDP	-4,8	-2,1	-2,1
Grüne	-0,4	-0,5	-0,5
Sonstige	-0,2	-0,9	-0,9
Mittlerer absoluter Fehler		2,7	1,4

eingesetzten Quotienten- und Differenzenmodelle haben zwei offensichtliche Schwächen, welche die Korrekturwirkung mindern, die aber beide überwunden werden können (vgl. dazu Rattinger/Ohr 1989). Zum einen ignoriert die „einfache“ Rückerinnerungsgewichtung das Faktum, daß zwischen Vorwahlzeitpunkt und Umfragezeitpunkt ein Teil des Elektorats verstorben ist. Selbst bei Abwesenheit sonstiger Verzerrungsmöglichkeiten (wie unwahrhaftigen Angaben, Stichprobenfehlern etc.) werden die befragten „Wiederholungswähler“ das Vorwahlergebnis nicht genau reproduzieren können, da die verschiedenen Altersgruppen selektiv von Mortalität betroffen sind und das Wahlverhalten eben auch vom Alter abhängt. Das erinnerte Vorwahlergebnis der Wiederholungswähler wird sich schon allein aus diesem Grund vom tatsächlichen Vorwahlergebnis unterscheiden. Zweitens behandeln manche Rückerinnerungskorrekturen die Wahlabsichten der Erstwähler wie die der Wiederholungswähler, verwenden also die gleichen Gewichtungsfaktoren, obwohl diese Faktoren ausschließlich aus dem amtlichen Vorwahlergebnis und den Rückerinnerungen der „Wiederholungswähler“ entstanden sind.⁷ Ein daraus eventuell entstehender Fehler wird um so größer sein, je höher der Erstwähleranteil ist.

1.3 Die Abhängigkeit der Abweichungen zwischen Wahlabsichten und Wahlergebnis vom zeitlichen Abstand zum Wahltag

1.3.1 Vorbemerkung

Beide Fehlerquellen (wie die korrespondierenden Lösungsmöglichkeiten) werden hier im folgenden jedoch ausgeblendet. Wir nehmen nur die einfachsten, globalen Quotienten- und Differenzenkorrekturen, das heißt, auch die Wahlabsichten der Erstwähler werden umgewichtet. Zwei Fragen sollen, wie schon eingangs erwähnt, untersucht werden: Sind Rückerinnerungsgewichtungen zu unterschiedlichen zeitlichen Abständen vom Wahltag verschieden effektiv, wenn man als Referenz das tatsächliche (spätere) Wahlergebnis nimmt?

absichtsfrage ableiten lassen.

Rückerinnerungsgewichtung kann ferner auch auf der *Individualebene* mit personenbezogenen Gewichten durchgeführt werden. Wenn diese personenbezogenen Gewichtungsfaktoren aus dem dargestellten Vergleich von amtlichen Stimmanteilen der Vorwahl und den Anteilen der Rückerinnerungsfrage gewonnen werden (vgl. Hentschel/Tennstädt 1985: 37-39), ergibt sich exakt die gleiche „Prognose“ wie bei der allgemein üblichen Gewichtung der aggregierten Wahlabsichten.

⁷Manche Institute schließen die Erstwähler von der Umgewichtung aus und gehen somit anders vor, als dies in Tab. 1.1 dargestellt wurde (vgl. Hofmann 1976: 7).

Was folgt daraus für die Bestimmung von Konfidenzintervallen, wobei wieder das spätere Wahlergebnis und nicht die aktuelle „Stimmung“ der wahlberechtigten Bevölkerung unser Referenzpunkt ist?⁸

Für unsere Fragestellungen könnte eine weitere potentielle Fehlerquelle der Rückerinnerungsmodelle von Bedeutung sein. Quotientenverfahren und Differenzenverfahren unterstellen eine identische (oder zumindest nicht allzu unterschiedliche) Verzerrung von Rückerinnerung und Wahlabsicht. Im Idealfall einer völlig identischen Verzerrung könnten die wahren aktuellen Wahlabsichten perfekt ermittelt werden. Sollen Rückerinnerungsmodelle zufriedenstellend funktionieren, muß sich die Verzerrung von Rückerinnerung und Wahlabsicht möglichst weitgehend entsprechen. Übertragen auf unsere folgende Analyse zeitlicher Abhängigkeiten bedeutet dies, daß die Rückerinnerungsverzerrung in etwa mit dem zeitlichen Verlaufsmuster der Wahlabsichtsverzerrung übereinstimmen muß. Wäre dies nicht der Fall, würden beispielsweise die Befragten kurz vor dem Wahltag ihre Wahlabsicht korrekt äußern, jedoch ihr Wahlverhalten bei der letzten Wahl zum Teil nicht korrekt wiedergeben, würden die Verfahren der Rückerinnerungsgewichtung bei geringer zeitlicher Distanz zum Wahltermin zu einer Verschlechterung der Anpassung an das Wahlergebnis führen.

Wenn man die Auswirkungen der zeitlichen Distanz zum Wahltag untersuchen will, benötigt man eine größere Anzahl von Wahlabsichtsdaten mit verschiedenen zeitlichen Abständen zum Wahltermin. Dafür steht uns ein Datensatz zur Verfügung, dessen Untersuchungseinheiten einzelne Umfragen sind und der insgesamt 229 Umfragestudien mit Wahlabsichten für die vier Bundestagswahlen 1976 bis 1987 enthält, ebenso die jeweiligen Rückerinnerungen und diverse technische Details der Studien (zur weiteren Charakterisierung s. Anhang 3). Dieser Datensatz wurde ergänzt durch 14 Wahlstudien der Forschungsgruppe Wahlen zur Bundestagswahl 1987. Aus diesen insgesamt 243 Studien wurden aus den oben genannten Gründen diejenigen ausgewählt, die einen zeitlichen Abstand von höchstens einem Jahr (genau: 400 Tagen) zum Wahltermin aufwiesen. Nach dieser zeitlichen Eingrenzung verbleiben 123 Studien, für die von uns jeweils drei verschiedene Gewichtungen der Wahlabsichten berechnet wurden: zuerst „rohe“, lediglich demographisch gewichtete Wahlabsichten, zweitens Umgewichtungen nach dem Quotientenverfahren, drittens nach dem Differenzenverfahren.

⁸Es ist also für unsere Untersuchung beispielsweise nicht von Interesse, ob sich Parteianteile von Erhebung zu Erhebung signifikant verändern. Für die politische Praxis kann genau darin das Hauptinteresse an kontinuierlicher Trendbeobachtung bestehen.

1.3.2 Ein erster Überblick

Im folgenden soll nach einem allgemeinen funktionalen Zusammenhang zwischen der Abweichung von Wahlabsichten und Wahlergebnis und der zeitlichen Distanz zum Wahltag gesucht werden. Dazu führen wir für die drei Gewichtungsverfahren jeweils Regressionsanalysen durch, wobei diese Abweichung die abhängige Variable und der zeitliche Abstand zum Wahltag die unabhängige Variable darstellt. Unser Maß für die Abweichung wird dabei ausschließlich der *mittlere* (Mittelung über die drei beziehungsweise vier Parteien und die „Sonstigen“) *absolute Fehler* sein. Der verwendete Datensatz enthält, wie erwähnt, Wahlabsichten und Fehlerwerte für drei Gewichtungsvarianten und für insgesamt vier Bundestagswahlen von 1976 bis 1987. Für den verfolgten Untersuchungszweck ist es erforderlich, etwaige *Niveaunterschiede* zwischen den mittleren Fehlern der vier Wahljahre zu eliminieren. Solche Niveaunterschiede gehen in erster Linie auf idiosynkratische Eigenschaften der einzelnen Wahlen zurück, die hier außer Betracht bleiben sollen. Niveaunterschiede behindern auch schon allein deswegen die Suche nach einem funktionalen Zusammenhang, weil je nach Wahljahr die Studien unterschiedlich auf der zeitlichen Abstandsachse verteilt sind.

Es gibt mindestens drei Möglichkeiten der Berücksichtigung von Niveaunterschieden der Fehler zwischen Wahlen: einmal die Durchführung einer wahlweisen Analyse, zweitens die Repräsentation der einzelnen Wahljahre durch Dummy-Variablen, drittens die Ausschaltung von Niveaunterschieden mittels Gewichtung. Wir haben uns aus Gründen einer anschaulicheren und sparsameren Ergebnispräsentation für die letzte Möglichkeit entschieden.⁹ Dabei werden die (mittleren) Fehler der einzelnen Studien mit dem Quotienten aus modellspezifischem Gesamtmittelwert und modell- und jahresweisem Mittelwert gewichtet. Die jahres- und modellspezifischen Mittelwerte werden so dem modellspezifischen Gesamtmittelwert angeglichen. Dieses Verfahren beseitigt die Niveaunterschiede zwischen den vier Wahlen und läßt dabei das Verhältnis zwischen den Abweichungen zweier Studien innerhalb eines Wahljahres vom jeweiligen Wahlergebnis konstant.

Tab. 1.2 gibt einen ersten Überblick sowohl über die Leistungsfähigkeit der drei Gewichtungsvarianten als auch über die Zeitabhängigkeit der Diskrepanzen zwischen Wahlabsichten und -ergebnissen.

Die letzte Zeile von Tab. 1.2 zeigt mit dem modellspezifischen Mittel aller einbezogenen Studien den grundsätzlichen Effekt der Rückerinnerungsgewichtung. Demnach erreicht sie nach dem Quotientenverfahren eine geringe Reduktion des mittleren Fehlers, nach dem Differenzenverfahren eine etwas stärkere

⁹Im Hinblick auf die weiter unten durchgeführte Regression der Fehlerwerte auf den zeitlichen Abstand entspricht die Varianzerklärung nach der Gewichtungsvariante im wesentlichen der zu den Dummy-Variablen *zusätzlichen* Varianzerklärung der zeitlichen Abstandsvariable.

Tabelle 1.2: Mittlere absolute Abweichung zwischen Wahlabsichten und Wahlergebnis nach Gewichtungsverfahren und zeitlichem Abstand zum Wahltag (n=123)

	Demographische Gewichtung	Quotienten- verfahren	Differenzen- verfahren
Abstand in Tagen			
0-30	2,24	1,84	1,40
31-60	2,73	1,57	1,19
61-90	2,84	1,89	1,52
91-120	3,05	1,60	1,40
121-150	3,52	2,16	1,55
151-180	2,92	2,85	2,19
180-270	2,79	3,01	2,91
271-400	2,16	2,83	3,01
Mittel	2,68	2,33	2,07

Verbesserung. Die Mittelung über alle 123 Studien verdeckt jedoch wichtige Unterschiede im Hinblick auf die Effektivität der Rückerinnerungsgewichtung. In einem „mittleren“ Distanzbereich von einem bis zu fünf Monaten (30 bis 150 Tagen) gibt es stets eine deutliche Verbesserung um mindestens einen Prozentpunkt, teilweise (vor allem beim Differenzenverfahren) sogar eine Verbesserung bis zu zwei Prozentpunkten. Weniger ausgeprägt, beziehungsweise zum Teil überhaupt nicht mehr vorhanden, ist der Reduktionseffekt dagegen bei großen zeitlichen Abständen: Ab einer Distanz von sechs Monaten gibt es im Mittel nur noch Verschlechterungen der Anpassung der Wahlabsichten an das Wahlergebnis durch den Einsatz der Rückerinnerungsgewichtung.

Zur genaueren Bestimmung dieses differentiellen Effektes der Rückerinnerungsgewichtung wollen wir nun die funktionale Form des Zusammenhangs zwischen Fehlern und zeitlichem Abstand bei dem Verfahren ohne Rückerinnerungsgewichtung sowie den beiden Varianten mit Rückerinnerungsgewichtung näher untersuchen.

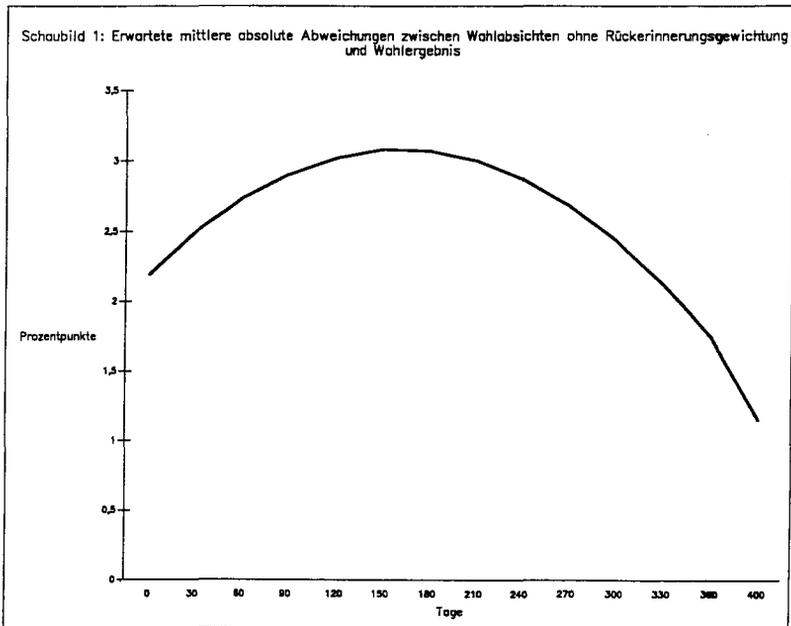
1.3.3 Detailanalyse

1.3.3.1 Verfahren ohne Rückerinnerungsgewichtung

Schon bei Betrachtung nur der Werte in Tab. 1.2 zeigt sich, daß „rohe“ Wahlabsichten dann am genauesten sind, wenn die Umfrage kurz vor der Wahl

oder aber in relativ großer zeitlicher Entfernung vom Wahltag stattfindet. Am schlechtesten ist das Grundmodell bei mittleren zeitlichen Distanzen. Auch die Regression des mittleren absoluten Fehlers auf den zeitlichen Abstand zum Wahltermin bestätigt diese Form der Nichtlinearität: Eine Modellspezifikation mit einer Kombination aus linearer Zeitvariable (gemessen in Tagen) und quadratischem Term ergibt mit einem R^2 von 0,15 im Vergleich zu vielen denkbaren funktionalen Verläufen die beste Datenanpassung. Schaubild Tab. 1.1 zeigt die aufgrund dieser Regressionssschätzung zu den jeweiligen zeitlichen Abständen erwarteten mittleren absoluten Abweichungen des Grundmodells vom Wahlergebnis.

Abbildung 1.1: Erwartete mittlere absolute Abweichungen zwischen Wahlabsichten ohne Rückerinnerungsgewichtung und Wahlergebnis



Auch wenn man in Rechnung stellt, daß die Datenanpassung nicht besonders gut ist und sich deswegen insbesondere bei großen zeitlichen Abständen vom Wahltag zum Teil wenig plausible Schätzungen für die Erwartungswerte errechnen lassen, bleibt doch auch bei der Regressionsanalyse das Grundmuster erhalten. Wie läßt sich dieser zunächst unerwartete Verlauf der Fehler bei dem Verfahren ohne Rückerinnerungsgewichtung theoretisch plausibilisieren? Man wird nach verschiedenen neueren Arbeiten der empirischen Wahlfor-

schung zur Frage der relativen Einflüsse langfristiger und kurzfristiger Faktoren auf das Wählerverhalten insgesamt von einem Übergewicht der Langfristfaktoren ausgehen können (Falter/Rattinger 1983, 1986; Jagodzinski/Kühnel 1990). Dabei kann in diesem Zusammenhang die theoretische Konzeptualisierung von Langfristfaktoren außer acht bleiben. Es muß hier nicht entschieden werden, ob Langfristfaktoren primär sozialstrukturell abgeleitet sind, ob Parteiidentifikation als affektive Bindung im Sinne des sozialpsychologischen Ansatzes der Wahlforschung zur theoretischen Fundierung vorzuziehen ist oder ob schließlich drittens Langfristfaktoren am besten im Rahmen eines „ökonomischen“ Erklärungsansatzes gefaßt werden können. Wichtig in diesem Kontext sind vielmehr die plausiblerweise anzunehmenden *zeitlichen Wirkungszeiträume* langfristiger sowie kurzfristiger Bestimmungsfaktoren des Wählerverhaltens. Kurzfristige Effekte werden bei großen zeitlichen Distanzen zum Wahltermin sicherlich noch nicht zum Tragen kommen. Dies dürfte sich dann ändern, wenn im Zuge des beginnenden Wahlkampfes zunehmend in der Öffentlichkeit Streitfragen thematisiert und Kandidaten präsentiert werden. Mittlere zeitliche Distanzen vor der Wahl werden somit vermutlich am stärksten von kurzfristig bedingten Abweichungen der Wahlabsicht von längerfristigen Neigungen betroffen sein. Ebenso scheint die Annahme plausibel, daß kurz vor der Wahl, gegen Ende des Wahlkampfes, die langfristigen Determinanten des Wählerverhaltens allmählich wieder an Bedeutung gewinnen, weil nun die Entscheidungssituation tatsächlich ansteht.

1.3.3.2 Verfahren mit Rückerinnerungsgewichtung

Verfahren der Rückerinnerungsgewichtung (nach dem Quotientenverfahren oder nach dem Differenzenverfahren) versuchen, durch die Korrektur der „rohen“ Wahlabsichten mit Hilfe der Relation aus Vorwahlergebnis und Rückerinnerungsergebnis solche aktuellen „meinungsklimatischen“ Kurzfristeinflüsse auszuschalten. Da die Wirksamkeit von Quotienten- und Differenzenverfahren mehrfach bestätigt wurde, ist ein funktionaler Verlauf der Fehler wie bei den „rohen“ Wahlabsichten nicht zu erwarten. Daß aber die Abweichungen bei Wahlabsichten mit Quotienten- oder Differenzengewichtung unabhängig vom zeitlichen Abstand sind, ist ebenfalls wenig wahrscheinlich. So lassen sich mehrere Fehlerquellen anführen, die mit dem zeitlichen Abstand variieren und deren Korrigierbarkeit durch die Rückerinnerungsgewichtungen unklar ist (vgl. dazu Rattinger/Ohr 1989: 295): Zum einen dürfte mit wachsender Entfernung vom Wahltermin die Zahl derjenigen steigen, die in ihrer Wahlabsicht noch „unentschieden“ sind. Vor allem aber ist anzunehmen, daß sich die „Unentschiedenen“ nicht nur unter denjenigen Befragten befinden, die auf die Wahlabsichtsfrage mit „weiß nicht“ geantwortet haben. Vielmehr wird vor allem bei größerer zeitlicher Distanz zum Wahltag zu erwarten sein, daß Probanden

„weichere“ Festlegungen ihrer Wahlabsicht vornehmen. Weiterhin sinkt mit zunehmendem Abstand zum Wahltermin die Sicherheit der *Wahlteilnahme* eines Befragten, umgekehrt steigen die Wahrscheinlichkeiten, daß ein Proband mit Wahlabsicht nicht an der Wahl teilnimmt, beziehungsweise ein Proband ohne Wahlabsicht trotzdem zur Wahl geht.

Eine letzte hier zu nennende Ursache für die Erwartung eines *Genauigkeitstrichters*, der sich bis zum Wahltermin ständig verengt, ist der sogenannte „last minute swing“: Die Wahrscheinlichkeit, daß ein Individuum aufgrund politischer Ereignisse doch noch „umkippt“, seine Parteipräferenz also noch ändert, wird bei Annäherung zum Wahltag abnehmen. Insgesamt sollte also die Fehlermarge politischer Gewichtungen mit wachsendem zeitlichem Abstand vom Wahltermin kontinuierlich zunehmen. Eine konstante Steigung, wie dies im linearen Regressionsmodell unterstellt wird, scheint allerdings nicht plausibel. So wird man nicht davon ausgehen, daß schon in unmittelbarer Nähe zum Wahltag eine Verschlechterung der Anpassung zwischen Wahlabsichten und -ergebnis erfolgt. Auch wird sich der Fehler bei größerer zeitlicher Entfernung nicht ständig weiter vergrößern, sondern irgendwann eine Obergrenze erreichen.

Einen funktionalen Verlauf, der diesen Annahmen entspricht und der zudem durch die empirischen Werte aus Tab. 1.2 nahegelegt wird, hat die *logistische* Funktion. Wenn man sich für das logistische Modell entscheidet, gilt es zwei Eigenschaften eines solchen Zusammenhangs zu beachten: Ein logistischer Verlauf wird einmal durch die Lage des Wendepunktes bestimmt und kann zum zweiten eine eher treppenförmige oder aber eine eher abgeflachte, linearisierte Gestalt annehmen. Beide Parameter lassen sich regressionsanalytisch schätzen. Der zweite Aspekt läßt sich durch eine unterschiedliche Skalierung des zeitlichen Abstandes testen, wobei mit (absolut) sehr großen Werten im Exponenten der logistischen Funktion ihr Verlauf eher treppenförmig wird, bei betragsweise kleinen Werten dagegen eher abgeflacht. Mit Skalierungen des Abstandes in Tagen, Wochen, Halbmonaten, Monaten, schließlich Quartalen testen wir insgesamt fünf Varianten. Gleichzeitig wird über das Kriterium der besten Datenanpassung der Wendepunkt der logistischen Funktion für jede Skalierung ermittelt. Tabelle 3 zeigt die Datenanpassungen für diese Funktionen sowie den Vergleich zu einem linearen Verlauf.

Tab. 1.3 ist zu entnehmen, daß der logistische Zusammenhang deutlich besser als ein linearer Verlauf an die Daten angepaßt ist. Dies gilt im übrigen auch im Vergleich mit anderen, ebenfalls plausiblen Verläufen (wie etwa der Quadratwurzelfunktion oder der natürlichen Logarithmusfunktion). Der funktionale Verlauf ist bei beiden Verfahren fast identisch: Die höchste *Varianzklärung* erreicht sowohl bei den Quotienten- als auch bei den Differenzmodellen eine Skalierung der Abstandsvariablen in Wochen, was einem eher treppenförmigen Verlauf entspricht. Allerdings sind die R^2 -Werte für die Ska-

Tabelle 1.3: Erklärung der mittleren absoluten Abweichung zwischen Wahlabsichten und Wahlergebnis durch den zeitlichen Abstand zum Wahltag^a

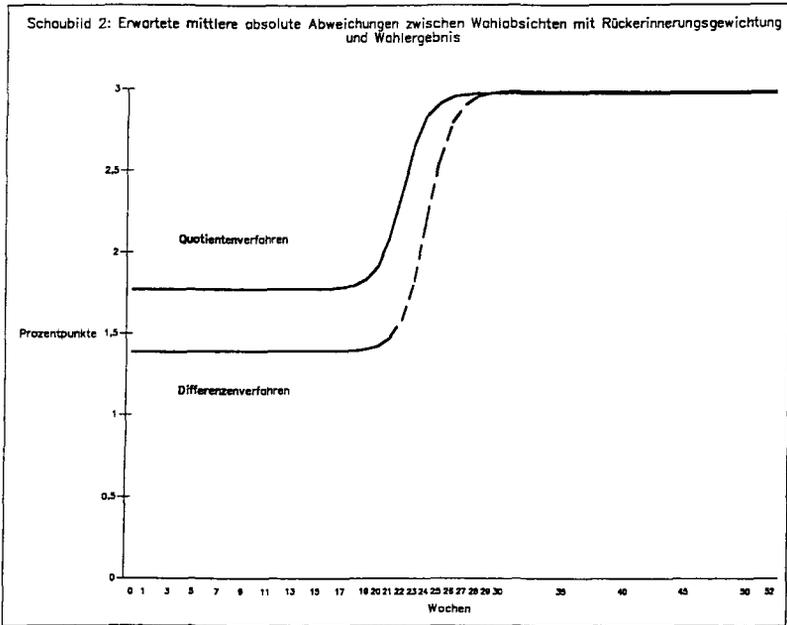
	Quotientenverfahren		Differenzenverfahren	
	Wendepunkt	R^2	Wendepunkt	R^2
a) Logistische Funktion, Skalierung in				
Tagen	150 T.	250	180 T.	427
Wochen	22 Wo.	263	24 Wo.	456
Halbmonaten	21 Wo.	257	24 Wo.	448
Monaten	21 Wo.	246	24 Wo.	433
Quartalen	16 Wo.	212	24 Wo.	386
b) Linear	—	177	—	350

a: Tab. 1.3 enthält in den Spalten 1 und 3 die Wendepunkte der logistischen Funktion in Tagen (T) bzw. Wochen (Wo) für die unterschiedlichen Skalierungen bei Quotientenverfahren bzw. Differenzenverfahren. Die Spalten 2 und 4 zeigen das Bestimmtheitsmaß R^2 , multipliziert mit 1000.

lierung der Abstandsvariablen in Halbmonaten und Monaten nur unwesentlich geringer, so daß ein eher „linearisierter“ Verlauf die Punktwolke fast ebenso gut annähert. Auch die Wendepunkte des jeweils besten Modells stimmen mit 22 beziehungsweise 24 Wochen fast überein. Die Datenanpassung insgesamt ist bei beiden Varianten befriedigend, vor allem bei den Fehlern des Differenzenverfahrens, wobei sich in der besseren Datenanpassung der Differenzenmodelle das schon in Tab. 1.2 sichtbare Faktum widerspiegelt, daß Differenzenmodelle kurzfristige Schwankungen insgesamt etwas besser in den Griff bekommen als Quotientenmodelle. Schaubild Tab. 1.2 stellt den zeitlichen Verlauf der nach der Regressionsschätzung erwarteten mittleren absoluten Fehler für Quotienten- und Differenzenverfahren dar.

Die untere Kurve zeigt den nach der Regressionsschätzung erwarteten Fehlerverlauf für Differenzenmodelle, die obere Kurve den Verlauf für Quotientenmodelle. Differenzenmodelle sind, wie auch schon bei den empirischen Werten

Abbildung 1.2: Erwartete mittlere absolute Abweichungen zwischen Wahlabsichten und Rückerinnerungsgewichtung und Wahlergebnis



zu erkennen war, bis zu einer Distanz von ca. sieben Monaten deutlich genauer als Quotientenmodelle. Erst ab einer Distanz von etwa sieben Monaten haben beide Verfahren fast identische erwartete Fehler. Der Verlauf nach dem jeweils besten Regressionsmodell ist nahezu treppenförmig: Bis zu einem Abstand von ungefähr fünf Monaten verändert sich der erwartete Fehler kaum, steigt dann zwischen fünf und sieben Monaten um ca. 1,5 Prozentpunkte an, um sich bei wachsender Entfernung kaum weiter zu vergrößern. Ein eher abgeflachter Verlauf (Skalierung in Monaten) hätte nach Tab. 1.3 eine „fast“ ebensogute Datenanpassung: Hier erhöht sich der erwartete Fehler etwa für Differenzenmodelle bis zu einer zeitlichen Distanz von fünf Monaten von 1,3 auf 1,8 Prozentpunkte (ohne Schaubild).

Zusammenfassend kann festgehalten werden, daß die *logistische Funktion* als funktionale Form des Zusammenhangs zwischen der Abweichung von Wahlabsichten und -ergebnis und zeitlichem Abstand zum Wahltermin empirisch gut abgesichert werden kann. Auch substantielle Überlegungen sprechen, wie oben ausgeführt, für einen solchen Zusammenhang. Ferner zeigt sich, daß vor allem bei geringeren zeitlichen Distanzen von weniger als einem halben Jahr

Differenzenmodelle den Quotientenmodellen überlegen sind, ein Befund, der sich auch für einen längeren Analysezeitraum nachweisen ließ (vgl. Rattinger/Ohr 1989). Der Vergleich zwischen den lediglich demographisch gewichteten Wahlabsichten und denjenigen mit Rückerinnerungsgewichtung (vgl. Tab. 1.2 und die Schaubilder Tabellen 1.1 und 1.2) ergibt bei zeitlichen Distanzen bis etwa zu einem halben Jahr einen deutlich niedrigeren Fehler der Rückerinnerungsmodelle. Bei größerer zeitlicher Entfernung vom Wahltermin schneiden die politisch gewichteten Wahlabsichten dagegen überraschenderweise schlechter als die „rohen“ Werte ab. Rückerinnerungsgewichtung führt hier also zu einer *Verschlechterung* der Anpassung der Wahlabsichten an die Endergebnisse.

Aus welchem Grund weisen nun Rückerinnerungsmodelle bei großen zeitlichen Abständen zum Wahltag relativ höhere Fehler auf? Die Logik der Rückerinnerungsgewichtung basiert, wie oben erwähnt, auf der Annahme einer weitgehend identischen Verzerrung von Wahlabsicht und Rückerinnerung in einer Stichprobe. Demnach muß sich die zeitliche Assoziation von Wahlabsichts- und Rückerinnerungsverzerrung hinsichtlich Stärke und funktionaler Form möglichst entsprechen, wenn die Rückerinnerungsgewichtung *unabhängig* vom zeitlichen Abstand zum Wahltermin zur Verringerung des Fehlers geeignet sein soll. Die zeitliche Abhängigkeit der (mittleren absoluten) Wahlabsichtsverzerrung wurde bereits in Schaubild Tab. 1.1 dargestellt, da die Wahlabsichtsverzerrung natürlich nichts anderes ist als der Fehler der „rohen“ Wahlabsichten. Eine nach unten offene Parabel konnte den zeitlichen Zusammenhang bei einem R^2 von 0,15 noch am besten annähern. Für die über die Parteien hinweg gemittelte absolute Verzerrung der Rückerinnerung läßt sich demgegenüber weder ein linearer noch ein nichtlinearer Zusammenhang mit dem zeitlichen Abstand ermitteln. Das Niveau der Rückerinnerungsverzerrung unterliegt im zeitlichen Verlauf keinen systematischen Änderungen. Auch die *parteispezifische* Gegenüberstellung von Wahlabsichtsverzerrung und Rückerinnerungsverzerrung führt im wesentlichen zum gleichen Befund: Die Verzerrung der Rückerinnerung läßt durchgängig kein systematisches zeitliches Verlaufsmuster erkennen, während die Wahlabsichtsverzerrung bei den Grünen und den Sonstigen (R^2 : 0,26 bzw. 0,10) einer ansteigenden logistischen Kurve, bei der FDP (R^2 : 0,20) einer fallenden logistischen Funktion folgt. Nebenbei bemerkt bedeutet das, daß die Fragen nach Wahlabsicht und Rückerinnerung nicht einfach Mehrfachmessungen der gleichen aktuellen Präferenz sein können.

1.3.4 Folgerungen für die Beurteilung der Güte der Anpassung von Wahlabsichtsbefragungen an spätere Wahlergebnisse

Welches Ausmaß an Vertrauen können wir in die Annäherung späterer Wahlergebnisse durch Wahlabsichten aus Umfragedaten haben? Wie können die bisher dargestellten Befunde zur zeitlichen Dimension dieser Fehler für die Beantwortung dieser Frage nutzbar gemacht werden? Das konventionelle Maß zur Abschätzung der Verlässlichkeit von Wahlabsichtsumfragen ist das *klassische Konfidenzintervall* um die geschätzten Parteianteile. Die im letzten Abschnitt durchgeführte Regression des mittleren absoluten Fehlers auf den zeitlichen Abstand zum Wahltag erlaubt es, diesem üblichen Konfidenzintervall ein alternatives, *zeitabhängiges Konfidenzintervall* gegenüberzustellen, ein Konfidenzintervall um die aufgrund des zeitlichen Abstandes erwarteten Fehler. Bei einem Vergleich zwischen beiden ist zu berücksichtigen, daß wir als Fehlermaß den *mittleren* absoluten Fehler verwenden, daß aber das übliche Konfidenzintervall eigentlich für jede Partei separat geschätzt werden muß. Dies ist erforderlich, weil die Ausdehnung des Konfidenzintervalls und damit die Schätzgenauigkeit unter anderem durch die Größe des Parteianteils bestimmt wird und sich so, *ceteris paribus*, bei größeren Parteianteilen auch größere Intervalle ergeben.

Die Varianz der Schätzfunktion für den wahren Parteienanteil p beträgt bekanntlich $p \cdot (1 - p) / n$. Dieser Ausdruck nimmt bei $p = 0,5$ den maximalen Wert von $0,25/n$ an. Damit wird auch, *ceteris paribus*, die Ausdehnung des Konfidenzintervalls maximal. Man kann nun eine in statistischer Terminologie konservative Schätzung des üblichen Konfidenzintervalls mit einem fiktiven (maximalen) Anteil von 50 % vornehmen. Bei 5 % Irrtumswahrscheinlichkeit und einem Stichprobenumfang von $n = 1000$ ergibt sich so eine Ausdehnung von $\pm 3,01$ Prozentpunkten um den geschätzten Parteienanteil. Bei Betrachtung nur der absoluten Abweichungen beträgt die Ausdehnung somit 3,01 Prozentpunkte. Dies ist gleichzeitig auch die Länge eines konservativen, einseitigen 95 %-Konfidenzintervalls um den Nullpunkt des *absoluten Fehlers*. Neben dieser konservativen Bestimmung des Konfidenzintervalls läßt sich auch ein *mittleres* Konfidenzintervall berechnen, das die Größenunterschiede der einzelnen Parteien in Rechnung stellt. So wissen wir aufgrund der bisherigen Bundestagswahlen, daß beispielsweise die FDP einen Stimmenanteil von sicherlich weniger als 15 Prozent erreichen wird. Zur Berechnung der Varianz $p \cdot (1 - p) / n$ für die FDP kann deshalb 0,15 als maximaler Wert für den wahren Anteil p angenommen werden. Die Ausdehnung des so ermittelten Konfidenzintervalls beträgt für die FDP, bei einem angenommenen Stichprobenumfang von 1000 und einem Konfidenzniveau von 95 Prozent, $\pm 2,21$ Prozentpunkte. Errechnet

man in dieser Weise für jede der vier betrachteten Parteien und die „Sonstigen“ ein 95 %-Intervall (s. dazu Anhang 2) und mittelt man die Ausdehnung der Intervalle, so erhält man eine durchschnittliche absolute Ausdehnung des üblichen Konfidenzintervalls von 2,32 Prozentpunkten.¹⁰

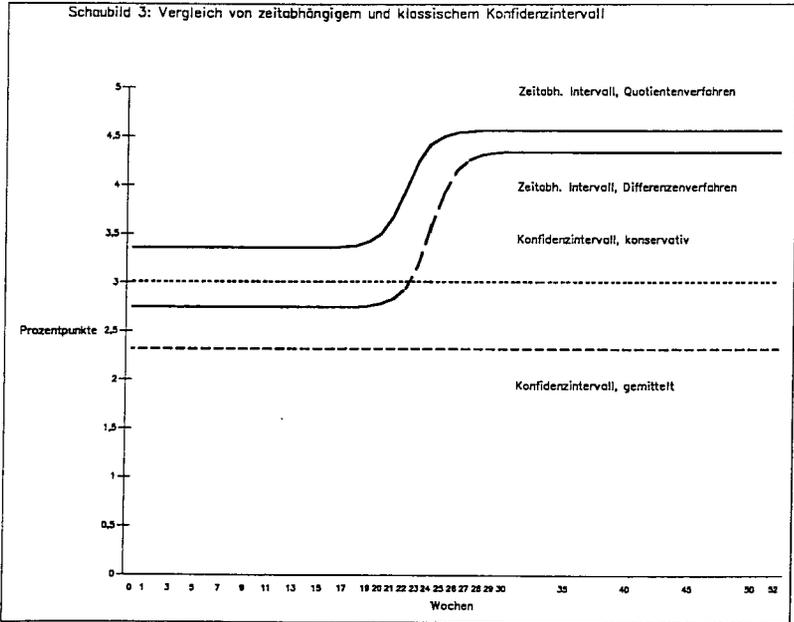
Nun wissen wir aber nach unserer bisherigen Analyse, daß politisch gewichtete Wahlabsichten mit sinkender Distanz zur Wahl näher an das Endergebnis rücken. Eine alternative, zeitabhängige Formulierung des Konfidenzintervalls geht deshalb aus von den Schätzungen der *zu jeder zeitlichen Distanz erwarteten Fehler*, wie sie in Schaubild Tab. 1.2 dargestellt sind. Diese Regressionsschätzungen für den Erwartungswert der mittleren Abweichung werden bei normalverteilten Störgrößen ungefähr einer Normalverteilung folgen.¹¹ Dabei wird die Stichprobenverteilung um den (wahren) Erwartungswert des Fehlers zentriert sein. In etwa 50 Prozent der Fälle wird der mittlere absolute Fehler geringer sein als erwartet, in etwa 50 Prozent der Fälle größer. Bei der Ermittlung des zeitabhängigen Konfidenzintervalls muß berücksichtigt werden, daß ersteres eine bessere Anpassung an das Wahlergebnis bedeutet. Zur Abschätzung der Verlässlichkeit der Ergebnisse der Wahlabsichtsfrage ist nur die Obergrenze des Fehlers von Interesse. Aus diesem Grunde muß das zeitabhängige Konfidenzintervall als einseitiges, oberes 95 %-Konfidenzintervall aufgefaßt werden. Dabei erhalten wir die obere Grenze des 95 %-Intervalls, wenn wir zu dem jeweils erwarteten Fehler (vgl. Schaubild Tab. 1.2) das Produkt aus dem 0,95-Perzentil der Standardnormalverteilung (1,64) und dem Standardfehler der Regressionsschätzung addieren. Dieser (geschätzte) Standardfehler beträgt bei den Quotientenverfahren 0,97 Prozentpunkte, bei den Differenzenverfahren 0,83 Prozentpunkte. Man erhält also die obere 95 %-Grenze, wenn man die logistischen Kurven aus Schaubild Tab. 1.2 um $1,64 \cdot 0,97$ beziehungsweise $1,64 \cdot 0,83$ nach oben verschiebt.

Vergleichen wir nun die absolute Ausdehnung des üblichen Konfidenzintervalls, konservativ sowie gemittelt, mit der Ausdehnung des zeitabhängigen Konfidenzintervalls für Quotienten- und Differenzenmodelle: Man sieht in Schaubild Tab. 1.3, daß sich bei zeitlichen Distanzen zum Wahltag bis zu etwa fünf Monaten die zeitabhängigen Konfidenzintervalle nicht allzu stark von den beiden Varianten des üblichen Konfidenzintervalls unterscheiden. Quoti-

¹⁰Die Mittelung der parteispezifischen Konfidenzintervalle ist statistisch gesehen nicht ganz korrekt, da die Schätzfunktionen für die Parteiateile nicht unabhängig voneinander sind.

¹¹Die theoretische Stichprobenverteilung der parteispezifischen *absoluten* Fehler weicht von einer Normalverteilung ab. Die *empirische* Verteilung der mittleren absoluten Fehler unserer Datenbasis erreicht bei den Differenzenmodellen eine recht gute Annäherung an eine Normalverteilung (Kolmogoroff-Smirnoff-Anpassungstest nicht signifikant auf dem 5 %-Niveau), während die Verteilung für die Quotientenmodelle stärker von einer Normalverteilung abweicht (Kolmogoroff-Smirnoff-Anpassungstest auf dem 5 %-Niveau signifikant).

Abbildung 1.3: Vergleich von zeitabhängigem und klassischem Konfidenzintervall



entenmodelle sind aber, gemessen am zeitabhängigen Intervall, auch in diesem zeitlichen Bereich, ungenauer als die beiden konventionellen Konfidenzintervalle nahelegen. Demgegenüber sind hier Differenzengewichtungen nach dem zeitabhängigen Intervallverlauf zwar ebenfalls etwas ungenauer, als es die gemittelte Form des klassischen Intervalls besagt, liegen jedoch unter der Ausdehnung des konservativen Konfidenzintervalls. In diesem zeitlichen Distanzbereich erhält man zum Konfidenzniveau von 95 Prozent eine Ausdehnung von 2,75 Prozentpunkten für das zeitabhängige Intervall der Differenzenmodelle. Bei geringen zeitlichen Abständen zum Wahltag wird also der mittlere absolute Fehler für Differenzenmodelle bei dem angegebenen Konfidenzniveau mit 95 %-iger Sicherheit unter 2,75 Prozentpunkten liegen, während das übliche, konservative Konfidenzintervall einen entsprechenden Fehler von rund 3 Prozentpunkten ausweist. Bei größeren zeitlichen Distanzen ab ca. sieben Monaten sind dagegen sowohl Quotienten- als auch Differenzenverfahren *wesentlich* ungenauer, als nach beiden konventionellen Intervallen zu erwarten wäre. Die Ausdehnung der zeitabhängigen Intervalle liegt nun bei etwa 4,3 bzw. 4,6 Prozentpunkten gegenüber den 2,32 bzw. 3,01 Prozentpunkten der konventionellen Intervalle.

Das Problem des konventionellen Konfidenzintervalls liegt darin, daß zu dem Zeitpunkt der Umfrage ja gar keine Wahl stattfindet. Es besteht noch ein mehr oder minder großer zeitlicher Vorlauf zum Wahltermin, der vom klassischen Konfidenzintervall überhaupt nicht in Rechnung gestellt wird. Das hier vorgestellte zeitabhängige Konfidenzintervall berücksichtigt demgegenüber den bisherigen Zusammenhang zwischen den Abweichungen von Wahlabsichten und -ergebnissen und der zeitlichen Distanz zur Wahl. Deswegen dürfte die Fehlermarge von Wahlabsichten zu einem gegebenen Zeitpunkt, damit einem bestimmten zeitlichen Abstand zum Wahltag, viel eher der Ausdehnung des zeitabhängigen Konfidenzintervalls entsprechen. Dies läßt sich mit unseren Daten einfach überprüfen, indem wir für das zeitabhängige Konfidenzintervall und die beiden Varianten des konventionellen Konfidenzintervalls die Anteile derjenigen Studien vergleichen, die *innerhalb* des jeweiligen Intervalls (zum Konfidenzniveau von 95 Prozent) liegen (s. dazu Tabelle 1.4).

Tabelle 1.4: „Empirisches Konfidenzniveau“ für das konventionelle und das zeitabhängige Konfidenzintervall nach dem Gewichtungmodell

	Konventionelles Konfidenzintervall		Zeitabhängiges Konfidenzintervall
	Konservativ	Mittelung	
Differenzenverfahren	82,9 (102)	64,2 (79)	94,3 (116)
Quotientenverfahren	79,7 (98)	58,5 (72)	95,9 (118)

Die erste Zahl gibt jeweils den Anteil der (123) Fälle an, die im Konfidenzintervall liegen. Die Zahl in Klammern bezeichnet die absolute Häufigkeit.

Beim zeitabhängigen Konfidenzintervall sind dies sowohl bei den Differenzen- als auch bei den Quotientenverfahren fast exakt 95 Prozent der Fälle. Das Konfidenzniveau ist hier also ein verlässliches Maß der Sicherheit. Anders verhält es sich mit dem konventionellen Konfidenzintervall: Bei der konservativen Berechnungsweise liegen nur, jeweils für Differenzen- und Quotientenverfahren, 82,9 beziehungsweise 79,7 Prozent der Studien im Intervall. Für die gemittelte Form des konventionellen Konfidenzintervalls betragen die entsprechenden Anteile sogar nur 64,2 beziehungsweise 58,5 Prozent. Das heißt, daß das vorgegebene Konfidenzniveau des konventionellen Konfidenzintervalls (in diesem Fall also 95 Prozent) eine Sicherheit über die Annäherung des Wahlergebnisses durch Wahlabsichten suggeriert, die in Wirklichkeit gar nicht gerechtfertigt ist. Auch unter diesem Blickwinkel ist daher das zeitabhängige Konfidenzintervall eindeutig überlegen. Diese Folgerung gilt zumindest dann, wenn man, wie wir hier, nicht am Ausmaß der Annäherung der

Wahlabsichtsfrage an die (essentiell unbekannt) Stimmungslage in der Population interessiert ist, sondern an der Beziehung zwischen Wahlabsichten und späterem Wahlverhalten.

1.4 Schlußbemerkungen

Einschätzungen politischer Stimmungslagen und Trends aus Wahlabsichtsbefragungen auf der Grundlage der Rückerinnerungsgewichtung können gerade kurz vor Wahlen meist eine erstaunlich hohe Genauigkeit für sich beanspruchen. Die empirische Wahlforschung ist von *theoretisch fundierten* Prognosemodellen mit einer den Rückerinnerungsmodellen nahekommenden Leistungsfähigkeit noch sehr weit entfernt. Gerade deswegen ist es umso dringlicher, Schwachpunkte wie auch Verfeinerungsmöglichkeiten der Rückerinnerungsgewichtungen herauszuarbeiten (vgl. Rattinger/Ohr 1989). Die meisten Meinungsforschungsinstitute sind in dieser Richtung beständig und kreativ tätig. Ebenso wichtig ist es aber, realistische Kriterien zur Beurteilung der „Genauigkeit“ der Anpassung von Wahlabsichten an Wahlergebnisse zu gewinnen. Das klassische Konfidenzintervall für die einzelnen Parteiateile stellt gegenüber der bloßen Angabe von Punktschätzungen sicherlich einen Fortschritt dar. Es ist nach wie vor unverzichtbar, wenn man feststellen will, ob die Neigung zu einer Partei sich gegenüber einem Vergleichszeitpunkt signifikant verändert hat, ob die Führung einer Partei vor einem Konkurrenten statistisch signifikant ist oder ob der Sprung über die Sperrklausel zum gegebenen Zeitpunkt als gesichert gelten kann.

Dieses konventionelle Konfidenzintervall bezieht sich aber nur auf die wahre — natürlich unbekannt — aktuelle politische Stimmungslage. Will man dagegen Wahlabsichtsdaten auch danach beurteilen, wie nahe sie am späteren Wahlergebnis liegen, so ist das in diesem Aufsatz vorgestellte *zeitabhängige Konfidenzintervall* vorzuziehen. Es beruht auf der sehr einfachen, plausibel begründbaren und empirisch untermauerbaren Erwartung, daß man aus Umfragen weiter vor einer Wahl mit geringerer Genauigkeit auf die späteren Ergebnisse schließen kann als unmittelbar davor. Unser zeitabhängiges Vertrauensintervall erlaubt es, für jeden Zeitpunkt (innerhalb des Vorwahljahres) eine realistischere Abschätzung der Fehlermarge der Wahlabsichten im Hinblick auf das Ergebnis der späteren Wahl vorzunehmen. Eine Übertragung dieser Analyse auf neuere und zukünftige Wahlen ist jederzeit möglich und sinnvoll. Wünschenswert ist dabei vor allem eine Verbreiterung der Datenbasis über die hier einbezogenen 123 Studien hinaus, um das zeitabhängige Konfidenzintervall noch genauer abschätzen und zu einer parteien- und wahlspezifischen Auswertung übergehen zu können.

Anhang 1: Gewichtungsverfahren

$P1_i$: Anteil der Wahlabsichten für die i -te Partei

$$P1_i = \frac{a_i}{A} \quad (1.1)$$

$P2_i$: Rückerinnerungsgewichtung der Wahlabsichten nach dem Quotientenverfahren

$$P2_i = \frac{a_i w_{i,t-1}}{r_i \sum \frac{a_i w_{i,t-1}}{r_i}} \quad (1.2)$$

$P3_i$: Rückerinnerungsgewichtung der Wahlabsichten nach dem Differenzenverfahren

$$P3_i = \frac{a_i}{\sum a_i} + \frac{w_{i,t-1}}{\sum w_{i,t-1}} - \frac{r_i}{\sum r_i} \quad (1.3)$$

wobei:

a_i : Zahl der Wahlabsichten für die i -te Partei

A : Gesamtzahl der Wahlabsichten ($= \sum a_i$)

r_i : Zahl der Rückerinnerungen an Stimmabgabe bei der Vorwahl für die i -te Partei

$w_{i,t-1}$: Zahl der Wähler der i -ten Partei bei der Vorwahl

Anhang 2: Berechnung des mittleren konventionellen Konfidenzintervalls

Angenommener maximaler (wahrer) Anteil für CDU/CSU: 50 %, SPD: 45 %, FDP: 15 %, Grüne: 10 %, „Sonstige“ 5 %.

Anhang 3: Datenbasis

Datensatz des „Arbeitskreises Empirische Wahlforschung“ mit Daten über Umfragestudien zu den vier Bundestagswahlen 1976 bis 1987 als Untersuchungseinheiten ($n=229$). Dieser Datensatz wurde von uns ergänzt durch 14 Wahlstudien der Forschungsgruppe Wahlen zur Bundestagswahl 1987. Von diesen insgesamt 243 Studien wurden diejenigen von der weiteren Analyse ausgeschlossen, die mehr als zwölf Monate (genau: 400 Tage) vor der jeweiligen Bundestagswahl erhoben wurden. Nach dieser zeitlichen Eingrenzung verblieben 123 Umfragen, für die jeweils die drei Gewichtungsvarianten berechnet wurden.

Literatur

- Buchanan, W., 1986: Election Predictions: An Empirical Assessment. *Public Opinion Quarterly* 50: 222-227.
- Falter, J.W./Rattinger, H., 1983: Parteien, Kandidaten und politische Streitfragen bei der Bundestagswahl 1980: Möglichkeit und Grenzen der Normal-Vote-Analyse. S. 320-421 in: M. Kaase/H.D. Klingemann (Hrsg.), *Wahlen und politisches System — Analysen aus Anlaß der Bundestagswahl 1980*. Opladen: Westdeutscher Verlag.
- Falter, J.W./Rattinger, H., 1986: Die Bundestagswahl 1983: Eine Normalwahlanalyse. S. 289-337 in: H.D. Klingemann/M. Kaase/(Hrsg.), *Wahlen und politischer Prozeß — Analysen aus Anlaß der Bundestagswahl 1983*. Opladen: Westdeutscher Verlag.
- Falter, J.W./Schumann, S., 1989: Methodische Probleme von Wahlforschung und Wahlprognose. *Aus Politik und Zeitgeschichte* B43: 3-14.
- Hentschel, H./Tennstädt, F., 1985: *Election Forecasts. Problems and Solutions* (Revised manuscript of a paper presented at the conference on 'Opinion Polls and Election Forecasts' (UIMP), Sevilla, Sept. 26-27, 1985).
- Hofmann, J.R., 1976: Geheime Wahl ohne Geheimnis. Wie das IfD Allensbach eine Wahlprognose entwickelt. *DAS PARLAMENT*, Nr. 38, vom 18. September 1976: 6-7.
- Jagodzinski, W./Kühnel, S.M., 1990: Zur Schätzung der relativen Effekte von Issueorientierungen, Kandidatenpräferenz und langfristiger Parteibindung auf die Wahlabsicht. S. 5-63 in: K. Schmitt (Hrsg.), *Wahlen, Parteieliten, politische Einstellungen — Neuere Forschungsergebnisse*. Frankfurt am Main: Peter Lang.
- Perry, P., 1979: Certain Problems in Election Survey Methodology. *Public Opinion Quarterly* 43: 312-325.
- Rattinger, H./Ohr, D., 1989: Wahlprognosen in einer Welt ohne Stichprobenfehler: Analytische Überlegungen und empirische Befunde. S. 282-331 in: J.W. Falter/H. Rattinger/K.G. Troitzsch (Hrsg.), *Wahlen und politische Einstellungen in der Bundesrepublik Deutschland. Neuere Entwicklungen der Forschung*. Frankfurt am Main: Peter Lang.
- Ulmer, F., 1987: Wahlprognosen und Meinungsumfragen: Der Orakelspruch mit dem repräsentativen Querschnitt. *Bild der Wissenschaft*, Januar: 89-100.