

Abschnitt II: Wahlbörsen und deren Einsatzmöglichkeiten in der Wahlforschung

Im zweiten Abschnitt, der die Kapitel 7 bis 10 umfaßt, werden Wahlbörsen primär aus politikwissenschaftlicher Sicht betrachtet. Dabei wird versucht zu klären wie gut politische Aktienmärkte als Prognoseinstrumente sind und wo strukturelle Unterschiede zur klassischen Wahlforschung, insbesondere zu Umfragen, liegen. In Kapitel 7 wird gezeigt, daß die moderne Umfrageforschung dann an ihre Grenzen stößt, wenn es um die Erstellung einer Wahlprognose geht. Während das Grundprinzip der Befragung – Repräsentativität einer Stichprobe für die Grundgesamtheit durch das Gesetz der großen Zahl bei einer Zufallsauswahl – funktioniert, bleibt man damit doch auf die Erhebung der Stimmung zu einem bestimmten Zeitpunkt beschränkt. Die Ableitung einer Prognose aus den Rohdaten der Befragung erfordert immer die Bearbeitung durch Experten.

Umfragen und Wahlbörsen werden einander im nächsten Kapitel unter verschiedenen Blickwinkeln gegenübergestellt. Die in früheren Kapiteln bereits erarbeiteten Stärken und Schwächen beider Instrumente werden in komprimierter Form aufgegriffen, um Möglichkeiten und Grenzen der politischen Aktienmärkte für eine präzisere und zuverlässigere Wahlforschung aufzuzeigen. Dabei wird insbesondere auf die Unterschiede in der jeweiligen Fragestellung der beiden Instrumente eingegangen. Es bleibt jedoch – wie bei Umfragen – die Schwierigkeit, daß eine Wahlentscheidung inhärent unsicher ist und daher nie exakt vorausgesagt werden kann.

Konkrete Designvorschläge zum erfolgreichen Einsatz von politischen Aktienmärkten sowie einen Ausblick auf deren Einsatzmöglichkeiten in der Meinungsforschung bringt das 9. Kapitel. Dabei wird argumentiert, daß die große Medienwirksamkeit zum Fluch der Wahlbörsen geworden ist, denn damit wurden diese auch für Manipulationen interessant. Die erfolgreichsten Wahlbörsen waren jene, die kaum beachtet wurden und so liegt vielleicht auch die Zukunft dieses Instruments im „Stillen“.

7 Klassische Wahlforschung und ihre Probleme

Demokratische Wahlen stellen ein relativ junges Phänomen dar. Sie finden in den Vereinigten Staaten seit ca. 220 Jahren, in den meisten europäischen Ländern meist seit weniger als hundert Jahren mehr oder weniger regelmäßig statt. Eine Folgeerscheinung – und damit noch jünger – ist die Wahlforschung, die nach dem Ersten Weltkrieg in den USA ihren Anfang nahm.

Dabei erwiesen sich Befragungen von Wahlberechtigten als relativ gutes Vorhersageinstrument und nahmen schnell eine zentrale Rolle in der Wahlforschung ein. Ende der 30er Jahren entwickelte George Gallup die Grundpfeiler der heute dominierende Methode der Prognose über kleine Stichproben, welche die bis dahin dominierenden Großerhebungen ablöste.

Doch Umfragen haben auch Nachteile, die früher unvermeidlich waren, heute aber eventuell behoben werden können. Kritikpunkte sind die fixe Fragestellung, die erforderliche Repräsentativität sowie die Gefahr von *self fulfilling prophecies*.¹ Insgesamt sind Erhebungen oft sehr unflexibel und steril, es gibt keinen echten Austausch mit den Befragten, die meist in eine sehr passive Rolle gedrängt werden. In der Regel erfolgt auch keine Rückmeldung an die Teilnehmer. Vor einigen Jahrzehnten war dies mit Rücksicht auf die Kosten nicht anders durchführbar, doch heute sind mit den Möglichkeiten auch die Ansprüche der Befragten gestiegen – sie sind es gewohnt, wirklich eingebunden zu werden, denn das Internet hat eine Kultur der Interaktivität geschaffen. Dieses Medium werden sich Wahlforscher früher oder später zunutze machen, denn stets reagierten sie auf das Erscheinen neuer Medien. Wahlforschung muß zeitgemäß sein, um ihre Berechtigung nicht zu verlieren.

¹ bereits durch die Art der Fragestellung kann das Ergebnis einer Erhebung verzerrt werden

7.1 Ökonometrie – eine Sackgasse der Wahlforschung?

Neben der Wahlprognose über den Markt, also Wahlbörsen, und der Vorhersage über Umfragen wurde Anfang der 90er Jahre von Volkswirtschaftlern versucht, den Ausgang einer Wahl über ökonometrische Modelle zu „berechnen“. Den Ausgangspunkt polit-ökonomischer Modelle bildet die plausible Überlegung, daß gesamtwirtschaftliche Daten wie Inflation, Arbeitslosigkeit und Wachstum Einfluß auf die Wiederwahlchancen einer Regierung bzw. eines Präsidenten haben.

Diese Faktoren werden in einer Formel mit (aus Vergangenheitsdaten geschätzten) Gewichtungen aggregiert und als Ergebnis erhält man nach mehr oder minder komplizierten Algorithmen den Stimmenanteil von Regierung bzw. Opposition. Das bekannteste derartige Modell entwickelte Ray Fair für die USA (vgl. Fair-Model, Internet). Das *Fair-Model* konnte jedoch in der Wahlvorhersage keine guten Ergebnisse erzielen: 1992 überschätzte Fair damit den Stimmenanteil von Bush um 9,8 Prozent, 1996 wurde ein Sieg Doles nicht ausgeschlossen² und 2000 wurde wiederum der spätere Verlierer Gore als Wahlsieger prognostiziert (vgl. Frey/Götte/Kucher 1996).

Es wurde in zahlreichen Studien eindeutig nachgewiesen und steht weitgehend außer Zweifel, daß die wirtschaftliche Situation eines Landes einen großen Einfluß auf die Wahlchancen von Regierung bzw. Opposition hat (vgl. etwa Frey/Schneider 1988). Dies bedeutet aber nicht, daß der Stimmenanteil einer Partei deswegen aus ökonomischen Variablen berechenbar ist. Viele weitere Faktoren spielen bei der Wahlentscheidung der Bürger eine Rolle. Polit-ökonomische Modelle werden nie alle diese Faktoren berücksichtigen können und sind daher wohl zu unflexibel für eine verlässliche Wahlprognose. Insgesamt haben sich die Modelle als nicht sehr robust erwiesen, so daß wir uns im folgenden auf den Vergleich der zwei erfolgreichsten Prognoseinstrumente der letzten Jahre konzentrieren: Wahlbörsen und Umfragen.

² je nach Modellspezifikation wurde einmal ein Sieg von Clinton, das andere mal der Triumph Dole's vorausgesagt, was den Verdacht der Beliebigkeit der Ergebnisse verstärkt

7.2 Beeinflussen Umfragen die Wahlentscheidung?

Warum gibt es überhaupt Wahlforschung? Ein wichtiger Grund ist sicher, daß Medien und Parteien schon vor einer Wahl wissen wollen, wie die Chancen der einzelnen Gruppierungen stehen und auf welche politische Konstellation man sich nach dem Urnengang einstellen kann.³ Dies ist ein verständliches Anliegen, und so ist die Meinungsforschung heute aus dem politischen Prozeß kaum noch wegzudenken. Parteien und Politikern können Umfragen zu speziellen Themen als Richtschnur ihrer politischen Handlungen dienen bzw. kann mit diesem Instrument auch im Nachhinein die Zustimmung zu einer beschlossenen Maßnahme erhoben werden.

Uns interessieren hier aber nicht Umfragen zu verschiedenen Problemstellungen, sondern nur Erhebungen zur Prognose eines zukünftigen Wahlergebnisses. Auf den folgenden Seiten wird erläutert, wie solche Befragungen durchgeführt werden, wie Vorhersagen erstellt werden und welche Probleme dabei auftreten können. Eine Frage die vorab zu klären ist, ist jene nach der Relevanz von Umfragen – haben diese einen Einfluß auf das Verhalten der Wähler?

Insbesondere in den USA, aber auch in Europa stellt die Mobilisierung von Sympathisanten den Schlüssel zum Wahlerfolg einer Partei dar. In den Vereinigten Staaten geht es fast nur darum, eigene potentielle Wähler zu den Urnen zu bringen. Das „Umstimmen“ der Anhänger anderer Parteien bildet eher die Ausnahme.⁴ Doch auch in Europa gibt es stets ein großes Reservoir an Nichtwählern (meist 20 bis 30 Prozent aller Wahlberechtigten). Jener Kandidat, dem es gelingt, seine Sympathisanten am besten zu mobilisieren, gewinnt in der

³ wie das Geschehen in Österreich nach der Nationalratswahl vom 3. Oktober 1999 zeigte, können selbst die besten Umfragen nicht vor Überraschungen bei einer nachfolgenden Regierungsbildung schützen

⁴ dies gelang beispielsweise Ende der 60er Jahre Nixon im bis dahin traditionell demokratischen Süden der USA, der seit damals überwiegend republikanisch wählt. Auch Reagan gelang es, viele Demokraten auf seine Seite zu ziehen (sogenannte *Reagan-Democrats*)

Regel auch die Wahl.⁵ Neben jeder Form der Wahlwerbung spielen bei der Wählermobilisierung auch Umfragen eine Rolle.

Wird glaubhaft versichert, daß jede Stimme für das Ziel einer Partei (bspw. den Parlamentseinzug bei Kleinparteien, „Platz Eins“ bei größeren Gruppierungen) entscheidend sein kann, so gehen die Anhänger dieser Gruppierung eher zur Wahl. Dies bezeichnen wir hier als ‚Mobilisierungseffekt.‘

Beispielsweise gelang es Al Gore und seinem Team, in den letzten Wochen vor der Präsidentenwahl 2000 viele Sympathisanten zu mobilisieren und so landesweit mehr Stimmen zu bekommen als sein Kontrahent, obwohl er wenige Wochen vor der Wahl noch zehn Prozent hinter Bush gelegen hatte. Ebenso verdankt es die ÖVP wohl einer guten Wählermobilisierung, bei der Nationalratswahl 1999 nur auf 26,91 % der Stimmen abgerutscht zu sein, nachdem Umfragen wesentlich größere Verluste prophezeit hatten.

Problematisch wird die Situation dann, wenn Umfrageergebnisse von Parteien instrumentalisiert werden, um eigene Wähler zur Stimmabgabe zu motivieren. Immer wieder tauchen Vorwürfe auf, einzelne Institute stünden einer Partei zu nahe und würden die Zahlen in deren Sinn „trimmen“.

Ähnlich wie beim Zauberlehrling hat man mit Umfragen ein grundsätzlich sinnvolles Instrument geschaffen, das sich jedoch als sehr mächtig – und damit interessant für Manipulatoren – erwiesen hat. Wie wir weiter unten noch sehen werden, können zu große Verzerrungen durch eine Art „sozialer Kontrolle“ der Institute weitgehend ausgeschlossen werden, doch selbst kleine „Korrekturen“ sind bereits zu kritisieren.

⁵ wie drastisch sich eine geringe Wählermobilisierung auswirken kann wurde jüngst durch den Wahlsieg des radikalen Likudführers Sharon in Israel demonstriert. Er erreichte bei 62 Prozent Wahlbeteiligung 63 Prozent der Stimmen – wurde also von 39 Prozent aller Wahlberechtigten gewählt. Dies entspricht fast exakt dem Anteil den sein Parteifreund Netanjahu zwei Jahre zuvor erzielte (38 Prozent aller Wahlberechtigten, 46 Prozent der abgegebenen Stimmen), damals jedoch bei 82 Prozent Wahlbeteiligung an Barak scheiterte. 2001 blieben die Wähler Baraks den Urnen fern und verursachten damit den Sieg Sharons

George Gallup, der Begründer der modernen Umfrageforschung sagte daß es die Mission von Umfragen sein müsse, die Stimme des Volkes hörbar zu machen, nicht sie zu verzerren (vgl. Gallup, Internet). Doch selbst eine unverzerrte Umfrage kann einen verzerrenden Einfluß auf die Öffentlichkeit haben.

Jeder Einzelne macht seine Wahlentscheidung bewußt und unbewußt auch von seinen Erwartungen abhängig. Glaubt ein Wähler, daß eine bestimmte Partei eine wichtige Funktion im Staat zu erfüllen hat, und befürchtet Verluste oder gar das Scheitern der Partei an einer Mindest-Prozenthürde, so wird er sehr wahrscheinlich dieser Partei seine Stimme geben.⁶ Dies sowie die Angst vor den Folgen einer zu starken FPÖ dürfte der ÖVP bei der Wahl 1999 viele Stimmen gebracht haben.

Einer der bekanntesten Effekte, die aus Erwartungen resultieren, ist der Mitläufereffekt (*bandwagon effect*). Jeder ist lieber bei den Siegern als bei den Verlierern, daher kommt es vor, daß viele Menschen dem erwarteten Wahlgewinner ihre Stimme geben und so für einen unerwartet hohen Triumph dieser Partei sorgen.

Umfragen können einen enormen Einfluß auf die Entstehung dieser Erwartungen haben, denn die Wahlforschung beeinflußt höchstwahrscheinlich den Ausgang der Wahl, deren Ergebnis sie vorherzusagen trachtet. Einen derartigen Zusammenhang bezeichnet man als selbstreflexives System, denn die Wahlforschung will etwas (Wahlverhalten) messen, das zum Zeitpunkt der Messung noch gar nicht feststeht bzw. sogar vom Ergebnis der Messung (der Prognose) abhängig ist. Damit wird eine Umfrage zum Grenzfall einer *self fulfilling prophecy*.

Die Vermutung, daß Umfragen selbst einen Einfluß auf den Ausgang einer Wahl haben können, macht es umso wichtiger zu klären, wie zuverlässig dieses Instrument bzw. die dahinterstehenden Institute sind, ob es zu Manipulationen kommt und ob bzw. wie diese verhindert werden können.

⁶ dies wird bisweilen etwas irreführend als „Mitleidseffekt“ bezeichnet

7.3 Eine kurze Geschichte der Umfragen

Systematische Datenerhebungen reichen weit in die Geschichte zurück. Die Bibel berichtet wiederholt von Volkszählungen bei den Juden und Römern.⁷ Auch bei den Ägyptern und in China gab es derartige Erhebungen schon vor Jahrtausenden, wobei solche Zählungen stets der Erfassung der Bevölkerung für den Militärdienst oder zur Besteuerung dienten.

Karl der Große verwendete um 800 wohl als erster standardisierte Fragebögen, um von Bischöfen des Reiches untereinander vergleichbare Antworten zu kirchenrechtlichen Fragen zu bekommen (vgl. Noelle-Neumann 1996, S. 38). Damit war der Kaiser seinen Zeitgenossen um Jahrhunderte voraus, denn erst ab dem frühen 17. Jahrhundert wurden statistischen Daten nicht nur für administrative Zwecke, sondern auch für die Sozialwissenschaften verwendet.

Die beiden Engländer Graunt und Petty begründeten um 1660 die „Politische Arithmetik“, die sich vorwiegend der quantitativen Messung und Beschreibung sozialer Vorgänge verschrieb. Dabei wurde bereits auf die Methode der repräsentativen Stichprobenziehung zurückgegriffen, wenn die Grundgesamtheit zu groß wurde. Süßmilch wendete diese Methoden später in Preußen an und erweiterte sie um soziologische Analysen und Ursachenforschung.

Die Zahlenlastigkeit der Politischen Arithmetik führte bald zu einer Gegenbewegung, die sich „Universitätsstatistik“ nannte und Ende des 17. Jahrhunderts in Deutschland entstand. Diese Forschungsrichtung, als deren Begründer Conring gilt, befaßte sich überwiegend mit der qualitativen Beschreibung sozialer Phänomene. Nach einer langen Entwicklungsphase wurden erst 100 Jahre später Achenwalls und Schlözer zu den bedeutendsten Vertretern dieses Wissenschaftszweigs (vgl. Diekmann 1995, S. 77ff.).

Vertreter der beiden Richtungen bekämpften sich bald verbal und in Publikationen und genau genommen tun sie (bzw. ihre Nachfolger) das bis heute, denn es wird weiterhin eifrig darüber gestritten, ob quantitativer oder qualitativer Forschung der Vorzug zu geben ist.

⁷ vgl. dazu 4. Buch Mose, 2, 32; 2. Buch Samuel, 24 sowie Lukas, 2, 1-3

Eine wesentliche Weiterentwicklung der empirischen Sozialforschung gelang dem belgischen Astronom und Physiker Quetelet, der um 1830 die damals relativ neue Wahrscheinlichkeitstheorie mit den bestehenden Methoden der Politischen Arithmetik vereinte. Die damit verbundene Anwendung der Gaußschen Glockenkurve auf Daten der Sozialstatistik ist eine der Grundlagen repräsentativer Umfragen

Redakteure der amerikanischen Zeitschrift „Literary Digest“ erkannten als Erste, daß dieses Instrumentarium auch zur Prognose eines Wahlergebnisses dienen könnte und veranstalteten daher in der Zwischenkriegszeit regelmäßig Wahlumfragen (*polls*) vor den Präsidentschaftswahlen. Dazu wurden Millionen von Probestimmzetteln mit der Bitte um Rücksendung an Wahlberechtigte verschickt.⁸ Die Methode war zwar sehr aufwendig und teuer, doch gelang es damit, bis 1936 jeweils den späteren Sieger der Wahl zu prognostizieren.

1936 versuchte der Forscher George Gallup konkurrierend dazu einen neuen Ansatz: er wandte die in demographischen Untersuchungen bewährte Methode der Stichprobe an, um eine Wahlprognose zu erstellen. Während „Literary Digest“ den Republikaner Landon eindeutig als Sieger sah, tippte Gallup auf eine Wiederwahl Roosevelts. Dabei hatte die Zeitschrift 10 Millionen Probestimmzettel versandt und 2,4 Millionen Antworten erhalten. Gallup griff hingegen auf eine kleine Quotenstichprobe zurück und konnte damit das tatsächliche Wahlergebnis sehr gut vorhersagen. „Literary Digest“ unterschätzte dagegen den Stimmenanteil Roosevelts um 19 Prozent, was den Untergang dieser aufwendigen Methode der Wahlprognose bedeutete.

Was hatte die Zeitschrift falsch gemacht? Grundsätzlich ist einer großen Stichprobe der Vorzug gegenüber einer kleinen wie der Gallups zu geben, doch darf es zu keinen krassen Verzerrungen gegenüber der Grundgesamtheit kommen. „Literary Digest“ griff für seine Erhebung auf die Adressen aus „Telephon und

⁸ die Zeitschrift erreichte dabei meist eine Rücklaufquote um 25 Prozent, was für schriftliche Erhebungen relativ gut ist

Auto“ zurück, doch damals besaß noch nicht jeder diese (Luxus)Güter. Daher wurden v.a. Wähler der gut situierten Mittelschicht erfaßt – und diese wählten überwiegend den Republikaner Landon, während die Mehrheit der Arbeiter Roosevelt den Vorzug gab. Die Auswahl der Befragten war also stark selektiv und damit verzerrend, was schließlich zur schlechten Prognose führte.

Nach diesem ersten großen Erfolg wurde die Gallup-Technik bzw. der Name „Gallup“ schnell zum Inbegriff von Meinungsumfragen schlechthin. Seine Methode der Verwendung relativ kleiner Stichproben hat sich weltweit in der Wahlforschung durchgesetzt. Dies, obwohl es bereits 1948 einen spektakulären und folgenreichen Fehlschlag gab: der demokratische Präsident Truman war nach seiner ersten Amtszeit nicht sehr populär und lag in allen Meinungsumfragen hinter seinem republikanischen Herausforderer Dewey. Am Tag nach der Wahl warteten viele Zeitungen (u.a. die New York Times) daher nicht mehr auf das offizielle Wahlergebnis, sondern verkündeten auf ihren Titelseiten schon vorab den Sieg Deweys. Umso größer war die Überraschung – und die Blamage – als Truman mit respektablem fünf Prozent Vorsprung im Amt bestätigt wurde (vgl. Diekmann 1995, S. 342).

Alle Umfrageinstitute hatten mit dem bis dahin bewährten Quotenverfahren gearbeitet, doch waren offenbar innerhalb der Quoten überwiegend republikanische Wähler befragt worden, was zu Verzerrungen führte. Dies gab den Befürwortern einfacher Zufallsstichproben ohne Quoten in der Folge starken Auftrieb.

Doch das Fiasko hatte noch weitere Konsequenzen: seit damals ist in den USA nur noch die Veröffentlichung der Rohdaten einer Erhebung erlaubt. Abgesehen von einer Gewichtung, um die demographische Repräsentativität sicherzustellen, ist jegliche Bearbeitung der Daten untersagt. Daher erklären selbst Umfrageinstitute in den USA heute, daß ihre Erhebungen keine Prognose des Wahlergebnisses, sondern lediglich Momentaufnahmen der politischen Stimmung sind (vgl. Forsythe et al. 1992, S. 1149).

Nach 1948 beschritt die Umfrageforschung in den USA und Europa unterschiedliche Wege. In den USA war, wie erwähnt, nur noch die Publikation von

Rohdaten erlaubt, so daß die Institute versuchten, diese immer besser und zuverlässiger zu erheben.

In Europa ging die Entwicklung in eine andere Richtung. Nach der Erhebung der Rohdaten ist der Wissenschaftler meist mit einem signifikanten Anteil an Antwortverweigerungen (üblich sind 20-30 Prozent) und Unentschlossenen (oft bis zu 50 Prozent) konfrontiert. Diese werden anschließend hochgerechnet, um eine Prognose für den Wahlausgang zu erhalten. Die dabei verwendeten Algorithmen und Verfahren werden von den Instituten meist geheim gehalten, weil dabei üblicherweise Erfahrungswerte und Hypothesen verwendet werden, die der Konkurrenz nicht bekannt werden sollen. Dadurch wird die Prognoseerstellung für Außenstehende jedoch undurchschaubar und in der Folge unglaubwürdiger. Da die Vorhersagen unterschiedlicher Institute zudem oft signifikant voneinander abweichen, werden immer wieder Spekulationen über Manipulationen und zu großer Parteinähe der Institute laut. Selbst wenn diese meist unbegründet sind, bleiben die Mechanismen hinter der Erstellung von Vorhersagen doch unbekannt oder zumindest unscharf.

Bevor wir diese Problematik genauer diskutieren, wenden wir uns einer wichtigen Eigenschaft von Umfragen, der Repräsentativität, sowie deren theoretischer Fundierung durch die Statistik zu.

7.4 Repräsentativität: Auswahlverfahren für Stichproben

Es gibt unzählige Möglichkeiten, Personen für eine Befragung auszuwählen, wobei für verschiedene Fragen unterschiedliche Methoden zum Einsatz kommen. Bei der Festlegung des Auswahlmechanismus sind daher vorab einige Fragen zu klären. Für Umfragen zur Wahlprognose sind a priori nur jene Methoden sinnvoll, die den Anspruch erheben, für die Grundgesamtheit (bspw. alle Wahlberechtigten Österreichs) gültig und repräsentativ zu sein. Sehen wir uns an, welche Möglichkeiten der Auswahl zur Verfügung stehen.

Die erste Frage, die geklärt werden muß, ist, ob alle Mitglieder der Grundgesamtheit befragt werden können bzw. sollen. Werden alle befragt, so spricht man von

einer Vollerhebung, andernfalls von einer Teilerhebung, wobei Umfragen vor Wahlen immer als Teilerhebungen (Stichproben) durchgeführt werden.

Im nächsten Schritt muß entschieden werden, welche Individuen der Bevölkerung befragt werden sollen. Erfolgt diese Ziehung völlig frei von Regeln, so spricht man von einer willkürlichen Auswahl. Nach dieser Methode wäre es beispielsweise gerechtfertigt, wenn ein Interviewer in einen Kindergarten gehen und einige Kinder befragen würde, um das Wahlverhalten der Gesamtbevölkerung zu erkunden. Dies ist offensichtlich nicht zielführend.

Gewisse Regeln (bspw. daß nur wahlberechtigte Personen befragt werden) sind daher notwendig, um ein gültiges Ergebnis zu erhalten. Basieren diese Regeln auf einem Zufallsprozeß, so spricht man von einer Zufalls- oder Wahrscheinlichkeitsauswahl, andernfalls von einer bewußten Auswahl. Letztere ist klarerweise sehr anfällig für Manipulationen bzw. nur schwer mit wissenschaftlicher Unabhängigkeit zu vereinbaren. Daher konzentrieren wir uns in der weiteren Analyse auf die Wahrscheinlichkeitsauswahl.

Jedes der bisher erwähnten Auswahlverfahren hat seine Berechtigung, je nach Fragestellung mag das eine oder andere Verfahren geeignet sein.⁹ Für eine Umfrage über das Wahlverhalten, die zudem repräsentativ für alle Wahlberechtigten sein soll, ist jedoch einzig eine Wahrscheinlichkeitsauswahl zulässig, weil die zufällige Auswahl der Teilnehmer für Repräsentativität sorgt (die Begründung hierfür liefert die Statistik – vgl. dazu den nächsten Abschnitt 7.5 Statistische Grundlagen für den Einsatz von Stichproben).

Das Ergebnis einer Wahrscheinlichkeitsauswahl bezeichnet man als Zufallsstichprobe; diese können wiederum in einfache und mehrstufige unterschieden werden.

⁹ so ist für die Erhebung von Konsumpräferenzen nichts gegen eine willkürliche Auswahl in einem Supermarkt einzuwenden, eine Vollerhebung mag das Betriebsklima in einem kleinen Unternehmen am besten erfassen

Bei einfachen Zufallsstichproben (*simple random samples*) wird aus der bewußt abgegrenzten Grundgesamtheit (z.B. alle wahlberechtigten Österreicher) in einem Schritt zufällig eine bestimmte Anzahl an Personen „gezogen“, wobei jedes Individuum dieselbe Wahrscheinlichkeit hat, in der Stichprobe zu sein. Dies kann direkt – etwa durch zufälliges Auswahl und Ansprechen auf der Straße, oder indirekt über „Stellvertreter“ – bspw. Karteikarten oder eine Liste aller Wahlberechtigten – erfolgen.¹⁰ Die wesentlichen Vorteile dieser Methode liegen in ihrer methodischen Einfachheit und im vergleichsweise geringen Informationsbedarf über Merkmale der Grundgesamtheit.¹¹

Im Gegensatz dazu wird bei einer mehrstufigen Zufallsstichprobe (*multi-level random sample*) die Grundgesamtheit nach geeignet erscheinenden Kriterien (Wahlbezirke, Gemeinden, etc.) in Gruppen (*primary sampling units*, PSUs) untergliedert. In einem ersten Schritt werden aus diesen PSUs zufällig einige gezogen, aus denen anschließend die zu befragenden Individuen (*second stage units*, SSUs) wiederum zufällig ausgewählt werden (für Details siehe Schumann 1990, S. 88ff.).¹²

Noch einen Schritt weiter gehen Quotenverfahren, bei denen die Stichproben nach vorgegebenen Quoten gezogen werden. Dabei wird die Bevölkerung nach geeignet erscheinenden Kriterien (Alter, Geschlecht, Beruf, Einkommen, etc.) eingeteilt. Die Interviewern müssen nun die zu befragenden Personen so auswählen, daß in ihrem Sample die Quoten der Gesamtbevölkerung widergespiegelt werden (z.B. 52 Prozent Frauen, 10 Prozent Studenten, etc.), um zuverlässig ein „verkleinertes Abbild der Grundgesamtheit“ zu erhalten.

¹⁰ bei sehr großen Grundgesamtheiten kommt fast nur die Auswahl über Stellvertreter zur Anwendung. Dabei ist darauf zu achten, daß für jedes Element der Grundgesamtheit die gleiche Wahrscheinlichkeit besteht, „gezogen“ zu werden

¹¹ da keine Gliederung der Grundgesamtheit folgt, benötigt ein Forscher auch keine weiteren Daten wie etwa Alter, Beruf, Wohnsitz, etc.

¹² dies ist der Vorgang bei einer zweistufigen Auswahl, es sind auch mehr Stufen denkbar, wodurch sich aber das Prinzip nicht ändert

Der Nachweis, daß bestimmte Merkmale in der Stichprobe mit derselben Häufigkeit auftreten, wie in der Grundgesamtheit, reicht aber noch nicht aus, um zu beweisen, daß eine Erhebung alle interessierenden Merkmale korrekt abbildet. Dies träfe nur dann zu, wenn alle anderen Merkmale innerhalb der Quotengruppen vollständig homogen verteilt wären, was jedoch realistischerweise kaum zutreffen wird. Dieses Verfahren setzt daher ein sehr detailliertes Vorwissen über die Grundgesamtheit voraus, womit es nur noch begrenzt als Zufallsauswahl anzusehen ist (vgl. dazu Schnell 1995, S. 286ff.).

Heute werden in der Wahlforschung für persönliche Interviews fast ausschließlich mehrstufige Zufallsstichproben und Stichproben nach Quoten eingesetzt, während bei telefonischen Befragungen auch die einfache Zufallsauswahl vorkommt. Die verschiedenen Institute geben dabei unterschiedlichen Methoden den Vorzug, so daß allgemeingültige Aussagen nur eingeschränkt möglich sind.¹³ Für unsere Betrachtungen ist dies jedoch nicht von zentraler Bedeutung, weil jede dieser Methoden – teils mit Einschränkungen – als repräsentativ gilt. Was bedeutet nun eigentlich „repräsentativ“, und wie wird begründet, daß Zufallsstichproben repräsentativ sind?

Schumann definiert den Begriff „repräsentative Umfrage“ als

„...eine Umfrage, die auf einer einfachen Zufallsstichprobe, einer mehrstufigen Zufallsstichprobe mit PPS-Design¹⁴ und gegebenenfalls Transformationen, auf einer proportional geschichteten Zufallsstichprobe oder auf einer disproportional geschichteten Zufallsstichprobe, die durch Designgewichtung proportionalisiert wurde, beruht.“ (vgl. Schumann 1990, S. 84)

¹³ so arbeitet in Österreich Gallup mit einfachen Zufallsstichproben, während OGM und IMAS auf eine mehrstufige, auf Quoten beruhende, Zufallsauswahl vertrauen

¹⁴ PPS steht für *probability proportional to size*, also für Stichproben, in denen die Wahrscheinlichkeit, daß eine Ausprägung „gezogen“ wird, deren Anteil in der Grundgesamtheit entspricht (vgl. Schumann 1990, S. 90)

Diese umständliche Definition kann man auch abkürzen und feststellen, daß eine Umfrage dann und nur dann als repräsentativ zu betrachten ist, wenn die Teilnehmer zufällig und frei von Verzerrungen ausgewählt bzw. wenn Verzerrungen ausreichend berücksichtigt und korrigiert wurden.

Ähnlich argumentiert auch Schnell, der meint:

„Zufallsstichproben stellen die einzige Gewähr dafür dar, daß aus Ergebnissen einer Stichprobe in bezug auf die Verteilung aller Merkmale (innerhalb bestimmter statistischer Fehlergrenzen) auf die Verteilung dieser Merkmale in der Grundgesamtheit geschlossen werden kann. Ein solcher „Repräsentationsschluß“ kann also nur gezogen werden, wenn der Auswahlmechanismus eine Zufallsauswahl ist. Die Bezeichnung einer Stichprobe als ‚repräsentativ‘ ist somit nur im Sinn des Prinzips der Zufallsauswahl zu verstehen: beide Begriffe sind im obigen Sinn synonym.“ (vgl. Schnell 1995, S. 286)

Das wissenschaftliche Fundament für die Behauptung, daß Zufallsstichproben repräsentativ sind, liefert die induktive Statistik, mit deren Hilfe man je nach Stichprobenumfang ein Konfidenzniveau für die Zuverlässigkeit eines Ergebnisses angeben kann.

7.5 Statistische Grundlagen für den Einsatz von Stichproben

Es ist für unsere Diskussion unverzichtbar, die wissenschaftliche Begründung für die unterstellte Repräsentativität von Stichproben zu hinterfragen, jedoch soll dies nicht zu einer langatmigen Abhandlung über Statistik führen. In diesem Kapitel werden lediglich die Prinzipien und die wichtigsten Überlegungen dargelegt, während jene, die sich eingehender mit dem Thema beschäftigen wollen auf spezielle Literatur zurückgreifen sollten.¹⁵

¹⁵ zu den generellen statistischen Grundlagen empfiehlt der Verfasser Bortz 1999, für eine gute Behandlung der Fehlerquellen bei Umfragen die hervorragende Arbeit von Schumann 1990

Für die Analyse betrachten wir eine einfache Zufallsstichprobe. Am Ende des Abschnitts wird dann noch kurz darauf eingegangen, inwiefern die Überlegungen auch für mehrstufige oder quotenbasierte Zufallsauswahlen gelten.

Das Wahlverhalten kann als dichotome Variable, präziser als Abfolge von dichotomen Variablen – also Variablen, die nur zwei Ausprägungen (ja/nein, 0/1) haben – gesehen werden.¹⁶ Mit den Informationen einer Stichprobe sollen dabei die entsprechenden Anteilswerte an der Grundgesamtheit geschätzt werden. Bestimmen wir zuerst die wichtigsten Variablen; zum besseren Verständnis werden beispielhafte Zahlen in Klammern angeführt

N (500) sei der Umfang einer Stichprobe und n (200) die Anzahl der Individuen in der interessierenden Kategorie (Wähler einer Partei). Ist P der tatsächliche Wähleranteil einer Partei, so bezeichnen wir p ($0,4 = 40$ Prozent) als Stichprobenschätzwert dieses Anteils, der sich als Quotient aus n und N ($200/500 = 0,4$) errechnet.

Selbst wenn von allen praktischen Problemen der Umfrageforschung abgesehen wird, ist dieser Schätzwert aus statistischen Überlegungen als unsicher anzusehen. Die entscheidende Frage ist dabei, wie unsicher dieser Wert ist.

Es ist einsichtig, daß zufällig überdurchschnittlich viele Wähler der Partei unter den Befragten sein könnten – ebenso wie das Gegenteil möglich ist. Ähnlich wie beim N -maligen Münzwurf die Anzahl n von „Kopf“ (bzw. der Anteil) binomialverteilt ist (für $P=0,5$), so ist auch hier die Stichprobenverteilung eine Binomialverteilung. Diese nähert sich für größere N sehr schnell einer Normalverteilung an, so daß wir in der Umfrageforschung getrost mit dieser Verteilung arbeiten dürfen.

¹⁶ Die Frage zur Parteipräferenz „Welcher Partei würden Sie am ehesten Ihre Stimme geben“ kann aufgliedert werden in mehrere dichotome Ja/Nein-Fragen: „Würden Sie am ehesten der SPÖ Ihre Stimme geben?“, „Würden Sie am ehesten der ÖVP Ihre Stimme geben“, usw., wobei jeder Wähler nur eine der Fragen mit „Ja“ beantworten würde – genau dies geschieht mit der üblichen Fragestellung, wo nur eine Partei genannt werden kann

Die Verwendung der Normalverteilung unterstellt mehrere Erhebungsvorgänge, deren Ergebnisse die charakteristische Form der Gaußschen Glockenkurve widerspiegeln sollten.¹⁷ Natürlich wird üblicherweise nur eine Stichprobe gezogen, aus welcher dann der Wert p berechnet wird. Daraus läßt sich anhand des Streuungsmaßes der Standardabweichung ein Fehlerintervall der Stichprobenschätzung bestimmen. Man schließt also zunächst von der Grundgesamtheit auf die Wahrscheinlichkeit von Stichprobenergebnissen, um die gewonnenen Informationen sodann in einem Umkehrschluß (Repräsentationsschluß) von der Stichprobe auf die Grundgesamtheit umzulegen. Dabei unterstellt man implizit, daß die Stichprobenanteile p entsprechend einer Gaußschen Glockenkurve um den wahren Wähleranteil P streuen. Ist N größer als 100 so kann der Standardfehler „ σ “ näherungsweise über die Formel

$$\sigma = \sqrt{\frac{p(1-p)}{N}}$$

berechnet werden. In unserem Beispiel wäre σ also die Wurzel aus $0,4 \cdot 0,6 / 1000$, was etwa 1,55 Prozent ergibt.

Bei der Normalverteilung befinden sich innerhalb bestimmter Intervalle um den Mittelwert bestimmte Wahrscheinlichkeitsanteile. Charakteristisch ist, daß innerhalb von $P \pm \sigma$ etwa 68 Prozent der Fälle liegen. Innerhalb von $\pm 1,96\sigma$ sind es 95 Prozent und im Intervall von $\pm 2,56$ Standardabweichungen um den Mittelwert befinden sich 99 Prozent der Fälle. Diese Eigenschaft macht man sich nun zunutze, um die Sicherheit einer Prognose anzugeben, indem man Konfidenzintervalle berechnet. Zwar ist es unmöglich, genau zu sagen, wie ein Wahlergebnis ausfallen wird, aber man kann Vertrauensbereiche angeben, innerhalb derer die erhobenen Werte zu 95% oder 99% die tatsächliche politische Stimmung widerspiegeln. Korrekterweise sollte bei jeder Umfrage die Bandbreite der möglichen Werte mit publiziert werden, was jedoch nur selten geschieht.

¹⁷ bei Wahlen werden die Prozentanteile für mehrere Parteien angegeben, wozu eigentlich eine Multinomialverteilung verwendet werden müßte. Dadurch werden die Fehlerintervalle größer

Um unser Beispiel fortzuführen: p beträgt 40%, der Standardfehler ist 1,55%. Der wahre Stimmenanteil der Partei liegt also mit 95-prozentiger Wahrscheinlichkeit zwischen 37% und 43%.

Die Größe des Vertrauensbereichs hängt primär von der Samplegröße ab, doch auch der Anteil eines Merkmals in der Grundgesamtheit hat Einfluß auf die Genauigkeit einer Schätzung. Die folgenden Tabellen zeigen für verschieden Samplegrößen die entsprechenden Standardfehler und Konfidenzintervalle bei einem Vertrauensgrad von 95 bzw. 99 Prozent, wobei die erste Tabelle für Mittel- und Großparteien gilt, die zweite für Kleinparteien

		N=500	N=1000	N=2000
Anteil	p	30 %	30 %	30 %
Standardfehler	σ	2,05 %	1,45 %	1,02 %
95% Konfidenzi.	$p \pm 1,96\sigma$	26,0-34,0 %	27,2-32,8 %	28,0-32,0 %
99% Konfidenzi.	$p \pm 2,56\sigma$	24,8-35,2 %	26,3-33,7 %	27,4-32,6 %

Tabelle 7: Verschiedene Stichprobengrößen und die entsprechenden Konfidenzintervalle für Mittel- und Großparteien

Je größer das Sample, desto zuverlässiger und präziser kann eine Aussage gemacht werden (vorausgesetzt Verzerrungen des Samples aufgrund seiner Größe können ausgeschlossen werden). Man sieht, daß der Befragungsumfang vervierfacht werden muß, um eine Halbierung des Konfidenzintervalls zu erreichen. Heute sind für seriöse Umfragen Samplegrößen um 1000 normal, wobei jedoch insbesondere Zeitungen häufig mit kleineren Samples arbeiten.¹⁸ Neben dem Befragungsumfang hat auch der Stimmenanteil der Partei Einfluß auf die Größe des Vertrauensbereichs: je kleiner eine Partei ist, desto genauer ist ihr Ergebnis abschätzbar, desto wichtiger werden aber auch Abweichungen.

¹⁸ dies führt zu teils unseriösen Prognosen, was insbesondere von den großen Umfrageinstituten, die ihren guten Ruf gefährdet sehen, kritisiert wird

		N=500	N=1000	N=2000
Anteil	p	6 %	6 %	6 %
Standardfehler	σ	1,06%	0,75 %	0,53 %
95% Konfidenzi.	p \pm 1,96 σ	3,9-8,0 %	4,5-7,5 %	5,0-7,0 %
99% Konfidenzi.	p \pm 2,56 σ	3,3-8,7 %	4,1-7,9 %	4,6-7,4 %

Tabelle 8: Verschiedene Stichprobengrößen und die entsprechenden Konvidenzintervalle für Kleinparteien

Damit können wir festhalten, daß der Schätzfehler einer Umfrage bei steigendem Stimmenanteil einer Partei tendenziell wächst.¹⁹ Der wichtigste Einflußfaktor ist jedoch der Stichprobenumfang, der positiv mit der Meßgenauigkeit korreliert ist.

Wie die Analyse gezeigt hat, wird zur Schätzung eines Anteils in der Grundgesamtheit der entsprechende Anteil in der Stichprobe herangezogen. Dies macht nur Sinn, wenn die Verteilung der betrachteten Merkmale in beiden Populationen annähernd gleich ist, was theoretisch am besten über eine einfache Zufallsauswahl sichergestellt wird. Jede Abweichung von der Zufälligkeit – ob durch die Einführung von Quoten oder mehrstufiger Auswahlverfahren – birgt das Risiko von Verzerrungen und damit unzuverlässigen Ergebnissen in sich. Während mehrstufige Wahrscheinlichkeitsauswahlen bei entsprechend gutem Design einer einfachen Zufallsauswahl noch sehr nahe kommen, weicht man mit Quoten schon sehr weit vom Ideal ab, weswegen dieses Verfahren auch am häufigsten kritisiert wird. Durch Quoten werden jene Verteilungen von Merkmalen, die eine Stichproben bei Zufallsauswahl ohnehin annehmen sollte,²⁰ künstlich erzeugt. Dieser Zwang kann – gemeinsam mit realistischere stattfindenden kleinen „Schummeleien“ der Interviewer – zu Verzerrungen führen, die für das Institut jedoch kaum erkennbar sind.

¹⁹ dieser Zusammenhang kann aus empirischen Daten jedoch kaum belegt werden, denn Schätzfehler treten in der Praxis unabhängig von der Größe einer Partei auf (siehe weiter unten)

²⁰ wie empirische Untersuchungen gezeigt haben, stimmen die Verteilungen und Anteile von Merkmalen in Zufallsstichproben und der entsprechenden Grundgesamtheit meist sehr gut überein

7.6 Umfragearten und deren technische Durchführung

Umfragen können prinzipiell auf drei Arten durchgeführt werden: mit persönlichen Interviews sowie als schriftliche²¹ oder telefonische Befragungen.²² Alle drei Typen werden heute intensiv und oft in der Meinungsforschung eingesetzt, jedoch werden für Umfragen zur Wahlprognose überwiegend persönliche Befragungen und Telefoninterviews verwendet, so daß wir uns bei den weiteren Ausführungen auf diese beiden konzentrieren wollen.²³ Sehen wir uns die beiden für unsere Überlegungen interessanten Instrumente kurz genauer an, wobei wir primär auf deren Brauchbarkeit hinsichtlich der für uns relevanten „Sonntagsfrage“ bezug nehmen. Die Analyse soll dabei so kurz und prägnant wie möglich gehalten werden, weil es nur um das Aufzeigen der generellen Problematik, nicht jedoch um deren detaillierte Behandlung geht.²⁴

7.6.1 Persönliche Interviews (*face-to-face* Interviews)

Die mündliche Befragung wird von vielen als die zuverlässigste und aussagekräftigste Methode der Datenerhebung in der Meinungsforschung angesehen, weil hier die Ausschöpfungsquote traditionell am höchsten ist und der Befrager individuell auf jeden Interviewten eingehen kann. Dies mag an sich stimmen, muß jedoch relativiert werden, wenn man die Realität der Theorie gegenüberstellt.

²¹ als eine Art der schriftlichen Befragung können auch Umfragen per e-mail gelten, da dabei nur das Medium gewechselt wird, das Prinzip jedoch dasselbe bleibt

²² eine neue Form stellen Umfragen per Internet dar, die jedoch nicht als repräsentativ gelten können, da die Befragten nicht ausgewählt werden, sondern selbst aktiv werden müssen

²³ viele der Ausführungen auf den nächsten Seiten beziehen sich auf Gespräche und e-mail-Kontakt des Verfassers mit Mitarbeitern von Umfrageinstituten in Österreich und Deutschland, ein Verweis auf die entsprechende Literatur erfolgt nur dort, wo dies notwendig erscheint

²⁴ viele klassische Probleme der Umfrageforschung (etwa Abhängigkeiten vom Interviewer, der Situation, Frageformulierung etc.) stellen sich bei der Behandlung der Sonntagsfrage nicht oder nur in geringem Ausmaß, so daß hier nicht näher darauf eingegangen wird. Für Interessierte verweise ich auf das hervorragende Buch von Diekmann aus dem Jahr 1995

Wir haben bereits gesehen, daß für eine Erhebung optimalerweise eine einfache Zufallsstichprobe aus der Grundgesamtheit gezogen werden sollte. Es ist jedoch kaum möglich, in einem Flächenstaat zu vertretbaren Kosten tausend oder mehr persönliche Interviews nach einer einfache Zufallsauswahl durchzuführen. Daher werden meist geschichtete Stichproben durchgeführt, bei denen zuerst zufällig einzelne Städte und Dörfer ausgewählt und anschließend an diesen Orten aus der Bevölkerung einfache Zufallsstichproben gezogen werden. Häufig wird jedoch auch das bereits diskutierte Quotenverfahren angewendet.

Klar ist, daß eine Umfrage mit mündlichen Interviews umso teurer und aufwendiger ist, je zufälliger und breiter gestreut sie durchgeführt wird. Bei der großen Zahl an Umfragen, die mittlerweile stattfinden, sind einfache Zufallsstichproben daher mit einem unvermeidbar hohen finanziellen Aufwand verbunden.²⁵ Durch die Verwendung anderer Auswahlverfahren zur Kostensenkung wird bewußt in Kauf genommen, daß es zu systematischen Verzerrungen kommen kann. Insbesondere durch Quoten werden die Prinzipien der Zufallsauswahl verletzt, was die Vorteile der größeren Zuverlässigkeit durch *face-to-face* Interviews wieder zunichte machen kann.

Besonders bei dieser Art von Befragungen ist auch der Einfluß der Anwesenheit Dritter nicht zu unterschätzen. Überraschend viele Menschen haben Hemmungen, ihre wahre Meinung – oder auch Parteipräferenz – in der Anwesenheit Dritter, speziell von Familienmitgliedern zu nennen (vgl. Diekmann S. 401). Andererseits erlaubt es der persönliche Kontakt bei der Befragung aber, mehr Techniken zur Motivation der Befragten zur Teilnahme anzuwenden. Diese können von einfacher Überredungskunst bis zu kleinen Geschenken reichen.

²⁵ in den USA wurden anlässlich der Präsidentenwahl 2000 1,4 Millionen (!) Umfragen durchgeführt

7.6.2 Telefonische Befragung

Waren *face-to-face* Interviews noch Mitte der 80er Jahre die meist verbreitete Umfragemethode, werden mittlerweile telefonischen Umfragen am häufigsten eingesetzt (vgl. Diekmann, S. 410ff.). Der wohl entscheidende Grund dafür liegt im geringeren – auch finanziellen – Aufwand.²⁶

Insbesondere technische Entwicklungen haben zum Aufschwung dieser Erhebungsmethode beigetragen. Neben der für die Repräsentativität einer Zufallsstichprobe notwendigen Voraussetzung einer nahezu lückenlosen Netzabdeckung, die heute sowohl in den USA als auch in Europa gegeben ist, ist es die Verfügbarkeit aktueller Telefonverzeichnisse auf CD-ROM, welche die schnelle Ziehung von Stichproben erlaubt. Nur Telefonumfragen werden heute regelmäßig als einfache Zufallsstichproben durchgeführt, weil dabei praktisch keine zusätzlichen Kosten entstehen und auch der Personalaufwand nicht höher ist als bei einer anderen Auswahl.

Während Telefonumfragen früher als unzuverlässig, aufgrund hoher Ausfallquoten verzerrt und qualitativ minderwertig angesehen wurden, zeigten empirische Vergleiche in den letzten Jahre, daß diese Befragungen mittlerweile ebenso gute Ergebnisse bringen wie persönliche Interviews.²⁷ Dies kann auch darauf zurückgeführt werden, daß *face-to-face* Interviews oft nach Quoten, telefonischen Befragungen aber als (zuverlässigere) einfache Zufallsstichproben gezogen werden.

Dennoch gibt es auch einige Problemfelder: die Verbreitung von Mobiltelefonen macht die Menschen zwar erreichbar, stellt die Institute jedoch vor die Frage wie sie dies berücksichtigen sollen. Es können reine oder gemischte Stichproben aus Festnetz- und Mobiltelefonbesitzern gezogen werden, ebenso wie man bei

²⁶ laut IHA-GfM Vademecum betragen die Kosten für ein Telefoninterview ca. ein Drittel des entsprechenden Werts bei einem persönlichen Interview (nach Diekmann 1995, S. 430)

²⁷ Gallup gibt die Kooperationsrate bei seinen telefonisch durchgeführten Interviews mit 67 Prozent an, was kaum unter der Quote bei persönlichen Interviews liegt (vgl. Gallup, Internet)

einer Netzdichte von über 90 Prozent weiterhin rein auf das Festnetz zurückgreifen kann. Entscheidend ist dabei die Frage, nach welchen Kriterien eine gemischte Auswahl erfolgen soll.²⁸

Ein weiteres Problem ist, daß die Netzdichte zwar hoch, aber nicht lückenlos ist, und daß jene, die kein Telefon haben, kaum zufällig verteilt sein dürften. Manche Gruppen besitzen kein Telefon (Obdachlose, Anstaltsinsassen), andere sind schlecht erreichbar (Soldaten, Schichtarbeiter), so daß man stets mit Verzerrungen rechnen muß. Zwar lassen sich diese ermitteln und abschätzen, jedoch nimmt mit jedem Eingriff des Instituts die Zufallskomponente ab, was die Repräsentativität der Stichprobe beeinträchtigt.

Ist es bei mündlichen Interviews auf der Straße noch einfach, absolute Anonymität glaubhaft zu machen, ist dies bei telefonischen Befragungen schon schwieriger. Zu eng ist der Zusammenhang von Telefonnummer und Adresse, so daß die Antwortquote insbesondere bei „kritischen“ Fragen bei telefonischen Erhebungen bisweilen erheblich niedriger ist als bei persönlicher Befragung.

Schließlich ist beiden Instrumente gemeinsam, daß ihr Ergebnis von vielen schwer quantifizierbaren Faktoren abhängig ist, selbst wenn man Probleme der Zufallsauswahl außer Acht läßt. Anlaß, Ort, Zeit, Dauer, soziale Rollen und die Art der Beziehung zwischen Befragtem und Befrager sind nur einige der Variablen, die ebenfalls zu berücksichtigen sind.²⁹

²⁸ selbst wenn man ein Gesamtverzeichnis aller Telefonanschlüsse zur Verfügung hat, ergibt sich das Problem von Doppeleinträgen jener, die beide Geräte besitzen, was diesen einer höhere Wahrscheinlichkeit gäbe, gezogen zu werden. Hat man getrennte Verzeichnisse aus denen je eine Zufallsstichprobe gezogen wird, so ist zu klären, welche Anteile aus welcher Gruppe gezogen werden sollen (in Westeuropa besitzen über 90 Prozent der Haushalte einen Festnetzanschluß, Anfang 2001 aber auch schon ca. 60 Prozent ein Mobiltelefon)

²⁹ für eine detaillierte Behandlung dieser Themen wird auf die entsprechende Fachliteratur verwiesen (Schumann 1990, Diekmann 1995, Schnell 1995)

Hier mag ein kurzer Vergleich mit Wahlbörsen angebracht sein: selbstverständlich sind auch bei diesem Instrument die Rahmenbedingungen wichtig (Universität vs. Zeitung als Veranstalter, Laufzeit etc.), aber generell ist deren Gestaltung ziemlich einheitlich und unabhängig vom Veranstalter, weil stets das Internet als einendes Medium und Handelsplattform dient. Durch das fast vollständige Fehlen persönlicher Kontakte ist auch deren Verzerrungswirkung durch personenspezifische Faktoren nicht so kritisch wie bei Befragungen.

7.7 Kritische Punkte der Umfrageforschung

Neben den bereits angesprochenen speziellen Problembereichen bei den einzelnen Umfragearten gibt es einige Bereiche, bei denen die moderne Demoskopie an die Grenzen ihrer Kunst stößt. Diese sind meist unabhängig vom Typ der Befragung und liegen in grundsätzlicheren Schwachstellen begründet.

Grob betrachtet lassen sich drei Hauptfehlerquellen unterscheiden:

- Zufallsfehler der Stichprobe – diese haben wir bereits behandelt
- systematische Fehler aufgrund des Verfahrens der Stichprobenauswahl, auch diese haben wir weiter oben schon betrachtet
- andere Verzerrungen, die meist im Verhalten der Befragten begründet liegen – diese letzte Kategorie von Verzerrungen sind das Thema des nun folgenden Abschnitts, in dem einige der wohl problematischsten Themen der modernen Umfrageforschung diskutiert werden.

7.7.1 Antwortverweigerung und Ausfälle

Eines der Hauptprobleme der Umfrageforschung stellen Antwortverweigerungen und andere Ausfälle dar.³⁰ Oft etwas unscharf, aber meist als Überbegriff für alle

³⁰ in diesen Gruppen werden Ausfälle aus den verschiedensten Gründen zusammengefaßt. Die Gründe für Ausfälle reichen von Todesfällen, Sprach- und technischen Problemen bis zu „Nicht-zugehörigkeit zur Grundgesamtheit“. Antwortverweigerungen können ebenfalls verschiedenste Gründe (Zeitmangel, Desinteresse, prinzipielle Befragungsablehnung) haben. Für uns ist nur die Unterscheidung in „Verweigerern“ (Gefahr von Verzerrungen) und „andere Ausfälle“ (in der

Arten von Antwortverweigerung und Unerreichbarkeit wird auch der Begriff *Nonresponse* verwendet. Einer der wesentlichen Unterschiede zwischen Zufalls- und Quotenstichproben besteht im Umgang mit diesen Ausfällen, so daß wir die Ansätze der beiden Methoden kurz getrennt betrachten.

Zufallsauswahl

Bei einer Zufallsstichprobe – egal ob einfach oder mehrstufig – werden *Nonresponses* meist ignoriert, weil unterstellt wird, daß diese bezogen auf die Grundgesamtheit zufällig sind und somit die Umfrage nicht verzerren. Trifft diese Annahme zu, so ist die (statistisch korrekte) Argumentation zulässig, daß eine Zufallsstichprobe aus einer Zufallsstichprobe wiederum eine Zufallsstichprobe darstellt (vgl. Schumann 1990, S. 87). Solange die Ausfälle tatsächlich zufällig erfolgen, ändert sich an der Aussagekraft der Stichprobe somit nichts.

Die Annahme, daß es so ist, mag auf den ersten Blick berechtigt erscheinen, ist jedoch bei genauerer Betrachtung nicht unproblematisch. Ausfälle erfolgen dann zufällig, wenn jedes ausgewählte Element dieselbe Wahrscheinlichkeit hat, nicht zu antworten. Dabei ist zu unterscheiden in Menschen, die nicht angetroffen werden oder nicht antworten können (Behinderung, Tod) und daher ausfallen, und solche, die nicht antworten wollen (Antwortverweigerer).³¹ In beiden Fällen sind Umstände denkbar, durch die Verzerrungen auftreten können.

Wird eine Umfrage (wie üblich) tagsüber durchgeführt, so werden Berufstätige kaum erreicht, während Hausfrauen und Pensionisten überproportional in der Stichprobe vertreten sein dürften.³² In diesem Fall hätten wir eine Verzerrung durch jene, die nicht antworten können.

Theorie zufällig und damit für die Repräsentativität einer Befragung unerheblich) relevant, wobei oft auch diese Unterscheidung nicht notwendig ist, so daß wir dann nur von „Ausfällen“ bzw. *Nonresponse* sprechen

³¹ der Anteil dieser Ausfälle beläuft sich üblicherweise auf 25-50 Prozent aller ausgewählten Personen, wobei Praktiker einen Anstieg dieser Quote diagnostizieren (vgl. Schnell 1995, S. 289)

³² dieses Argument dient dem IMAS-Institut als Rechtfertigung für die Verwendung von Quoten

Auch ist es vorstellbar, daß die Anhänger einer Partei häufiger keine Antwort geben, als die Parteigänger anderer Gruppierungen, was zu Fehlern aufgrund von willentlichen Antwortverweigerungen führen kann.³³

Es ist evident, daß viele Ausfälle nicht zufällig erfolgen. „Systematische Ausfälle“ sind solche, bei denen Variablen des Untersuchungsgegenstandes (bspw. Parteipräferenz, Einkommen, Beruf) mit den Ursachen des Ausfalls zusammenhängen. Nehmen wir (realistischerweise) an, daß derartige Ausfälle auftreten, so müssen wir davon ausgehen, daß diese zu signifikanten Verzerrungen bei den Ergebnissen führen können. In diesem Zusammenhang ist auch das Problem der Anonymitätssicherung zu nennen. Viele Befragte sind nicht prinzipiell kooperationsunwillig, befürchten jedoch, daß vertrauliche Daten (Einkommen, Ladendiebstahl etc.) an Behörden weitergegeben werden könnten, was dazu führt, daß es bei derartigen Fragen oft hohe Nonresponseraten gibt (zu Relevanz und Ausmaß der Verzerrungen durch *Nonresponse* vgl. Kurz 1987).

Ein Umfrageinstitut kann hier selbstverständlich „glättend“ und „korrigierend“ eingreifen, die Tür zu Willkür und subjektiver Auslegung ist damit jedoch weit aufgestoßen.³⁴ Es ist zu bedenken, daß der Grund für Ausfälle und somit auch die Art der Verzerrung in der Regel unbekannt oder zumindest unsicher ist. Selbst wenn eine Regelmäßigkeit gefunden wurde, ist kaum feststellbar, ob diese zeitlich beständig oder nur vorübergehend ist.

Quotenauswahl

Bei Quotenstichproben tritt das Problem der Ausfälle ebenso auf, kann jedoch besser gemessen und gegebenenfalls korrigiert werden.³⁵ Bei der Auswahl der zu befragenden Individuen wird versucht, jede Gruppe entsprechend ihrem Anteil an

³³ dies konnte in Österreich lange Zeit in Bezug auf Wähler der FPÖ, die sich überdurchschnittlich oft nicht deklarierten, beobachtet werden

³⁴ auf meine Anfrage hin gaben österreichische Umfrageinstitute an, solche Verzerrungen einzurechnen und aufgrund ihrer „langjährigen Erfahrungswerte“ in der Prognose zu berücksichtigen

³⁵ dafür muß bei dieser Art von Befragungen mehr Aufwand zur Datenerhebung betrieben werden

der Grundgesamtheit zu erfassen. Treten nun in einer Gruppe überdurchschnittlich viele Ausfälle auf, so hat das Umfrageinstitut zwei Möglichkeiten, diese Verzerrung zu korrigieren: es können aus dieser Gruppe weitere Individuen zufällig ausgewählt und befragt werden, bis die Responsequote der betreffenden Gruppe jener der anderen Gruppen entspricht. Alternativ können die Teilnehmer der unterrepräsentierten Gruppe in der Gesamtumfrage überproportional gewichtet werden, um das Gleichgewicht wieder herzustellen (dies wird als *Redressment* bezeichnet, vgl. dazu Schnell 1995, S. 290f.).

Diese Verfahren setzen jedoch implizit voraus, daß die *Nonresponses* innerhalb der Gruppe zufällig und unsystematisch erfolgten, was jedoch eine unrealistische und unhaltbare Annahme ist. Wie es scheint, wird durch Korrekturmaßnahmen das Problem der Ausfälle meist nur verdeckt, jedoch nicht gelöst.

Allen Methoden zum „Ausgleich“ von Ausfällen ist bedauerlicherweise gemein, daß sie aufgrund ihrer Willkürlichkeit wissenschaftlich teils unseriös, zumindest aber nicht unproblematisch sind. Relevant für die weiteren Betrachtungen ist, daß Antwortverweigerungen bzw. Ausfälle in einer Stichprobe zu Verzerrungen und somit zu einer Beeinträchtigung der Vorhersagegenauigkeit führen können. Dies ist ein grundsätzliches und kaum behebbares Problem der Demoskopie.

Eines der Hauptziele der Institute ist es daher, die Nonresponsequote zu senken, was primär durch eine Erhöhung der Kontaktversuche erreicht wird. Andere Möglichkeiten sind die Kombination verschiedener Erhebungsmethoden,³⁶ oder die „Bezahlung“ der Befragten.³⁷

Eine weitere, berechtigterweise umstrittene Methode beruht darauf, eine Antwort „abzuschätzen“. Wenn auf eine direkte Frage (bspw. „Welche Partei würden Sie wählen?“) keine Antwort zu erhalten ist, so kann eventuell über die Antworten auf

³⁶ wenn eine Person persönlich nie anzutreffen ist, so ist auch eine Kontaktaufnahme über Telefon oder Brief möglich

³⁷ immerhin leisten diese den wohl wertvollsten Beitrag zum Gelingen einer Umfrage. Da die Institute in der Regel für ihre Arbeit bezahlt werden, ist ihnen auch zuzumuten, die Interviewten zu entlohnen. Allerdings erhöht dies die ohnehin enormen Kosten von Befragungen weiter

ähnliche Fragen (Kanzlerfrage, Entscheidung bei der letzten Wahl, Zustimmung zu politischen Positionen, etc.) versucht werden, das Verhalten des Befragten abzuschätzen. Dies ist jedoch eine Methode, die nur sehr behutsam angewendet werden sollte, da sie mit sehr viel Unsicherheit verbunden ist.

Das Problem der Ausfälle ist eines der wesentlichsten in der Umfrageforschung. Erfreulicherweise sind politische Aktienmärkte davon nicht betroffen, so daß diese hier eine Schwachstelle der klassischen Wahlforschung ausgleichen könnten. Während bei Umfragen der Veranstalter aktiv auf Personen zugeht, um sie zu interviewen, muß sich jeder Teilnehmer an einer Wahlbörse selbst aktiv anmelden, womit naturgemäß das Problem von Ausfällen nicht auftreten kann. Bedenklich ist hier eher der Gegensatz zwischen sehr aktiven und inaktiven Händlern. Dadurch erlangen die aktiven mehr Einfluß auf den Preis, was auf den ersten Blick dem Prinzip der Gleichgewichtung widerspricht.³⁸

Bei genauerer Betrachtung stellt sich jedoch heraus, daß in diesem Ungleichgewicht gerade die Stärke eines politischen Aktienmarktes liegt. Die Intensität des Handels ermöglicht es, die Intensität des Interesses der betreffenden Person an der Wahl auszudrücken, was bisher in der Wahlforschung kaum möglich war. Ebenso wie in der Realität aktive Menschen andere von ihrer Meinung überzeugen können oder sich direkt politisch engagieren, sind es am Markt die aktivsten Händler, die den Preis am meisten beeinflussen.

7.7.2 Aufrichtigkeit der Befragten: Prestigeantworten

Eine Befragung ist nur dann sinnvoll, wenn einige unverzichtbare Grundvoraussetzungen erfüllt sind. Dazu zählt neben der Kooperationswilligkeit v.a. die Aufrichtigkeit der Befragten. Glaubt ein Interviewer, daß die interviewte Person „schwindelt“, so ist das Interview abzubrechen, weil nicht mit zuverlässigen Ergebnissen gerechnet werden kann. Das häufigste Problem stellen dabei sogenannte „Prestigeantworten“ dar.

³⁸ bei einer Umfrage wird jeder Befragte im Sample üblicherweise gleich gewichtet, eine Ausnahme stellen Gewichtungsmaße zwecks Proportionalisierung dar

Umfrageforscher fanden im Laufe der Jahre heraus, daß manche der von ihnen erhobenen Daten nicht mit statistischen Vergleichsdaten übereinstimmten – etwa wenn laut Befragung 150 Kilo Altpapier pro Person gesammelt wurden, der tatsächliche Wert aber nur bei 80 Kilo liegt. Ebenso konnte in einigen europäischen Ländern festgestellt werden, daß die Umfragedaten für weit rechts stehende Parteien oft wesentlich unter deren späterem Wahlergebnis lagen.

Zur Erklärung dieses Phänomens greifen Wissenschaftler auf die Psychologie zurück. Alle Menschen scheinen nach sozialer Anerkennung zu streben und Mißbilligung durch andere vermeiden zu wollen. Dies dürfte ein wichtiger Grund sein, wieso bei Befragungen überdurchschnittlich oft jene Antworten gegeben werden, die von einzelnen oder der Gesellschaft als „sozial erwünscht“ eingestuft werden (man bezeichnet solche Antworten als Prestigeantworten, weil die Befragten mit beschönigenden Auskünften ihr Ansehen steigern wollen).³⁹

Für die Wahlumfrageforschung dürfte dies ein kleineres Problem sein als für andere Felder der Meinungsforschung,⁴⁰ dennoch kommt es auch hier zu Verzerrungen. Die Umfrageforscher versuchen diese abzuschätzen und in ihre Prognose einzubeziehen, jedoch ist dieser Prozeß notwendigerweise subjektiv und wissenschaftlich unzuverlässig. Noch schwieriger wird das Problem, wenn man bedenkt, daß derartige Tabu- oder Prestigeantworten sich mit den Jahren verändern können.⁴¹

³⁹ Diekmann berichtet von einer als Test für Prestigeantworten durchgeführten Umfrage in Sachsen, wo nach der Bekanntheit des Staatssekretärs Dieter Köstritz gefragt wurde. Acht Prozent der Befragten gaben an, von ihm gehört zu haben, wobei insbesondere die Gebildeten den Namen des Politikers als geläufig bezeichneten (18 Prozent der Interviewten mit einem höherem Bildungsabschluß) – insgesamt eine beachtliche Zahl, wenn man bedenkt, daß es Herrn Köstritz gar nicht gibt... (vgl. Diekmann 1995, S. 386)

⁴⁰ so wurden die größten Diskrepanzen bei Befragungen zum Sexual- sowie zum Umweltverhalten beobachten

⁴¹ die Grünen waren in den 80er Jahren „tabu“, und sind heute „in“, was auch deren aktuellen Höhenflug in den Umfragen miterklären könnte

An Wahlbörsen sorgen das Gewinnstreben der Händler und die Anonymität des Marktmechanismus für die Offenlegung der wahren Meinung der Teilnehmer. Anreize zur Offenbarung von ‚Prestigeantworten‘ gibt es nicht.

7.7.3 Aufrichtigkeit der Interviewer

Nicht nur die Wahrheitstreue der Befragten, auch jene der Befrager spielt für den Erfolg einer Erhebung eine maßgebliche Rolle. Dabei ist Praktikern klar, daß es immer wieder zu „Schummeleien“ kommt. Deren Ursache liegt meist nicht in einer beabsichtigten Beeinflussung des Ergebnisses der Umfrage, sondern in einer Zeitersparnis für den Interviewer. Insbesondere bei persönlichen Befragungen sind die Interviewer weitgehend unbeaufsichtigt vom Institut „im Feld“, um Daten zu erheben. Dabei ist zu bedenken, daß viele Institute überwiegend Studenten, die nebenbei etwas Geld verdienen wollen, als Befrager einsetzen. Meist werden diese nach der Anzahl der fertiggestellten Interviews bezahlt, was für diese Anreize schafft, in möglichst kurzer Zeit möglichst viele Menschen zu befragen.

Während einfache Zufallsstichproben meist über Telefon durchgeführt werden, und kaum Möglichkeiten zur Zeitersparnis einräumen, ist es bei *face-to-face* Interviews einfach, Befragungs- und Suchzeit zu reduzieren. Insbesondere Quoten, die verlangt werden, lassen sich leicht erfüllen, wenn Alter, Beruf und Lebensstellung der Befragten falsch angegeben werden.⁴²

Ebenso läßt sich die Befragungszeit reduzieren, wenn nur einige Kernfragen gestellt werden und der Rest des Fragebogens vom Interviewer in „kreativer Schreibtischarbeit“ ausgefüllt wird. Zwar kann nicht eindeutig gesagt werden, ob dieses Problem Umfragen systematisch verzerren, dennoch ist von den Instituten darauf zu achten, daß ihre Interviewer sich an die vorgegebenen Richtlinien halten.

⁴² ist es beispielsweise für den Interviewer schwierig, einen verwitweten Pensionisten zu finden, so kann er auf seinen glücklich verheirateten Großvater zurückgreifen und ihn „zwangs-verwitwen“

Wir wollen im Rahmen dieser Arbeit nicht näher in diese Problematik und mögliche Lösungsansätze dafür eingehen. Es sollte lediglich aufgezeigt werden, daß Umfragen abgesehen von technischen und statistischen Anforderungen auch noch weitere Problembereiche aufweisen, was bei der Betrachtung von Ergebnissen stets im Gedächtnis sein sollte.

Fast alle Wissenschaftler und Mitarbeiter von Umfrageinstituten sind sich einig, daß die mittels einer Befragung erhobenen Rohdaten das spätere Wahlergebnis nicht perfekt widerspiegeln. Zu viele unwägbarere Einflüsse verzerren die Stichprobe, so daß meist eine Bearbeitung der Daten erfolgt, um zur Prognose zu gelangen. Sehen wir uns diesen Prozeß etwas genauer an.

7.8 Verarbeitung der Rohdaten zur Prognose

Die vom Umfrageinstitut erhobenen Befragungsergebnisse werden als Rohdaten bezeichnet, womit bereits angedeutet wird, daß es noch zu einer Verarbeitung der Daten kommt, bevor eine Wahlprognose vorliegt.⁴³ Insbesondere jene Wähler, die ihre Wahlabsicht nicht deklariert haben, müssen zugeordnet werden. Dabei gilt es abzuwägen, wieviele dieser Wähler schließlich nicht zur Wahl gehen und wie sich die anderen auf die einzelnen Parteien verteilen werden. Doch auch andere Verzerrungen werden von den Instituten mehr oder minder bei der Erstellung einer Prognose berücksichtigt.⁴⁴

Wer hat sich nicht deklariert? Was werden die Unentschlossenen wählen? Wie hoch wird die Wahlbeteiligung sein? Welcher Partei wird es gelingen, ihre Wähler optimal zu mobilisieren?

⁴³ die folgenden Ausführungen beziehen sich auf Umfragen in Europa, in den USA werden aufgrund der erwähnten gesetzlichen Bestimmung stets Rohdaten veröffentlicht

⁴⁴ Diekmann meint dazu treffend „Die ... Rohwerte ... werden von den Meinungsforschungsinstituten mit mehr oder minder obskuren Gewichtungungsverfahren korrigiert.“ (vgl. Diekmann 1995, S. 389)

So oder so ähnlich lauten die Fragen, die ein Umfrageforscher beantworten muß, um aus Rohdaten eine Prognose zu erstellen. Es hat sich gezeigt, daß eine einfache Hochrechnung der Rohdaten nicht zielführend ist. Beispielsweise war es in Österreich lange so, daß überdurchschnittlich viele spätere Wähler der FPÖ ihre Wahlabsicht nicht bekanntgaben, während die Grünen als „Umfragekaiser“ bisweilen sogar höhere Werte in den Rohdaten erhielten, als dann bei der Wahl.⁴⁵ Berücksichtigt ein Umfrageforscher derartige Faktoren (gesteht also der FPÖ viele der unentschlossenen Wähler, den Grünen hingegen wenige zu), so fällt er damit ein Urteil, das schwer zu begründen und praktisch unmöglich als „richtig“ zu beweisen ist.

Der Verfasser kontaktierte verschiedene Umfrageinstitute in Österreich und Deutschland um zu erfahren, wie diese aus Rohdaten eine Prognose erstellen. Die Antworten waren stets kryptisch und vage und lassen sich wohl am besten als „durch Erfahrungswerte und genaue Datenanalyse“ zusammenfassen.⁴⁶ Diese „Definition“ zeigt deutlich, wie unscharf eine derartige Datenauslegung erfolgt.⁴⁷ Die meisten Institute stimmen darin überein, daß die Rohdaten von Umfragen bei gleichbleibender Fragestellung im Zeitablauf zu volatil sind und ein „glättender“ Eingriff notwendig ist.⁴⁸ Zu kritisieren ist, daß jeder Eingriff das Urteil eines Experten voraussetzt, wobei dieses Urteil falsch sein kann.

⁴⁵ auch bei der SPÖ in Österreich bzw. der SPD in Deutschland konnte über die Jahre ein „*over-reporting*“ festgestellt werden, ohne daß dafür ein schlüssiger Grund gefunden worden wäre

⁴⁶ meist werden neben der Wahlabsicht auch andere Indikatoren – Zufriedenheit mit einzelnen Parteien, bisherige Wahlentscheidungen, Kanzlerdirektwahl, Sympathien und Antipathien etc. erfaßt, was eine Zuordnung erleichtert. Als Richtwerte dienen Erfahrungswerte aus bisherigem Wahlverhalten und anderen Zeitreihen

⁴⁷ das ist den Instituten und ihren Mitarbeitern nicht vorzuwerfen – worauf sollten sie sich anders stützen, als auf eine bestmögliche Einschätzung aller relevant erscheinenden Faktoren

⁴⁸ das Gallup Institut in den USA bezeichnet die teilweise enormen Schwankungen hingegen als „*real*“. Dort ist man der Meinung, daß die Stimmungen in der Bevölkerung tatsächlich so stark

Betrachten wir als Beispiel die „Sonntagsfrage“ in Deutschland anhand der monatlich vom „Politbarometer“ für das ZDF ermittelten Daten. Die folgende Abbildung zeigt die Umfrageentwicklung der beiden großen Parteien seit der letzten Bundestagswahl. Die dünneren Linien entsprechen dabei den Rohdaten der Erhebungen (bereinigt um *Nonresponses*, die meist um 25 Prozent lagen), die dickeren Linien geben die Prognose des Instituts unter Einbeziehung „längerfristig wirksamer Faktoren“ wieder. Man erkennt deutlich, daß die Daten der kurzfristigen Erhebungen wesentlich mehr schwanken als die „geglätteten“ längerfristigen Prognosen, womit die Intention der mitarbeitenden Forscher erfüllt wurde.⁴⁹

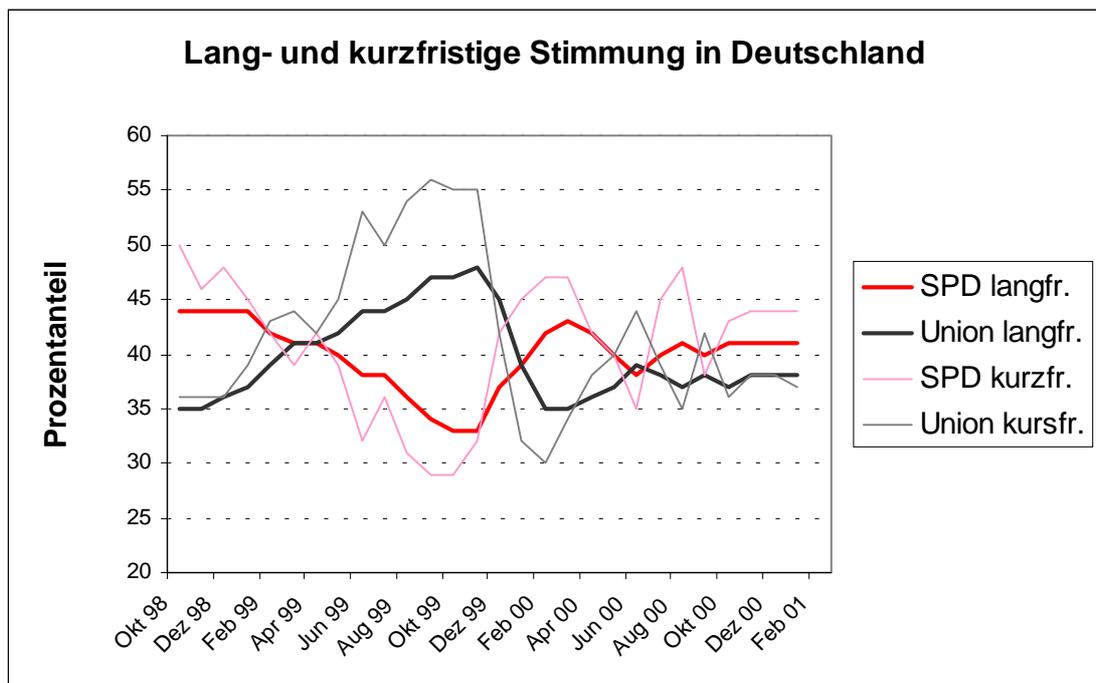


Abbildung 25: Lang- und kurzfristige politische Stimmung seit der letzten Bundestagswahl in Deutschland; Quelle: ZDF Politbarometer

schwanken, wie es erhoben wird. Immerhin stehen die Forscher damit konsequent hinter ihren Arbeits- und Erhebungsmethoden, was in Europa offensichtlich nicht gegeben ist

⁴⁹ die Standardabweichungen betragen langfristig für SPD und Union 3,1 bzw. 4,0 Prozent, kurzfristig hingegen 6,1 bzw. 7,4 Prozent

Unabhängig davon, welche Gewichtung- und Glättungsmechanismen eingesetzt werden und wie deren Einsatz begründet wird, die Kritik bleibt, daß jede Bearbeitung der Rohdaten das Grundprinzip der Umfrageforschung – die Repräsentativität durch das Gesetz der großen Zahl bei Zufallsauswahlen – untergräbt.

Egal, wie sehr sich die Institute auch um Objektivität bemühen, im Endeffekt gibt es immer einen oder mehrere Experten, die die Rohdaten entsprechend ihrer Meinung „formen“ und interpretieren. Im schlimmsten Fall sind die Experten nicht unabhängig, sondern stehen einer Partei nahe,⁵⁰ was zu bewußt oder unbewußt verzerrten Ergebnissen führen kann.

Doch das Problem ist ein grundlegendes und existiert unabhängig von der Absicht der Wissenschaftler. Selbst der unabhängigste und „objektivste“ Umfrageforscher, der vorstellbar ist, muß, wenn er Rohdaten bearbeitet, Urteile fällen. Unzählige Faktoren fließen in das endgültige Wahlergebnis ein. Der Verfasser ist überzeugt, daß es nicht möglich ist, diese auch nur annähernd zu erfassen und einzuschätzen – Trends ändern sich, das Wetter am Wahltag ist unkalkulierbar, und die menschliche Psyche ist zu vielschichtig, als daß sich die Wahlentscheidung tausender schwankender Wähler vorhersagen ließe.

Jeder Experte, der glaubt, all diese Faktoren zu kennen und einschätzen zu können unterliegt einer Hybris. John Stuart Mill meinte bereits vor 150 Jahren sinngemäß er hätte lieber die Durchschnittsmeinung von 100 Dummen als die Meinung eines Klugen, weil das Urteil der Masse selbst dann besser sei, wenn jeder einzelne weniger wüßte als der Experte (vgl. Schredelseker 2000, S. 270f).⁵¹

⁵⁰ häufig werden Umfragen von Parteien in Auftrag gegeben und bezahlt, ein Umstand, der zu „beschönigten“ bzw. gewünschten Ergebnisse beitragen kann

⁵¹ ähnlich argumentierte auch von Hayek (vgl. dazu den dritten Abschnitt diese Arbeit)

Dies ist eine starke Aussage, die man auf sich wirken lassen muß. Stimmt man ihr zu, so sollte man die weitreichenden Konsequenzen bedenken: eine Prognose, die auf bearbeiteten Rohdaten einer Umfrage beruht, ist immer eine Expertenmeinung und damit die Meinung eines Einzelnen. Vielleicht ist damit der amerikanische Weg von kaum bearbeiteten Umfragen der „bessere“.

Im Gegensatz zu Umfragen findet die Prognose auf einer Wahlbörse durch das Zusammenspiel vieler Händler statt. Die Informationen der einzelnen Teilnehmer werden am Markt aggregiert, was nach Mill und von Hayek zu einem effizienten Ergebnis führen sollte. Folgt man dieser Argumentation, so ist dem Prinzip der Wahlbörse der Vorzug vor bearbeiteten Umfragen zu geben. Die Unsicherheit des Wahlausgangs ist jedoch systeminhärent und weder durch eine Verbesserung von Umfragen noch Wahlbörsen zu beheben.

Die bisherigen Ausführungen bezogen sich in erster Linie auf Umfragen in Europa. Hier erheben die Institute den Anspruch, ein zukünftiges Wahlergebnis zu prognostizieren. In den USA tun sie das – auch in offiziellen Mitteilungen der Umfrageinstitute – schon lange nicht mehr (vgl. Gallup, Internet). Der Grund hierfür ist jedoch nicht die Bescheidenheit der Institute, sondern die Gesetzeslage, die infolge des bereits erwähnten Prognosedebakels im Jahre 1948 wesentlich restriktiver ist als in den meisten anderen Ländern. In den USA sind veröffentlichte Umfragedaten daher im Zeitablauf oft wesentlich volatiler als in Europa.⁵² Insbesondere kurzfristig wirksame Phänomene wie der berühmte *convention boost* beeinflussen die Momentaufnahmen der politischen Stimmung.⁵³

Weiters wirken sich unterschiedliche Erhebungsmethoden der einzelnen Institute mehr aus, so daß deren Werte weiter voneinander abweichen. Die folgende Grafik zeigt die Umfragewerte zweier Institute sowie die Kurse an der IEM-Wahlbörse

⁵² allerdings verwendet bspw. das Gallup-Institut in den USA drei- und sechs-Tages Durchschnittswerte, um die Volatilität ihrer Umfragewerte zu verringern

⁵³ dabei steigen die Umfragewerte eines Kandidaten nach dem Parteikonvent (*convention*) seiner Partei regelmäßig um bis zu 20 Prozent, pendeln jedoch innerhalb weniger Wochen wieder auf ihren vorherigen Wert zurück

anlässlich der US-Präsidentenwahl 1988. Um die Daten vergleichbar zu machen wird dabei jeweils der Vorsprung von Bush gegenüber Dukakis in Prozentpunkten abgebildet. Man sieht deutlich, wie sehr die Umfragewerte innerhalb weniger Wochen schwanken. Gleichzeitig ist zu erkennen, daß der Kursverlauf an einer Wahlbörse wesentlich stabiler ist.

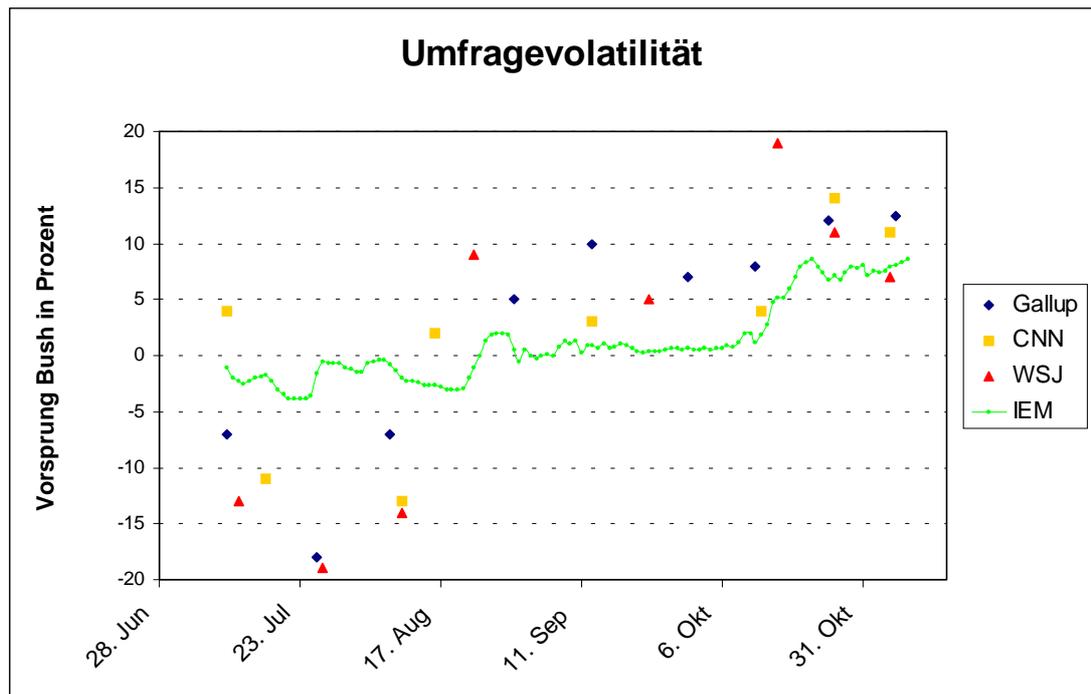


Abbildung 26: Volatilität von Umfragen in den USA; Quelle: Forsythe et al. 1992, S. 1150

Mit dem Verzicht auf dezitierte Wahlprognosen dürften Umfrageinstitute in den USA genau das als Ergebnis deklarieren, das sie wirklich erheben: ein aktuelles Stimmungsbild. Die veröffentlichten Daten spiegeln die geäußerte Wahlabsicht der Befragten, nicht jedoch eine Prognose des Wahlergebnisses wider. Die obige Grafik zeigt auch, daß die Umfrageergebnisse verschiedener Institute so unterschiedlich sind, daß nicht alle als glaubwürdig eingestuft werden können. Andererseits schwanken auch die Daten eines Instituts innerhalb weniger Tage so stark, daß die Zuverlässigkeit der Ergebnisse, insbesondere wenn man sie als Prognose auffaßt, bezweifelt werden muß.

Im Gegensatz zu den amerikanischen Organisationen behaupten Institute in Europa sehr vehement, das zu liefern, was sie doch nie zuverlässig bereitstellen können: eine wirkliche Prognose des Wahlergebnisses, die auf Daten und nicht auf einer Expertenmeinung beruht. Sehen wir uns im nächsten Abschnitt an, wie zuverlässig diese Prognosen sind bzw. sein können.

7.9 Zuverlässigkeit von Umfragen als Prognoseinstrumente

Wie wir weiter oben gesehen haben, sollte bei einem Sampleumfang von ca. 1000 die Abweichungen bei kleinen Parteien mit 95 % Wahrscheinlichkeit höchstens 1,5 % betragen, bei größeren dürften sie etwa doppelt so groß sein.

Diese Analyse kann man nun einen Schritt weiterführen und vergleichen wie weit die Umfrageinstitute in ihren Prognosen von den tatsächlichen Wahlergebnissen abweichen und wie wahrscheinlich derartige Abweichungen sind. Natürlich gibt es gute und weniger gute Umfragen,⁵⁴ doch im langjährigen Schnitt bewegt sich der MAE, also die durchschnittliche absolute Abweichung vom Mittelwert, in Europa zwischen 1,6 und 2,0 Prozentpunkte (vgl. dazu die Tabellen in 5.7.3 Prognosegenauigkeit und mögliche Ursachen für schlechte Vorhersagen sowie 8.4 Umfragen und Wahlbörsen im empirischen Vergleich).⁵⁵

Legt man eine in Europa typische Umfrage mit ca. 1000 Befragten und eine Wahl mit fünf Parteien zugrunde, so ergibt sich nach der Formel aus 7.5 Statistische Grundlagen für den Einsatz von Stichproben, ein Standardfehler von ca. 1,25 Prozent. Im Schnitt beträgt der beobachtete MAE somit ungefähr eineinhalb Standardfehler, was bei einer Normalverteilung nur mit einer Wahrscheinlichkeit von ca. 13 Prozent eintreten sollte. Es ist also extrem unrealistisch, daß die

⁵⁴ als überdurchschnittlich gut hat sich etwa in den letzten vier Jahrzehnte das deutsche Allensbach-Institut erwiesen, das bei den 11 Bundestagswahlen von 1957 bis 1994 einen durchschnittlichen MAE von 0,81 Prozent aufwies (vgl. Noelle-Neumann 1996, S. 27)

⁵⁵ in den USA ist die Abweichung größer – laut Gallup durchschnittlich drei Prozent – was sowohl in der Nicht-Bearbeitung der Umfragen als auch in der geringeren Zahl der Parteien begründet sein dürfte, die Wahrscheinlichkeit, daß es zu so großen Abweichungen zufällig kommt, liegt bei nur 6 Prozent

beobachteten Abweichungen zufällig sind. Bei Prognosen, die aus Umfragen abgeleitet werden, dürften also systeminhärente Gründe zu Abweichungen vom Wahlergebnis beitragen.

Die bisherigen Ausführungen haben gezeigt, daß anhand von Befragungen die Wahlabsicht einer Grundgesamtheit innerhalb statistischer Bandbreiten erfaßt werden kann. Jedoch ist eine Erhebung immer auf den Zeitpunkt bezogen, zu dem sie stattgefunden hat. Eine Umfrage die zwei Wochen vor der Wahl durchgeführt wird, sagt nur etwas über die Stimmung zu diesem Zeitpunkt, nicht jedoch über das spätere Wahlergebnis, aus.

Die einzig gültige Vorhersage des Wahlergebnisses liefern daher *exit polls*, die am Wahltag durchgeführt werden. Diese Befragungen sind im übrigen der stärkste Beweis für das Funktionieren der Methodik von Umfragen, denn sie spiegeln das Wahlergebnis meist sehr gut wider, obwohl – oder gerade weil – sie von den Instituten nicht bearbeitet werden.⁵⁶ *Exit polls* – egal ob als Zufallsauswahl oder nach Quoten ausgewählt – belegen somit, daß sowohl die statistischen Grundlagen als auch die Erhebungsmethoden von Umfragen ihre Berechtigung haben. Der Fehler liegt offensichtlich in dem Schritt, aus einer Erhebung eine Prognose abzuleiten, also die Daten zukunftsbezogen zu interpretieren.

Eine Prognose ist genau genommen nichts anderes als ein Expertentip, der auf Umfragedaten aufbauend, von den Institutsmitarbeitern gemacht wird. Damit ist die Vorhersage begründet in der „persönlichen Einschätzung“ von Experten.⁵⁷ Nur wenn eine Umfrage am Wahltag erfolgt (bspw. *exit poll*), oder wenn es aufgrund anderer (nur schwer vorstellbarer) Umstände zwischen der Erhebung und der Wahl zu keiner Meinungsänderung der Wähler mehr kommen kann, darf mit einer zuverlässigen Prognose gerechnet werden.

⁵⁶ eine von Verfasser analysierte *exit poll* zur Nationalratswahl 1999 wies einen MAE von 0,54 Prozent auf

⁵⁷ selbst wenn diese zur Erstellung der Prognose auf Regeln und Mechanismen zurückgreifen – dann liegt die Subjektivität nur eine Ebene höher bei der Erstellung dieser Regeln

Das grundsätzliche – und wohl unbehebbares – Problem der Umfrageforschung besteht darin, daß die der Prognose zugrundeliegende Zufallsstichprobe aus einer anderen Grundgesamtheit gezogen wird, als der Gesamtheit derer, die schließlich wählen: manche können nicht zur Wahl gehen, obwohl sie es vorhatten (Unfall, Tod, etc.), andere schwanken und bleiben dann aufgrund von Wetter, Fernsehprogramm oder anderen unkalkulierbaren Einflüssen doch der Wahlurne fern.

Selbst wenn derartige Faktoren außer Acht gelassen werden, so gibt es viele Wähler, die sich erst kurz vor der Wahl für eine Partei entscheiden oder ihre Meinung in den letzten Tagen noch einmal ändern, obwohl sie sich schon festgelegt hatten.⁵⁸ Aus diesen Gründen MÜSSEN Wahlprognosen die aus Umfragedaten abgeleitet sind, als exakte Vorhersagen scheitern.

Dies beraubt die Umfrageforschung jedoch nicht ihrer Berechtigung als Meinungsbarometer und Orientierungshilfen für Parteien und Politikinteressierte. Jedoch sollten die Umfrageinstitute so ehrlich sein zu sagen, daß sie das Wahlergebnis nicht exakt vorhersagen können und in Zukunft Wahlprognosen bewußt als bedingt Vorhersage mit dem dazugehörigen Konfidenzniveau und den entsprechenden Bandbreiten angeben.

Wenden wir uns nun dem Vergleich von Wahlbörsen und Umfragen zu. Dabei werden wir im Verlauf der Analyse noch einmal zum Thema der Prognosegenauigkeit zurückkehren.

⁵⁸ nach allen dem Verfasser bekannten Untersuchungen nimmt der Anteil dieser *late deciders* kontinuierlich zu, ebenso steigt der Anteil der Wähler, die sich nicht fix an eine Partei gebunden fühlen

8 Wahlbörsen und Umfragen im Vergleich

Nach der eingehenden Analyse des Aufbaus und der Funktionsweise von Wahlbörsen im ersten Abschnitt sowie von Umfragen im vorhergehenden Kapitel werden die beiden Instrumente in diesem Kapitel direkt miteinander verglichen. Hauptziel ist es, Stärken und Schwächen dieser Methoden der Wahlprognose zu erkennen und Ansätze herauszuarbeiten, wie die Defizite des einen Instruments durch die Vorteile des anderen behoben werden können, um die Wahlforschung insgesamt präziser und zuverlässiger zu machen.

Welchen Beitrag können Wahlbörsen in der Wahlforschung leisten? Vergessen wir nicht, daß die Fragestellung bei Wahlbörsen eine andere ist, als bei Umfragen. Während bei Erhebungen meist die eigene Meinung der Teilnehmer zu einem Thema gefragt ist („Was würden Sie wählen...?“), befindet man sich mit politischen Aktienmärkten auf einer Metaebene – man stellt implizit eine zukunftsbezogene Frage zweiter Ordnung („Was glauben Sie, daß alle anderen wählen werden?“).

Der zweite wesentliche Unterschied ist jener der Vorgehensweise: bei Umfragen strebt man Repräsentativität an und vertraut dabei auf das Gesetz der großen Zahl bei Zufallsstichproben. Die größte Schwachstelle dabei ist, daß eine Prognose prinzipiell unmöglich ist, weil jede Erhebung nur für den Erhebungszeitpunkt ein gültiges Stimmungsbild vermittelt. Die Erstellung einer Vorhersage erfordert somit den Eingriff von „Experten“, womit aber das Prinzip der Repräsentativität durch Zufallsauswahl verletzt wird.

Die Funktion der Experten bei Umfragen, nämlich die Einbeziehung langfristiger Erwartungen, übernimmt an einer Wahlbörse jeder einzelne Händler. Bei einem PSM vertraut man auf den Markt als Aggregationsinstrument von Information, wobei auch eine Börse mit einem nicht repräsentativen Händlerfeld eine gute Prognose erbringen sollte. Zum größten Problem von Wahlbörsen wurde ausgerechnet ihr anfänglicher Prognoseerfolg, denn dieser war die Ursache eines erheblichen Medienechos, was die Märkte wiederum für Manipulationen durch Sympathisanten einzelner Gruppierungen interessant machte.

Der große Vorteil eines PSM ist, daß keine „Experten“ in den Preisbildungsprozeß eingreifen, womit er ein demokratischeres Instrument als eine Umfrage darstellt. Mehr noch, die Struktur des Handels erlaubt es besonders interessierten Händlern, größeren Einfluß auf die Preisbildung und damit die Prognose zu gewinnen. Somit bietet die Börse eine Möglichkeit, die Intensität einer Überzeugung auszudrücken, was bei Umfragen zur Wahlforschung kaum möglich ist.⁵⁹

Ein weiterer Unterschied zwischen Umfragen und Wahlbörsen liegt in der Art der Fragestellung. Bei Umfragen ist die Frageformulierung oft einer der schwierigsten Teile einer Erhebung. Auch wenn dies bei PSMs aufgrund der meiste sehr eindeutigen Frage („Wie wird die Wahl ausgehen?“) eine geringere Rolle spielt, sollte die Bedeutung des Designs eines Marktes nicht unterschätzt werden.⁶⁰

Ein großer Vorteil von Umfragen liegt in ihrer Flexibilität – grundsätzlich ist fast jede Ausprägung einer Meinung erhebbar. Die Aussage bei der Grundform einer Wahlbörse ist hingegen sehr begrenzt, denn „prognostiziert“ wird „nur“ das Wahlergebnis. Jedoch gibt es auch bei PSMs mittlerweile eine Vielzahl verschiedener Designs (vgl. IEM, Internet).

8.1 Die Sonntagsfrage

Es geht nicht darum, das Instrument „Wahlbörse“ mit der Meinungsforschung generell zu vergleichen, sondern um eine Gegenüberstellung mit jenen Erhebungen, die mit dem Ziel einer Wahlprognose gemacht werden. Solche Befragungen werden allgemein unter dem Schlagwort „Sonntagsfrage“ zusammengefaßt. Diese Frage lautet mit leichten Abwandlungen stets „Welche Partei würden Sie am ehesten wählen, wenn am nächsten Sonntag ...wahl wäre.“

⁵⁹ während es bei vielen Erhebungen Skalen zum Beschreiben der Intensität einer Ausprägung gibt („Wie hoch würden Sie auf einer Skala von 1 bis 7 ...?“), erlaubt die Frage nach der Wahlentscheidung stets nur eine Intensität der Antwort (dichotome Variable)

⁶⁰ wir haben bereits gesehen, daß die Kursentwicklung an VS- und WTA-Märkten sehr unterschiedlich ist, auch wenn sie auf dasselbe Ereignis bezogen sind. Auch andere Designelemente spielen oft eine beträchtliche Rolle

Die manchmal vorgebrachte Kritik, daß damit das Wahlverhalten am nächsten Sonntag, nicht aber am Wahltag abgefragt würde, geht ins Leere, denn kein „normaler“ Mensch würde nächsten Sonntag der einen Partei den Vorzug geben, wenn er schon wüßte, daß er am Wahltag eine andere wählen würde.⁶¹

Die Fragestellung ist damit jener an den PSMs ähnlich, jedoch nicht ident, denn an Wahlbörsen wird die Meinung über den Wahlausgang, bei Umfragen die eigene Wahlabsicht gefragt. Dadurch haben politische Märkte stets einen starken Bezug auf den Wahltag, während mit Umfragen die momentane Stimmung erhoben wird.

Trotz der Einfachheit der Fragestellung bei der Sonntagsfrage sind auch dabei kritische Entscheidungen zu treffen. Nennt man etwa in Österreich nur die Partei oder auch den Spitzenkandidaten? Nennt man in den USA nur die Präsidentschaftskandidaten oder auch deren Stellvertreter, erwähnt man die Partei, für die sie kandidieren?

Gallup hat dabei die einleuchtende Regel entwickelt, die Situation bei der Wahl möglichst gut zu kopieren – da am Stimmzettel die Namen der Präsidentschafts- und Vizepräsidentschaftskandidaten sowie deren Partei aufscheinen, erwähnt auch Gallup in seinen Umfragen diese Merkmale.

Beginnen wir nun die vergleichende Analyse der beiden Instrumente bei ihrem wichtigsten Element – den Teilnehmern.

⁶¹ die Frage wird auf den nächsten Sonntag bezogen, weil bei der Frage nach dem Verhalten am Wahltag die *Nonresponsequote* enorm ansteigt, weil viele Befragte dann "Weiß ich noch nicht." antworten

8.2 Motivation der Teilnehmer

Jede Handlung wird von einem Individuum nach ihren Kosten und Nutzen bewertet. Getätigt werden nur Handlungen, bei denen der erwartete Nutzen höher ist, als die anfallenden Kosten.

Dieser Ansatz sollte jedoch nicht zu formal gesehen werden: natürlich wird nicht bei jeder kleinen Tätigkeit im Geiste genau berechnet, welcher Aufwand notwendig ist, um zum gewünschten Ergebnis zu kommen, oft greift der Mensch auf Prozeduren und Gewohnheiten zurück.⁶² Dies liegt daran, daß auch ein Zeitaufwand als „Kosten“ zu betrachten ist, so daß es u.U. nicht sinnvoll ist, stundenlang darüber nachzudenken, wie man minimale Nutzenverluste beheben könnte.

Betrachten wir vor diesem Hintergrund nun Wahlbörsen und Umfragen, um festzustellen, warum Menschen daran teilnehmen.

8.2.1 Wahlbörsen

Die Motivationsgründe für Teilnehmer an Wahlbörsen wurden im ersten Abschnitt bereits eingehend behandelt. Kürzestmöglich kann man sagen, daß die Händler aus eigenem Antrieb an einem politischen Aktienmarkt teilnehmen, was sich schon in der aktiven Anmeldung zur Börse niederschlägt. *Traders* nehmen die erheblichen Zeitkosten auf sich,⁶³ um sich selbst zu beweisen, daß sie am Markt erfolgreich sind. Weitere wichtige Gründe sind Spaß am Spiel und das Entdecken eines neuen Instruments. All dies sind Motive, die nicht auf andere bezogen sind, sondern dem „inneren“ Antrieb der Händler entspringen.

⁶² nur die wenigsten Menschen stellen lange Überlegungen an, bevor sie von der Couch aufstehen, um sich ein Glas Wasser zu holen – sie tun es jedoch genau dann, wenn der Nutzen (Durstlöschen) größer ist als die Kosten (Bequemlichkeitsverlust)

⁶³ die monetären Kosten sind meist vernachlässigbar und beschränken sich auf Gebühren für das Surfen im Internet sowie die Anmeldegebühr, die man jedoch – bereinigt um den Gewinn oder Verlust (Erwartungswert null) – am Wahltag wiederbekommt

8.2.2 Umfragen

Gegenteilig ist die Situation bei Umfragen, welche man als von „außen“ induziert bezeichnen kann. Die Teilnahme erfolgt nicht durch eine aktive Handlung des Befragten, sondern die Kontaktaufnahme geht vom Umfrageinstitut aus.⁶⁴ Auch bei Umfragen hat der Teilnehmer erhebliche Kosten in Form von Zeitaufwand, sowie manchmal in Form von Verärgerung oder Beschämung über sinnlose oder provokante Fragen, die er lieber nicht gefragt worden wäre.⁶⁵

Der Nutzen der Teilnahme an einer Erhebung kann in einem besonderen Interesse am Befragungsthema, dem Glauben an die Nützlichkeit wissenschaftlicher Forschung oder auch einfach in der Abwechslung durch eine neue Erfahrung liegen (vgl. Schnell 1995, S. 293). Menschen wollen prinzipiell soziale Mißbilligung, wie sie eine Antwortverweigerung auslösen könnte, vermeiden, und nehmen darum vielfach höflichkeitshalber an einer Befragung teil. Anderen scheint es wiederum Spaß zu machen, befragt und „wichtig genommen“ zu werden (vgl. Diekmann 1995, S. 377).

8.2.3 Verlässlichkeit von Ergebnissen

Die unterschiedlichen Motivationsgründe bei PSMs und Umfragen haben auch weitreichende Konsequenzen für die Verlässlichkeit der jeweiligen Ergebnisse. Zwar haben die meisten Menschen in westlichen Kulturen eine Norm verinnerlicht, Fremden gegenüber eine „wahre“ Antwort zu geben, jedoch gilt dies nur solange, als die Kosten der Aufrichtigkeit einen (individuellen) Schwellenwert nicht überschreiten.⁶⁶ Insbesondere die Frage nach dem

⁶⁴ eine Ausnahme – und eventuelle Entwicklungsmöglichkeit – bilden Umfragen über das Internet, bei denen der Befragte die aktive Rolle übernimmt. Der entscheidende Nachteil ist, daß solche Umfragen bisher nicht repräsentativ sind

⁶⁵ sind diese Kosten für eine Person zu hoch, so verweigert der Befragte entweder einzelne Antworten oder überhaupt die Teilnahme an der Erhebung

⁶⁶ dieser Schwellenwert variiert durch persönliche Erfahrungen und unterschiedliche Einschätzungen dessen, was „statthaft“ ist. Noch stärker sind diese Schwankungen, wenn man unterschiedliche Kulturen betrachtet

Wahlverhalten hat sich dabei in der Praxis als kritisch herausgestellt. Daher ist bei Umfrageergebnissen stets zu überprüfen, ob die erhobenen Daten zuverlässig sind. Ein weiterer kritischer Punkt sind Antwortverweigerungen, weil diese eine Stichprobe verzerren können.

Bei Wahlbörsen stellen Verweigerungen hingegen kein Problem dar, denn die Teilnahme an einem Markt ist ohnehin freiwillig und erfordert ein aktives Bemühen des Händlers. Verliert dieser später das Interesse, so nimmt er eben nicht mehr am Markt teil, was aber kaum zu Verzerrungen führen dürfte. Ein großer Vorteil ist, daß für eine gute Prognose keine Repräsentativität der Teilnehmer erforderlich ist, weil die *traders* nicht ihre Wahlabsicht, sondern ihre Meinung über die Wahlabsicht anderer in ihrem Verhalten am Markt ausdrücken.⁶⁷ Schließlich gibt es keinen Grund, warum ein Händler nicht seine wahre Meinung über den Wahlausgang zeigen sollte, denn durch schlechte Handelsentscheidungen riskiert er, sein eingesetztes Geld zu verlieren. Die Wahlabsicht der Teilnehmer bleibt an einer Wahlbörse unerheblich und unbekannt, so daß auf diesbezügliche Hemmungen keine Rücksicht genommen werden muß.

Beide von uns betrachteten Instrumente weisen verschiedene Stärken und Schwächen auf. Betrachten wir als nächsten Aspekt, wie robust sie jeweils gegen Manipulationen sind und wie Veranstalter solche Verzerrungen vermeiden können

8.3 Manipulierbarkeit und Beeinflussung

Wir haben bereits weiter oben festgehalten, daß Umfragen u.U. selbst den Wahlausgang beeinflussen, wenn Wähler ihre Wahlentscheidung auch mit Rücksicht auf Umfragewerte treffen. Der Einfluß der Meinungsforschung ist zwar schwer meßbar, doch sind sich Medien und Parteien einig, daß er erheblich ist, womit Meinungsforschungsergebnisse für Manipulationen interessant werden.

⁶⁷ zwar ist prinzipiell keine Repräsentativität erforderlich, jedoch haben Studien gezeigt, daß die parteipolitische Überzeugung die Wahrnehmung eines Menschen beeinflussen kann (vgl. Forsythe et al. 1992). Dies kann zu Fehleinschätzungen am Markt führen, die bestehen bleiben können, wenn sehr viele Teilnehmer der Verzerrung unterliegen

Daher soll hier kurz beleuchtet werden, wie robust Umfragen bzw. Wahlbörsen gegen Beeinflussungen sind. Dabei ist der Verursacherkreis von Manipulationen bei den beiden Instrumenten jeweils unterschiedlich: bei Umfragen haben die Teilnehmer kaum einen Einfluß auf die Prognose, denn die Daten werden von einem Institut erhoben und ausgewertet. Auch die Prognose wird vom Veranstalter erstellt, so daß nur die Mitarbeiter der Institute das Ergebnis nachhaltig beeinflussen können.

Bei Wahlbörsen ist die Situation eine andere: der Veranstalter stellt in der Regel nur die Handelsplattform zur Verfügung, greift aber nicht aktiv ein. Die Prognose kommt alleine durch die Transaktionen der Händler zustande.

Die größte Gefahr von Manipulationen liegt stets beim aktiven, einflußreichen Teil – bei Umfragen also bei den Instituten, bei Wahlbörsen hingegen bei den Teilnehmern.

8.3.1 Wahlbörsen

Die Möglichkeit von Manipulationen bei PSMs war bereits ein zentrales Kapitel im ersten Abschnitt, so daß hier nur in komprimierter Form auf einige wesentliche Punkte eingegangen wird. Wie bereits festgestellt, gehen Manipulationen bei Wahlbörsen praktisch immer von Händlern aus. Bewußt herbeigeführte Verzerrungen durch Veranstalter sind bisher nicht bekannt geworden und wohl auch nicht vorgekommen. Selbst wenn gewinnorientierte Unternehmen wie etwa Zeitungen als Betreiber von Wahlbörsen auftreten, so bleibt es ihr Hauptinteresse, daß die Börse funktioniert, den Teilnehmern Spaß macht und daß die Schlußkurse eine gute Prognose des Wahlergebnisses liefern. Jeder Eingriff könnte diesen Zielen zuwiderlaufen, weswegen die Veranstalter i.d.R. so wenig wie möglich am Markt intervenieren.

Den Händlern ist leider kein so gutes Zeugnis auszustellen, sowohl Teilnehmergruppen als auch Parteien haben wiederholt versucht, Märkte zu manipulieren.

Während einzelne Händler kaum gewinnbringende Manipulationen durchführen können, ist diese Möglichkeit bereits bei kleinen Händlergruppen gegeben. Ziel ist

es dabei meist, den Gewinn eines Händlers maximieren um einen ausgesetzten Preis für den erfolgreichsten Teilnehmer zu erhalten. Dabei versuchen die manipulierenden Händler üblicherweise die Kontrolle über den Kurs einer Aktie zu erlangen, um die Kontrollmechanismen des *Double Auction Markets* außer Kraft zu setzen, und das gemeinsame Spielkapital auf einem Teilnehmer zu vereinen. Da diese Manipulationen in den Kursdaten leicht erkennbar sind, können derartige Händler meist relativ rasch identifiziert und aus dem Markt ausgeschlossen werden.

Durch ein gutes Marktdesign läßt sich die Gefahr derartiger Manipulationen auf ein sehr geringes Maß reduzieren. Neben einem einheitlichen Einsatz aller Teilnehmer spielt dabei eine angemessene Teilnehmerzahl eine zentrale Rolle. Beide Maßnahmen begrenzen die Marktmacht des einzelnen und erschweren damit die bewußte Verursachung von Verzerrungen.

Wesentlich schwieriger zu bekämpfen sind nicht-gewinnorientierte Manipulationen durch Parteien oder deren Sympathisanten. Werden die Kurse einer Wahlbörse häufig in Massenmedien erwähnt, so gewinnen sie ebenso einen Einfluß auf die Wahlentscheidung einzelner, wie das schon bei Umfragen festgestellt wurde. Mehr noch, manche Leser von Zeitungen oder Seher von Fernsehsendungen könnten die Kurse für Umfragewerte halten. Damit werden die Kurse ebenso attraktiv für Beeinflussungen durch Parteien, wie Umfragen.⁶⁸

Beteiligt sich beispielsweise eine Gruppe von Parteigängern an einer Wahlbörse, so verfügt sie u.U. über erhebliche Marktmacht. Setzen die Händler diese Macht nicht entsprechend ihrer Meinung über den Wahlausgang, sondern zur Unterstützung „ihrer“ Partei ein, so kann es zu erheblichen Verzerrungen kommen.

Problematisch sind derartige Manipulationen v.a. deshalb, weil sie länger wirksam, schwerer zu erkennen und kaum zu bekämpfen sind. Die beste Abhilfe bieten hier, wie bei allem Manipulationen, möglichst hohe Händlerzahlen sowie ein einheitlicher Einsatz zur Reduktion der Marktmacht des einzelnen.

⁶⁸ während es meist sehr schwierig, oder zumindest teuer ist, Umfragen zu manipulieren, ist dies bei Wahlbörsen schon mit geringen Geldmengen möglich

8.3.2 Umfragen

Bei Umfragen ist der Einfluß des einzelnen Befragten meist so gering, daß davon kaum eine verzerrende Wirkung ausgehen kann. Existieren jedoch in größeren Wählergruppen Tendenzen, falsche Antworten zu geben bzw. gar nicht zu antworten, so können diese sehr wohl zu Verzerrungen führen.⁶⁹ Diese werden jedoch nicht bewußt von jemandem ausgelöst oder gesteuert und sind somit für unsere Analyse absichtlich herbeigeführter Manipulationen irrelevant, selbst wenn sie die Qualität eines Befragungsergebnisses beeinträchtigen können.

Die Macht, auf das Ergebnis einer Umfrage bewußt einzuwirken, haben nur die Mitarbeiter des befragenden Instituts. Dabei kann zwar von einzelnen Interviewern ein verzerrender Einfluß ausgehen, wirklich bedenklich sind jedoch v.a. bewußte Verzerrungen durch jene Experten, die letztendlich aus den Rohdaten eine Prognose formen.

Die üblicherweise beträchtliche Quote an *Nonresponses* gibt den Mitarbeitern von Instituten einen erheblichen Spielraum bei der Erstellung einer Wahlvorhersage. Den Experten eine bewußte Manipulationsabsicht zu unterstellen, wäre dumm, doch ist zu bedenken, daß Umfragen normalerweise im Auftrag von Medien oder Parteien durchgeführt werden. Dies kann die Meinungsforscher bei der Erstellung der Prognose bewußt wie unbewußt beeinflussen. Überlegen wir kurz, was die Auftraggeber erwarten.

Die Medien wollen „Sensationen“, um hohe Auflagen zu erzielen. Daher ist ihnen eine Prognose mit markanten Aussagen und möglichst überraschenden Ergebnissen am liebsten. Da die Institute das wissen und auf Medien als Kunden zumindest teilweise angewiesen sind, könnten die Rohdaten im Zweifelsfall etwas extremer in eine Richtung interpretiert werden.⁷⁰

⁶⁹ man denke hier nur an die lange vorhandene Tendenz von FPÖ-Wählern, sich nicht zu deklarieren, wodurch die Rohdaten dieser Partei stets zu niedrig waren

⁷⁰ dies ist den Instituten nicht vorzuwerfen, zumal sie mit gewagten Thesen auch richtig liegen können – OGM sah im österreichischen Wahlkampf 1999 bereits im Juli die FPÖ vor der ÖVP, was sich schließlich als korrekt erwies. Damit war das Institut der Konkurrenz um sechs Wochen voraus

Auch die zweite wichtige Kundengruppe, Parteien, verleiten Meinungsforscher möglicherweise zu einer tendenziösen Auslegung von Daten. Ist ein Spitzenkandidat mit schlechten Umfragedaten konfrontiert, so kontert er gerne mit den Werten eines anderen Instituts, bei dem seine Partei besser liegt.

Wahrscheinlich treten Manipulationen dennoch nur in sehr geringem Maß auf - zwar kann es bei der Verarbeitung von Rohdaten zu Prognosen zu geringen Verzerrungen kommen, doch sollte der Einfluß der „sozialen Kontrolle“ bzw. der Disziplinierung durch den Markt nicht unterschätzt werden. Ein Umfrageinstitut, das Ergebnisse zugunsten einer Partei manipuliert, oder auch nur in den Ruf kommt dies zu tun, wird sehr schnell als „Parteiinstitut“ abgestempelt.

In der Folge hätte es am freien Markt Probleme, Aufträge zu erhalten. Schlimmer noch, durch eine Verzerrung von Ergebnissen würde sich das Institut auch für jede Partei überflüssig machen, denn Befragungen, denen man nicht trauen kann, sind wertlos. Mit dem Verlust des Vertrauens entzöge sich ein derartiges Unternehmen selbst seine Existenzgrundlage, so daß sich schon aus diesem Grund jeder Mitarbeiter vor bewußte Verzerrungen hüten wird. Leichte „Beschönigungen“ von Zahlen dürften also vorkommen, systematische Manipulationen von Umfragen bilden in der westlichen Welt aber die Ausnahme.⁷¹

Zusammenfassend ist festzustellen, daß sowohl bei Umfragen als auch bei Wahlbörsen das Problem von Manipulationen gegeben ist, daß drastische Verzerrungen jedoch eher die Ausnahme bilden dürften. Wenden wir uns nun der Praxis zu und betrachten wir, wie Umfragen und politische Aktienmärkte im empirischen Vergleich abschneiden.

⁷¹ so liegen die Umfragewerte verschiedener Institute in Österreich in der Regel zwar nicht zu weit auseinander, jedoch sind Abweichungen um zwei bis drei Prozent keine Seltenheit. Einzig die Umfragewerte für Richard Lugners „Unabhängige“ vor der NRW 1999 lassen aufforchen: diese lagen bei gleichem Erhebungszeitraum zwischen einem (Spectra) und vier (Gallup) Prozent. Ebenso schwankten die Werte des Liberalen Forums bei den einzelnen Instituten zwischen 2 und 5 Prozent. In diesen Fällen könnten parteipolitische Präferenzen der Institute bzw. ihrer führenden Mitarbeiter in die Prognosen eingeflossen sein

8.4 Umfragen und Wahlbörsen im empirischen Vergleich

Nachdem in den bisherigen Ausführungen bruchstückhaft immer wieder empirische Daten und Vergleichswerte von Umfragen und Wahlbörsen vorkamen, um unterschiedlichste Aspekte zu beleuchten, bilden solche Gegenüberstellungen das zentrale Thema dieses Kapitels. Dabei soll erarbeitet werden, welche Faktoren einer Wahlauseinandersetzung von welchem Instrument besser erfaßt werden können und inwieweit die beiden Methoden voneinander unabhängig sind. Es kann vorausgesetzt werden, daß Wahlbörsen keinen signifikanten Einfluß auf Umfragen haben. Selbst an den größten Wahlbörsen haben stets weniger als ein Promille aller Wahlberechtigten eines Landes teilgenommen; selbst wenn man die Publikation der Kurse in Medien berücksichtigt, ist anzunehmen, daß die Werte maximal von einem Prozent aller Wähler bewußt wahrgenommen werden.⁷²

Während Wahlbörsen keinen Einfluß auf Umfragen haben, ist die entgegengesetzte Beziehung umstritten. Veranstalter von Wahlbörsen betonen regelmäßig, daß die Kursbildung unabhängig von Umfragen erfolgt, während Kritiker behaupten, daß die Märkte nur die Entwicklungen von Befragungen nachvollziehen. Fest steht, daß die Teilnehmer an politischen Aktienmärkten auf Umfragen zurückgreifen können und diese wohl auch in die Bildung ihrer Erwartungen über den Wahlausgang einbeziehen.

8.4.1 Wie reagieren Wahlbörsen auf Umfragen?

Wie bereits im ersten Abschnitt gezeigt wurde, spiegeln Wahlbörsen nicht einfach die Daten und Trends aus Umfragen wider. Die Händler beziehen vielmehr ihre Erwartungen in ihr Handelsverhalten ein, sie denken langfristig. Betrachten wir anhand zweier aktueller Märkte, wie sich die Kurse an Wahlbörsen im Vergleich zu den jeweiligen Umfragen entwickelten, um daraus abzuleiten, wie abhängig oder unabhängig von Befragungsdaten die Preisbildung an PSMs ist.

⁷² bei den größten Wahlbörsen in Deutschland (12.000 Händler) und Österreich (1500 Händler) entsprechen die Teilnehmerzahlen nur ca. 0,03 Prozent aller Wahlberechtigten. Bei einer Umfrage mit Samplegröße 1000 wäre damit unter den Befragten wahrscheinlich kein, maximal ein Händler der Wahlbörse (im Schnitt 0,3)

Österreich

Die Einbeziehung langfristiger Überlegungen in die Handelsentscheidungen an Wahlbörsen konnte man bereits an den ersten derartigen Märkten beobachten. Als Beispiel wird in dieser Arbeit der Markt zur letzten Nationalratswahl (NRW) in Österreich vom 3. Oktober 1999 herangezogen. Die Wahl war zum einen vom „Kampf um Platz zwei“ zwischen der FPÖ und der ÖVP, zum andern vom Zittern um das Erreichen der 4-Prozent Hürde durch das Liberale Forum gekennzeichnet. Als zusätzlicher Unsicherheitsfaktor kam die Kandidatur der „Unabhängigen“ Richard Lugners (DU) hinzu; nachdem der Bauunternehmer ein Jahr zuvor bei der Präsidentschaftswahl knapp 10 Prozent der Stimmen erhalten hatte, schien für seine Gruppierung der Parlamentseinzug in Reichweite.

Die folgende Abbildung zeigt den Kursverlauf an der Wahlbörse der Tageszeitung „Die Presse“ (Linien) sowie alle ab Ende Juni 1999 veröffentlichten Umfragewerte der Institute Gallup, OGM und Spectra (Punkte).

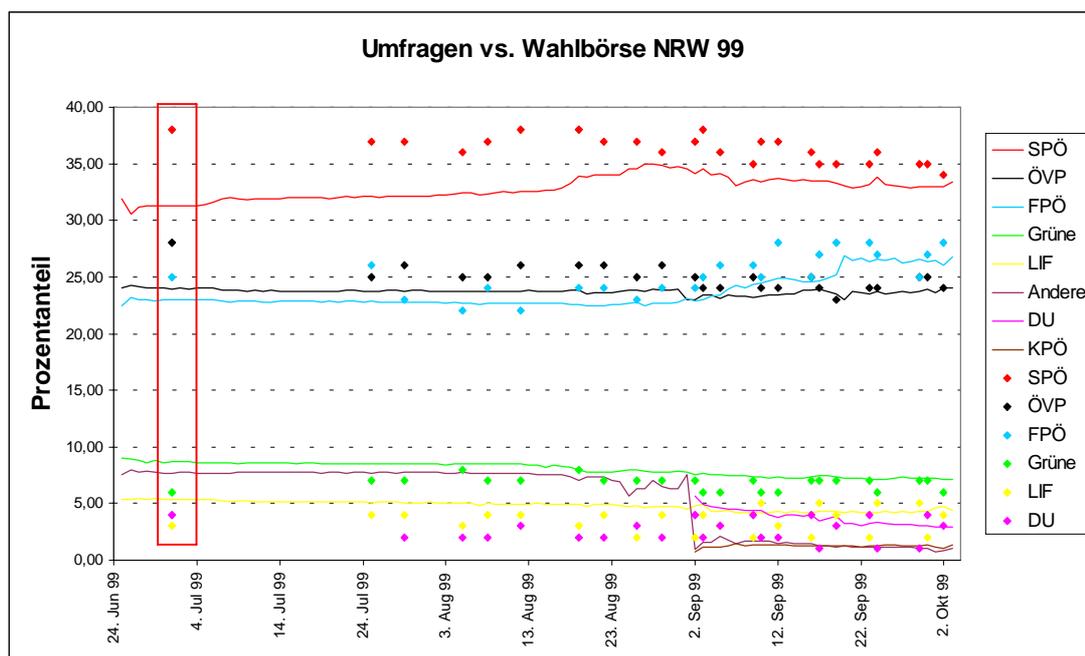


Abbildung 27: Umfragen vs. Wahlbörsenkurse anlässlich der Nationalratswahl 1999 in Österreich; Quelle: „Die Presse“, Umfrageinstitute

Schon auf den ersten Blick ist zu erkennen, daß Umfragewerte und Börsenkurse nur in Ausnahmefällen übereinstimmten. Die Kurse der drei größeren Parteien

bewegten sich fast durchgehend unter den entsprechenden Umfragewerten, während die Aktien der kleineren Parteien meist über den jeweiligen Umfragewerten gehandelt wurden. Diese Diskrepanz war zu Beginn des Wahlkampfes am größten (rote Markierung), und nahm kurz vor dem Wahltag ab, wobei die Kurse der größeren Parteien tendenziell eher anstiegen (in Summe von 78,4 in der ersten Woche auf 84,2 Prozent am Wahltag), während jene der kleineren Parteien tendenziell sanken (von 22,0 Prozent auf 16,9 Prozent).

Die folgende Abbildung zeigt detaillierter den Kursverlauf und die Umfragewerte der Grünen zwischen dem 24. Juli und dem Wahltag. Das Wahlergebnis ist als unterbrochene rote Linie dargestellt

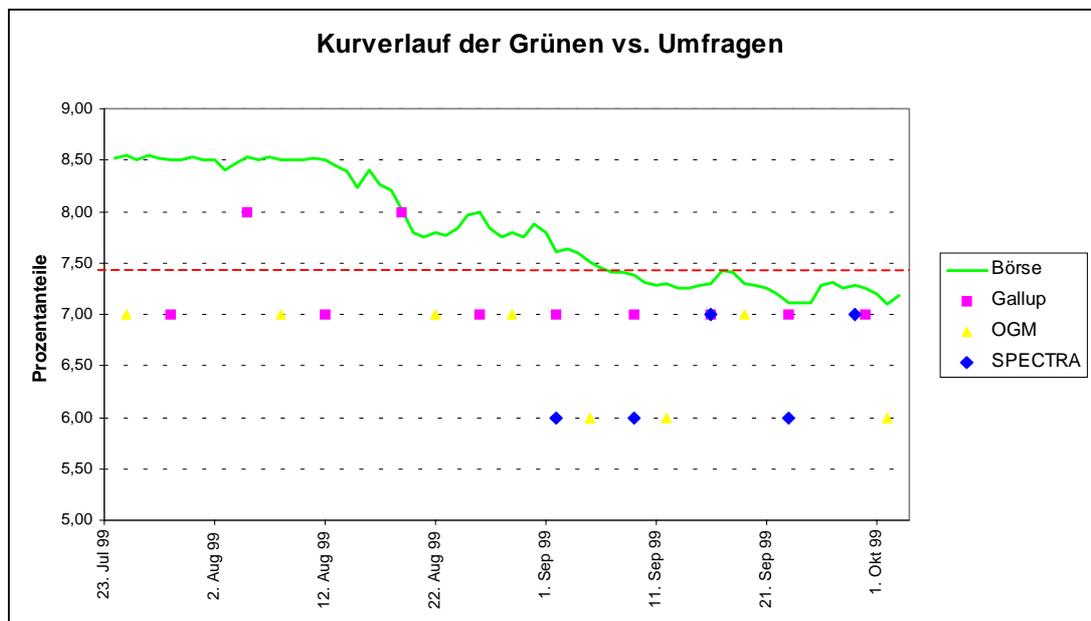


Abbildung 28: Kursverlauf und Umfragewerte der Grünen im Vergleich;
Quelle: „Die Presse“, Umfrageinstitute

Wie bereits im ersten Abschnitt ausgeführt, dürften der Optionscharakter sowie die bessere Eignung für Spekulationen Gründe für die anfänglich hohen Kurse von Aktien kleinerer Parteien sein (vgl. Abschnitt 5.6.1 Überbewertung kleiner Aktien). Gegen Ende des Wahlkampfes sinken die Kurse dann, so daß sie als seriöse Prognose betrachtet werden können.

Als indirekte Folge davon liegen die Kurse der größeren Parteien eher unter den entsprechenden Umfragewerten. Dies erkennt man in der folgenden Abbildung, die den Kursverlauf und die Umfragewerte der SPÖ darstellt. Die Aktie wurde über den gesamten Verlauf des Wahlkampfes geringer als in den Umfragen gehandelt, was später durch das Wahlergebnis (unterbrochene grüne Linie) gerechtfertigt wurde.

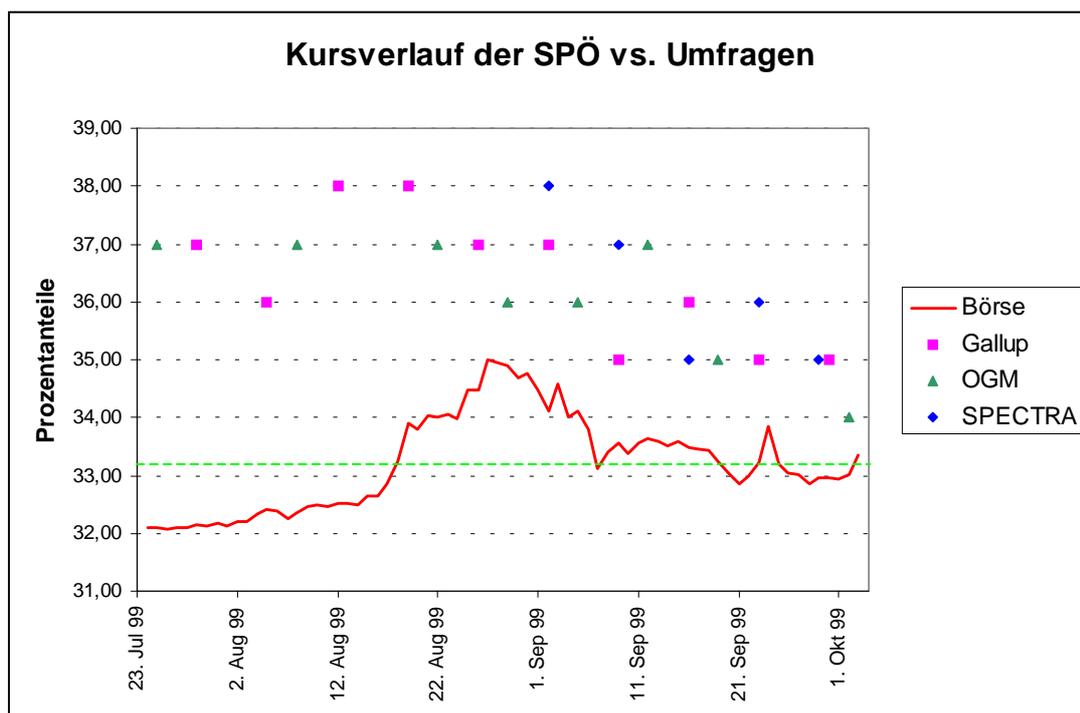


Abbildung 29: Kursverlauf und Umfragewerte der SPÖ im Vergleich;
Quelle: „Die Presse“, Umfrageinstitute

Damit gelang es der Wahlbörse schon frühzeitig, eine gute Prognose des Wahlausgangs zu liefern. Bereits im Juni, als die Umfragen noch sehr weit neben dem späteren Wahlergebnis der SPÖ lagen (37 bis 38 Prozent vs. einem Wahlergebnis von 33,15 Prozent), wurde die SPÖ zu ca. 32,50 gehandelt. Der Markt erfüllte die Vorhersagefunktion in diesem Fall gut. Bei den Aktien kleinerer Parteien traten jedoch zeitgleich die bekannten Verzerrungen (Überbewertung) auf. Somit könnte die frühzeitige gute Vorhersage des SPÖ-Wahlergebnisses eher zufällig durch die Unterbewertung der Aktien größerer Parteien entstanden sein und sollte nicht überbewertet werden.

Eines der bemerkenswertesten Ereignisse des Wahlkampfes war das „Überholen“ der ÖVP durch die FPÖ. Beide Instrumente diagnostizierten diesen Wechsel auf Platz zwei fast zeitgleich in der ersten Septemberwoche. Bemerkenswert ist, daß der Wechsel in den Umfragen durch eine „Wählerwanderung“ – Verluste bei der Volkspartei und Gewinne bei den Freiheitlichen“ erfolgte, während der gemeinsame Stimmenanteil der beiden Parteien konstant blieb. An der Wahlbörse konnte im selben Zeitraum die FPÖ zulegen, was jedoch v.a. auf Kosten der „Unabhängigen“ geschah,⁷³ der Kurs der ÖVP blieb dabei praktisch konstant (der Kursverlust der ÖVP vom 1. September erwies sich als kurzfristig, und war zwei Wochen später vollkommen ausgeglichen)

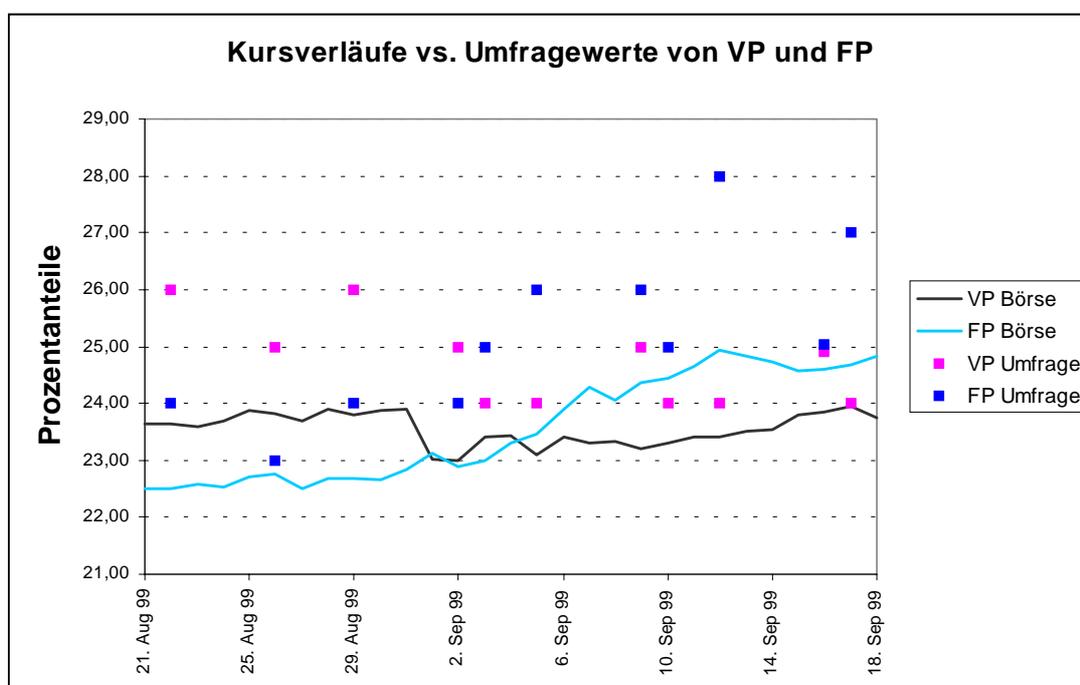


Abbildung 30: Kursverläufe und Umfragewerte von ÖVP und FPÖ Ende August und Anfang September; Quelle: „Die Presse“, Umfrageinstitute

⁷³ der Korrelationskoeffizient zwischen der FPÖ und der Aktie der „Unabhängigen“ betrug vom 2. September (erstmalige Notierung der „DU“-Aktie) bis zum Wahltag $-0,95$, sie bewegten sich also fast gegenläufig. Die Korrelation von FPÖ und ÖVP war hingegen nur leicht negativ ($\rho = -0,23$)

Diese Tendenz einer hohen Wählermobilisierung bei den „bürgerlichen“ Parteien, der später durch das Wahlergebnis bestätigt wurde,⁷⁴ „erkannte“ die Wahlbörse somit besser als die Umfragen.

Für uns bleibt festzuhalten, daß sich, nach den vorliegenden empirischen Daten zu urteilen, die Kurse an diesem politischen Aktienmarkt ziemlich unabhängig von Umfrageentwicklungen bildeten und veränderten. Aufgrund der relativ geringen Schwankungen sowohl von Umfrage- als auch Börsenwerten wäre es jedoch zu früh für eindeutige Aussagen, so daß wir zuerst noch einen Blick auf einen volatileren Markt an den IEM werfen.

USA

Wesentlich deutlicher wird der Kontrast zwischen Umfragen und Wahlbörsen bei WTA-Märkten, wie sie in den USA üblich sind. Als Beispiel ziehen wir das spannende Rennen zwischen George W. Bush und Al Gore um das Weiße Haus im Jahr 2000 heran. Der Wahlausgang war dabei bis zum Ende höchst ungewiß, was Umfrageinstitute und Märkte gleichermaßen vor eine große Herausforderung stellte. Daß der Markt dabei – wie die folgende Analyse zeigen wird – weitgehend unabhängig von Umfragedaten agierte, macht den Vergleich umso lohnender.

Bereits im Frühling und Sommer 2000 zeigte sich die unterschiedliche Entwicklung der beiden Instrumente, denn selbst als Bush in Umfragen regelmäßig um mehr als 10 Prozent vor seinem Konkurrenten lag, wurden an der Wahlbörse beide Kandidaten praktisch zum selben Preis gehandelt, wobei Gore teilweise sogar favorisierte war (Grafik in Kapitel 5.1 Dauer von Märkten).

Die Marktteilnehmer lassen sich in der frühen Phase des Wahlkampfes offenbar nur in geringem Maß von Umfragen leiten, sie verharren in einer abwartenden Position.⁷⁵ Dies kann durchaus als rational angesehen werden, weil der Ausgang

⁷⁴ Freiheitliche und Volkspartei erreichten zusammen fast 54 Prozent der Stimmen, während historisch zwischen 48 und 52 Prozent üblich waren

⁷⁵ so bleiben die Kurse in den USA meist um 49 Prozent für die beiden Kandidaten der etablierten Parteien, während in Europa meist das letztmalige Wahlergebnis, korrigiert um zu erwartende Verschiebungen, als Richtschnur dient

einer Wahlauseinandersetzung zu unsicher ist, um so früh prognostiziert werden zu können. Im Fall der US-Wahl 2000 kam hinzu, daß der republikanische Kandidat zwar in den Umfragen führte, daß aber die Erfahrung vergangener Wahlen gezeigt hatte, daß ein amtierender Vizepräsident bei guter Wirtschaftslage fast immer gewählt wurde. Daher dürfte sich die Umfrageschwäche von Gore lange Zeit nicht in den Kursen niedergeschlagen haben. In der ersten Phase eines Wahlkampfes bewegen sich Wahlbörsen somit praktisch unabhängig von Umfrageergebnissen.⁷⁶

Später schrumpfte der Vorsprung Bushs in den Umfragen, doch an der Wahlbörse stieg sein Kurs beständig. Wie ist dies zu erklären bzw. was hatte sich geändert?

Eine genaue Analyse der letzten fünf Wochen vor der Wahl mag hier Aufschluß bringen. In dieser Zeit veröffentlichte Gallup jeden Tag eine aktuelle Umfrage (*tracking poll* als Durchschnittswert der Befragungen der vorangegangenen drei Tage), so daß wir einen guten Vergleichsmaßstab für die Kurse an der IEM-Wahlbörse haben.⁷⁷ Die folgende Grafik bildet die Verläufe der Kurse an den IEM sowie die Umfragewerte von Bush und Gore zwischen 1. Oktober und 7. November 2000 ab. Die grüne Linie markiert dabei den Vorsprung Bushs in den Umfragen und wurde zur Verdeutlichung der Trends eingefügt. Anfang Oktober lagen beide Kandidaten in der Wählergunst gleichauf, doch an der Wahlbörse führte Gore deutlich mit ca. 60 Prozent Siegeswahrscheinlichkeit. Selbst als Bush in der folgenden Woche in den Umfragen bis zu elf Prozent Vorsprung aufwies, stieg Gores Kurs an der Wahlbörse auf knapp 70 Prozent. Wenige Tage später

⁷⁶ der Korrelationskoeffizient zwischen IEM-Kursen und den Gallup-Umfragewerten war in dieser Zeit (Ende April bis Ende August) annähernd null

⁷⁷ als Datensätze werden dabei die jeweiligen Tagesschlußkurse am WTA-Markt der IEM bzw. die Umfragewerte von Gallup am Tag ihrer Veröffentlichung verwendet. Alternativ könnte man die Gallupwerte auch am Tag ihrer Erhebung mit den IEM-Daten vergleichen, wobei die festgestellten Trends dabei insbesondere länger vor der Wahl besser übereinstimmen. In den letzten Wochen vor der Wahl ähneln sich jedoch die Bewegungen an IEM und Gallup am Tag der Veröffentlichung mehr. Dies könnte auf eine gesteigerte Umfrageabhängigkeit wenige Wochen vor einer Wahl hindeuten

schien es, daß die Käufer von Gore-Aktien Recht behalten würden, denn Bushs Vorsprung in den Umfragen sank auf nur noch ein Prozent. Erstaunlicherweise stieg in diesen Tagen aber der Kurs der Bush-Aktie von ca. 32 auf fast 40. Der Markt ignorierte hier also die Umfrageentwicklung, bzw. bewegte sich sogar gegenläufig zu dieser.⁷⁸

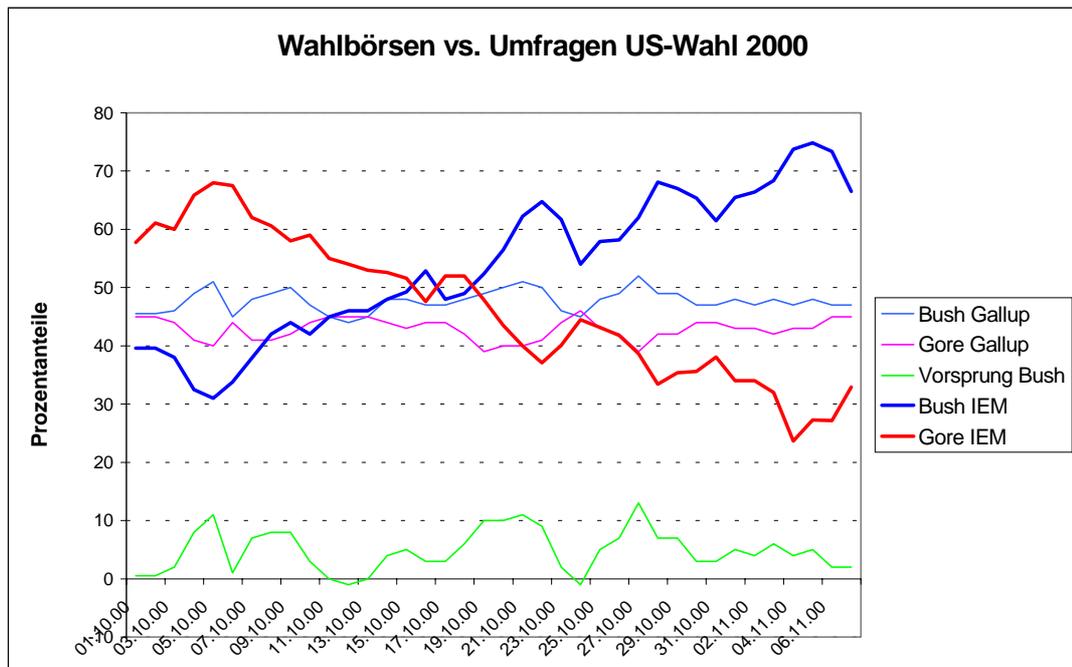


Abbildung 31: Vergleich von Wahlbörsen und Umfragen anhand der US-Wahl 2000; Quelle: IEM, Gallup

In den folgenden Wochen (bis 21. Oktober) führte Bush meist in den Umfragen, jedoch war sein Vorsprung dabei geringer als zu Beginn des Monats. Dennoch stieg sein Kurs beständig, so daß der Wert der Bush-Aktie Mitte Oktober den der Aktie auf Gore überstieg und nach diesem Zeitpunkt an den IEM stets höher als jene des Demokraten gehandelt wurde.

Es scheint, daß in dieser Zeit viele Händler ihre Strategie an der Börse änderten. Während bis dahin Gore fast unabhängig von Umfragewerten als Favorit galt, wurde nun Bush – wiederum weitgehend unabhängig von aktuellen Umfragedaten

⁷⁸ in der Zeit vom 1. bis zum 12. Oktober betrug der Korrelationskoeffizient der Bush-Aktie mit den Umfragewerten des Republikaners $-0,40$

– als wahrscheinlicher Sieger gehandelt. Der Markt, der lange in einer unentschiedenen Position verharret war, bildete nun eine seriöse Prognose des Wahlergebnisses. Während anfangs Erfahrungswerte und „Abwarten“ die Preisbildung bestimmten, wurden wenige Wochen vor der Wahl die Wahlchancen aufgrund aktueller Entwicklungen, Umfragen und Programme abgeschätzt. Das Ergebnis des Marktes war, daß der Sieger am Wahltag Bush heißen würde.

Das Erstaunliche ist, daß sich an dieser Haltung auch in den letzten zehn Tagen vor der Wahl nichts änderte. In dieser Phase konnte Gore Bushs Vorsprung in den Umfragen kontinuierlich von 13 auf 2 Prozent verringern.⁷⁹ Die Händler an der Wahlbörse beeindruckte dies jedoch nicht; Gores Kurs fiel weiter, während der Kurs des Republikaners auf über 75 Prozent Siegeswahrscheinlichkeit stieg. Viele Händler glaubten wohl wirklich an des Sieg von Bush, andere waren möglicherweise nur nicht gewillt, zu den gegebenen Kursen etwas an ihrem Portfolio zu ändern. Die folgende Abbildung zeigt noch deutlicher die teils sehr unterschiedliche Entwicklung der beiden Instrumente.

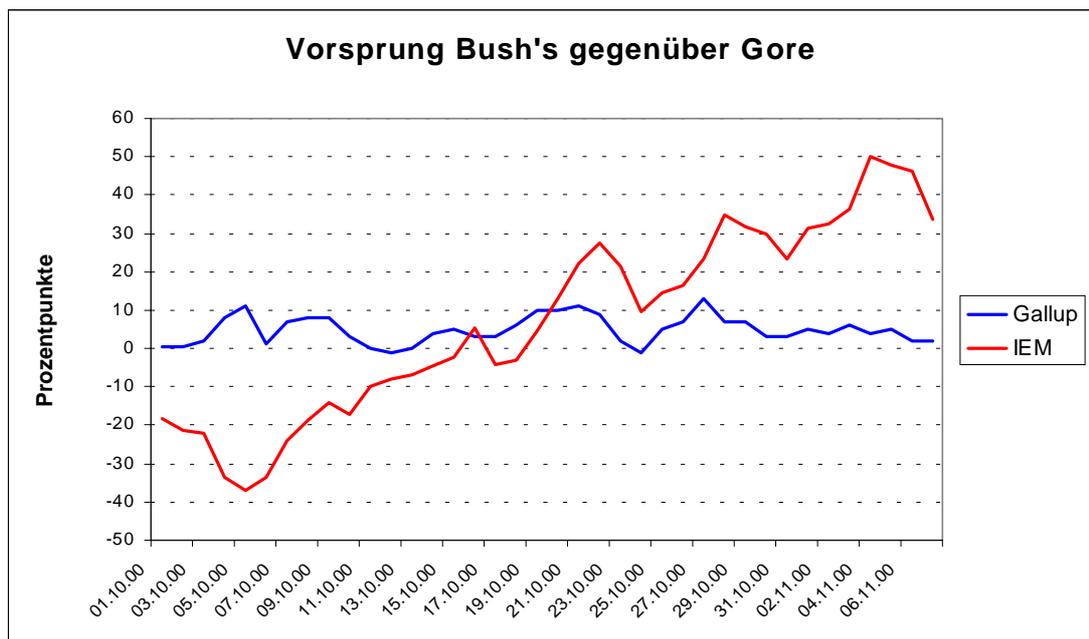


Abbildung 32: Vorsprung von Bush gegenüber Gore bei Gallup bzw. an den IEM; Quelle: IEM, Gallup

⁷⁹ in einigen Umfragen anderer Institute führte Gore sogar vor seinem republikanischen Gegner

In der Grafik wird nur noch die Differenz zwischen den Umfrage- bzw. Kurswerten von Bush und Gore dargestellt. Wieder erkennt man deutlich, daß die Kursverläufe sich nur ausnahmsweise ähneln.⁸⁰ Während beide Kandidaten Anfang Oktober und am Wahltag bei Gallup fast gleichauf liegen, wandelt sich in der selben Periode an der Wahlbörse der 20-Prozent Rückstand Bush's in einen Vorsprung von bis zu 50 Prozent. Relativ ähnlich sind sich die Kursverläufe nur vom 11. Oktober bis zum 2. November, während vorher wohl Erfahrungswerte zu einer Favorisierung von Gore führten, und nachher offenbar kein Meinungsumschwung an der Wahlbörse mehr stattfand.

Zusammenfassend läßt sich feststellen, daß Wahlbörsen offenbar drei Phasen durchlaufen. Nur in der mittleren Phase scheinen Umfragen eine wichtige Rolle für die Händler zu spielen, während sie vorher und nachher weitgehend ignoriert werden.

Die erste Phase ist eine des Abwartens, und dauert bis etwa zwei Monate vor der Wahl, der Markt verharrt in der Ausgangsstellung, die Kurse stellen noch keine seriöse Prognose des Wahlergebnisses dar. Daher ist eine lange Laufzeit für Wahlbörsen nicht erforderlich und hat bisher auch nicht zu besseren Ergebnissen geführt.

In der zweiten Phase kommt es meist zu intensivem Handel und sehr volatilen Kursen. Die Akteure haben sich ihre Meinung über den Wahlausgang gebildet und handeln dementsprechend am Markt. Angebot und Nachfrage sorgen dafür, daß die Meinungen und Informationen der einzelnen Händler aggregiert werden und eine gute Prognose des Wahlergebnisses liefern. Eine – aber keineswegs die einzige – Informationsquelle für Händler stellen dabei Umfragen dar.⁸¹

⁸⁰ der Korrelationskoeffizient beträgt +0,09, was den geringen Zusammenhang der beiden Datenreihen ebenfalls belegt

⁸¹ bei meiner Umfrage zur EU-Wahl 1999 in Österreich gaben 24 Prozent der Händler an, mit ihren Handelsentscheidungen auf Umfragen zu reagieren

Die letzte Woche vor der Wahl ist manchmal als eine neue Phase (Endphase) mit anderen Charakteristika als der zweiten zu identifizieren. Sie ist durch weiterhin hohe Umsätze, aber relativ geringe Kursbewegungen gekennzeichnet. Der Markt hat sein Gleichgewicht gefunden, an einem WTA-Markt steht der Sieger praktisch fest, und auch die Stimmenanteile der Parteien lassen sich häufig sehr gut abschätzen. Der Handel bleibt trotz der geringen Kursschwankungen jedoch aufgrund gesteigerten Interesses an der unmittelbar bevorstehenden Wahl und knapp plazierter limitierter Orders meist sehr aktiv. Ist die Wahl bis zuletzt extrem offen, so entfällt diese Phase.

Bei der in diesem Kapitel näher betrachteten Wahl in den USA ist diese Phase nur ansatzweise in der ersten Novemberwoche auszumachen. Zwar nahmen die Kursschwankungen in dieser Periode ab, jedoch fand der Markt kein wirklich stabiles Gleichgewicht – mit Recht, wie das knappe Wahlergebnis zeigte.

Finanzwirtschaftlich interessant ist, daß der Korrelationskoeffizient zwischen den Kursen an den IEM und den Gallupwerten dabei in der ersten und letzten Phase annähernd null war, in der zweiten Phase jedoch mit 0,48 (Bush) bzw. 0,35 (Gore) deutlich positiv ausfiel.⁸² Nur in dieser Handels- und Kursbildungsperiode bewegen sich Umfragewerte und Kurse also überwiegend in die selbe Richtung.⁸³

Zu klären ist nun, warum das geschieht – folgen Wahlbörsen zumindest in dieser Phase den Umfragen, oder bilden sie getrennt voneinander dieselben Trends. Falls sie unabhängig Trends erkennen, welches Instrument ist dabei schneller?

⁸² die erste Phase reicht dabei bis zum 10. Oktober, die zweite von 11. bis 31. Oktober und die dritte Phase umfaßt die erste Novemberwoche 2000

⁸³ eine gute Analyse für Deutschland liefert Brüggelambert (1999) ab Seite 155, wobei er einen Einfluß der Umfragen auf die Kurse an Wahlbörsen nicht ausschließen kann, keinesfalls wird den Umfragen aber bedingungslos gefolgt

8.4.2 Wie schnell werden wichtige Neuigkeiten als solche erkannt?

Eine entscheidende Frage ist, welches Instrument früher oder „besser“ fähig ist, einschneidende Neuigkeiten und Ereignisse als solche aufzuzeigen. Diese Frage ist nicht einfach zu klären, denn meist ist auch im Nachhinein nicht ohne weiteres festzustellen, wann eine wesentliche Änderung der Stimmung in der Bevölkerung eingetreten ist.

Wann überholte die FPÖ die ÖVP? Sowohl Umfragen als auch Wahlbörsen stellten diese Veränderung relativ präzise in der ersten Septemberwoche 1999 fest. Doch haben tatsächlich in diesem Zeitraum viele Wähler ihre Meinung geändert?

Exakt läßt sich der Zeitpunkt einer Änderung in den Präferenzen der Wähler nur dann ausmachen, wenn dieser Umschwung durch ein konkretes Ereignis (*event*) ausgelöst wurde oder zumindest zu einem feststellbaren Zeitpunkt erstmals auftrat.

Dazu muß das *event* einem großen Teil der Wähler bekannt, bzw. zugänglich sein. Beispiele sind etwa Fernsehduelle von Spitzenkandidaten, oder prägnante Äußerungen der Parteiführer, die den Wahlberechtigten von den Medien vermittelt werden. Um festzustellen, welches Instrument die Wirkung derartiger *events* besser abbilden kann, betrachten wir nun die Entwicklung von Wahlbörsen und Umfragen nach prägnanten Ereignissen und konzentrieren uns dabei wieder auf die beiden Wahlen zum österreichischen Nationalrat 1999 und zur US-Präsidentschaft im November 2000.

Aufgrund ihrer Struktur des ununterbrochenen Handels können politische Aktienmärkte schneller auf Neuigkeiten wie etwa Fernsehdebatten reagieren als Umfragen.⁸⁴ Dieser Zeitvorsprung ist jedoch für unsere Analyse unerheblich; es wäre sinnlos, diese Reaktionen miteinander zu vergleichen und dann festzustellen,

⁸⁴ jedoch können Kursreaktionen an Wahlbörsen durch limitierte Orders „gedämpft“ werden, so daß es einige Zeit dauern kann, bis sich eine signifikante Änderung in den Kursen voll niederschlägt

daß Umfragen zwei Tage länger brauchen, um Neuigkeiten zu erkennen. Für diese Analyse wird daher bei Umfragen stets der Zeitpunkt der Erhebung der Daten herangezogen. Am aussagekräftigsten sind *events*, in deren Folge sich Umfragen und Wahlbörsen in unterschiedliche Richtungen bewegten, doch werden wir sehen, daß solche Ereignisse sehr selten sind.

Österreich

In Österreich fallen die Ausschläge sowohl bei Umfragen als auch an den Wahlbörsen meist relativ schwach aus,⁸⁵ so daß es meist schwierig ist, eindeutig unterschiedliche Bewegungen festzustellen. Erschwerend kommt hinzu, daß es keine täglichen Umfragedaten gibt. Daher ist eine vergleichende Analyse nur sehr eingeschränkt möglich.

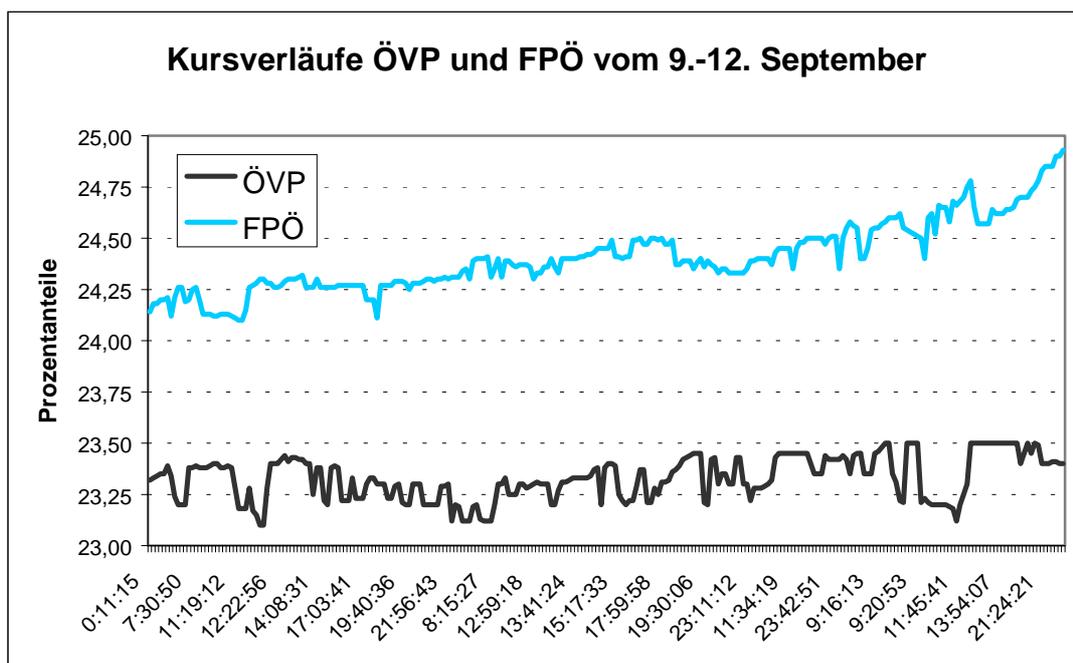
Die zwei wahrscheinlich markantesten Einzelereignisse des Wahlkampfes zur NRW 1999 waren die Ankündigung von ÖVP-Obmann Wolfgang Schüssel vom 9. September, bei einem Abrutschen auf „Platz drei“ in die Opposition zu gehen, sowie die Fernsehdebatte der fünf Parteiführer am Donnerstag vor der Wahl (30. September). Da nach der Fernsehdebatte keine Umfragerohdaten mehr erhoben wurden, wird dieses Ereignis hier nicht eingehender behandelt. Wenden wir uns also den Ereignissen Anfang September 1999 zu.

Am 9. September, etwa vier Wochen vor der Wahl, verkündete der Spitzenkandidat der Volkspartei, daß er seine Partei in die Opposition führen würde, falls sie in der Wählergunst auf Platz drei zurückfiele. Damit reagierte er auf Umfragewerte, die etwa eine Woche vorher ein Abrutschen seiner Partei hinter die FPÖ registriert hatten.

Die Abbildung auf der nächsten Seite zeigt die Entwicklung der Kurse von ÖVP und FPÖ am Tag der „Oppositionsansage“ sowie an den drei folgenden Tagen. Betrachtet man die Kursentwicklung an der Wahlbörse, so scheint er damit v.a.

⁸⁵ bei den Wahlbörsen liegt das v.a. in ihrer Konstruktion als VS-Märkte begründet, bei Umfragen in der Glättung durch die Institute

erreicht zu haben, daß den Umfragen, also der Stärke der FPÖ mehr Glauben geschenkt wurde. Der Kurs der ÖVP bewegte sich auch in den Tagen nach der Ankündigung kaum, während jener der FPÖ seinen damals schon bestehenden Aufwärtstrend beschleunigt fortsetzte. Inwiefern dies jedoch mit dem *event* zusammenhing, läßt sich nur schwer feststellen.



Deutlicher als am VS-Markt der Stimmenanteile war die entsprechende Reaktion am parallel stattfindenden WTA-Markt über die zukünftige Regierungskoalition: am 9. September verdoppelte sich die Aktie für eine SPÖ-Alleinregierung von 5,6 auf 12,6 Prozent, während die Aktien für die Große und die Kleine Koalition (ÖVP und FPÖ) um 6,4 bzw. 3,2 Prozent sanken. Dies deutet darauf hin, daß der Markt die Ankündigung durchaus ernst nahm und auch Platz drei für die ÖVP für durchaus realistisch hielt. Leider liegen dazu aber keine vergleichbaren Umfragewerte vor, so daß kein Vergleich möglich ist. Wenden wir uns daher nun den (in dieser Hinsicht aussagekräftigeren) Märkten in den USA zu.

USA – Erfahrungen an den IEM

Beim Vergleich von Umfragen und PSMs anhand des US-Wahlkampfes 2000 können wir ab Anfang Oktober auf tägliche Umfragedaten zurückgreifen, was eine präzisere und aussagekräftigere Analyse erlaubt.⁸⁷

Bis Anfang September bewegten sich die Kurse sowohl für Bush als auch für Gore unabhängig von Umfragen und Wahlkampfneuigkeiten stets um 50 Prozent. Daran änderten auch die Parteikonvente im August nichts, obwohl die Umfragen in deren Folge um ca. 20 Prozentpunkte fluktuierten. Die Händler an den IEM erkannten den Effekt dieser Versammlungen (*convention boost*) als kurzfristig und ignorierten ihn folgerichtig. Hier erwies sich der Markt als wesentlich robuster als Umfragen.

Jene Ereignisse, die von den meisten Wählern verfolgt wurden, waren die drei Fernsehdebatten zwischen Bush und Gore, die am 3., 11. und 17. Oktober stattfanden. Gegenüber dem steif, aber souverän auftretenden Gore galt Bush eher als Außenseiter in den Debatten, so daß viele Beobachter erwarteten, er würde die Wahl bei diesen Konfrontationen verlieren.

Sehen wir uns die Umfragewerte kurz vor und nach diesen Debatten an und vergleichen sie mit den entsprechenden Kursen an den IEM. Jede der drei grünen

⁸⁷ als Daten dienen dabei die Werte der *tracking poll* von Gallup sowie die mittleren Handelskurse des WTA-Marktes an den IEM

Linien bezeichnet eine der Fernsehdebatten. Beachtenswert ist, daß Gore vor den Debatten an den IEM deutlich in Führung war, obwohl sein Kontrahent in den Wochen davor in den Umfragen kontinuierlich vorne lag.

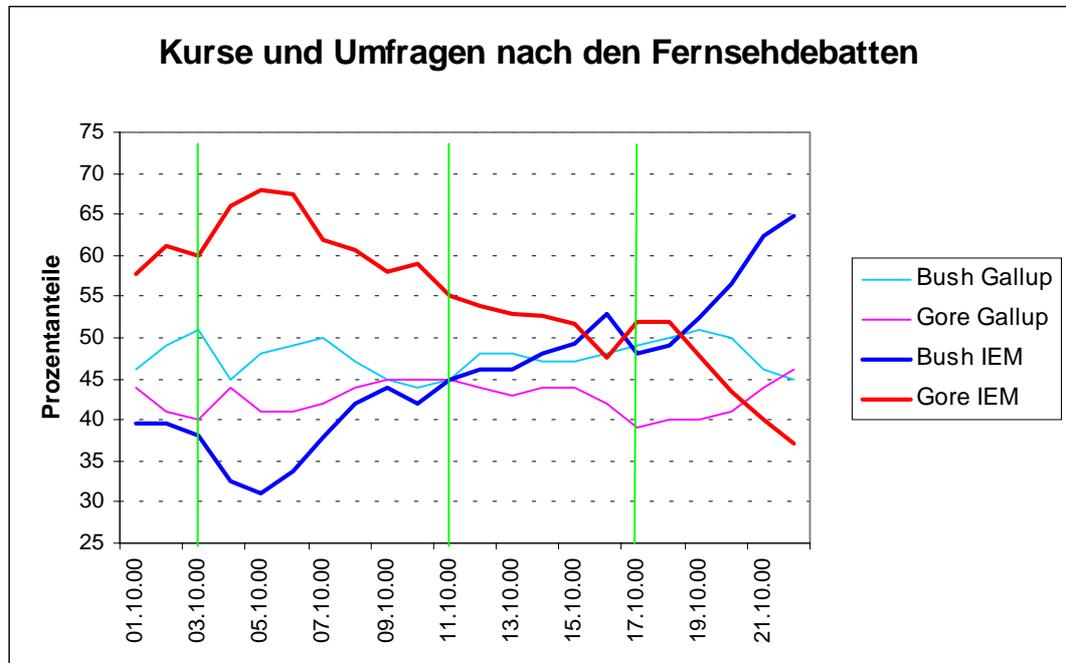


Abbildung 34: Kurse und Umfragewerte in den USA vor und nach den Fernsehdebatten; Quelle: IEM, Gallup

In der ersten Debatte am 3. Oktober hatte nach übereinstimmender Meinung der meisten Beobachter Gore die überzeugenderen Argumente und konnte Sympathien gewinnen. Dies wurde am folgenden Tag auch in den Umfragen reflektiert, denn Gore konnte dabei um fast 5 Prozent zulegen, während Bush ebensoviel verlor. Auch an den IEM stieg der Kurs des Demokraten deutlich von 60 auf 66 Prozent. In der Woche bis zur nächsten Debatte entwickelten sich Umfragen und die Wahlbörse jedoch unterschiedlich. Während Bush in den Umfragen nach einem Zwischenhoch sogar hinter Gore zurückfiel, stieg sein Kurs am Markt beständig (von 31 auf 40 Prozent). Am Markt nahm offensichtlich die Zahl derer zu, die einen Sieg von Bush für wahrscheinlich hielten, obwohl aus den Umfragen eher das Gegenteil zu schließen war.

Die Außenpolitik war das Schwerpunktthema der zweiten Konfrontation – angeblich ein Gebiet, über das Bush nicht so gut Bescheid weiß. Doch der Republikaner präsentierte sich überraschend souverän und konnte Gore wiederholt in die Defensive drängen, so daß die Mehrzahl der Kommentatoren Bush als Sieger der Debatte sah. Parallel dazu verbesserten sich seine Umfragewerte, so daß er wieder 10 Prozent Vorsprung vor dem Demokraten aufwies. Auch an der Wahlbörse konnte der Republikaner stark zulegen und überholte seinen Gegner einen Tag vor der dritten Debatte. Damit entwickelten sich nach der zweiten Debatte Umfragen und Wahlbörsen im Gleichschritt. Die dritte Debatte brachte schließlich unterschiedliche Trends, die wir näher betrachten wollen.

Nach der dritten TV-Konfrontation der beiden Kandidaten waren sich die Kommentatoren nicht einig, wer besser abgeschnitten hatte, doch eine Mehrheit meinte, daß Gore überzeugender war. Darauf deuteten auch die Umfragen der nächsten Tage hin. Bushs Vorteil blieb zwar noch zwei Tage erhalten, bröckelte dann aber ab und war nach fünf Tagen von elf Prozent Vorsprung in einen Rückstand von einem Prozentpunkt umgekehrt. Ganz anders die Entwicklung an den IEM. Der Kurs des Republikaners stieg in den Tagen nach der Debatte steil an. In weniger als einer Woche erhöhte sich der Preis der Bush-Aktie von 48 auf 65 Prozent.

Damit haben wir erstmals zwei gegenläufige Trends bei Umfragen und Wahlbörsen, wobei in den Umfragen der spätere Verlierer favorisiert wird, während die Bewegung am Markt in die richtige Richtung verläuft. Dieses Ergebnis soll jedoch weder überbewertet noch verallgemeinert werden, denn der Ausgang dieser Wahl war so knapp, daß eine Wahlprognose einem Glücksspiel ähnlicher war als einer wissenschaftlichen Untersuchung.

Ein einschneidendes Ereignis war auch die Wahl selbst, denn der Ausgang war noch Wochen später ungewiß. Da die Wahlbörse an den IEM noch drei Tage nach der Wahl am 7. November offen gehalten wurde, können wir betrachten, wie die Händler auf diese Unsicherheit reagierten.

In der Abbildung werden die Tageshöchst-, Tiefst- und Schlußkurse für Gore gezeigt. Man sieht, daß die Aussichten der Kandidaten nach der Wahl falsch

eingeschätzt wurden: Gore wurde mit über 90 Prozent Siegeswahrscheinlichkeit gehandelt, während sein Kurs vor der Wahl meist um 30 Prozent lag. Man erkennt aber auch, daß die Kursschwankungen, also die Volatilität, am Wahltag und dem Tag darauf mit 70 bzw. 97 Prozent sehr hoch ist, was die am Markt herrschende Verunsicherung widerspiegelt.

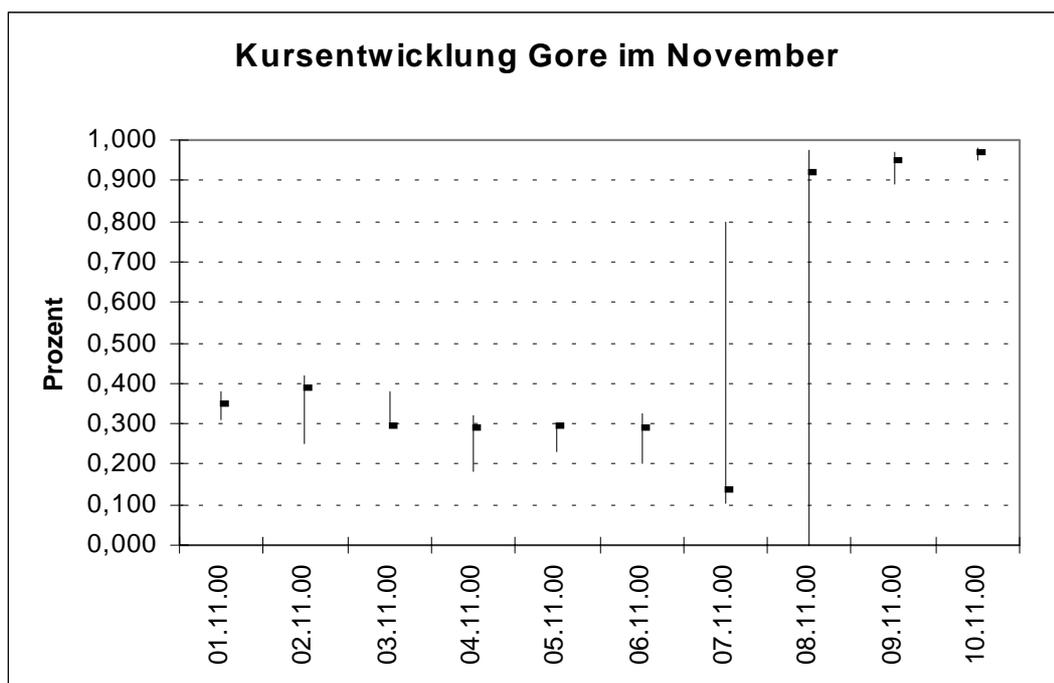


Abbildung 35: Kursentwicklung Gore im November; Quelle: IEM

Die entsprechende Grafik für Bush sieht fast spiegelbildlich aus, so daß hier auf ihre Darstellung verzichtet wird. Bedauerlicherweise ist kein Vergleich mit Umfragen möglich, da entsprechende Daten nicht vorliegen, doch wird an dem Beispiel auch so deutlich, daß auch Wahlbörsen falsch liegen können.

Andere Beispiele

In diesem Zusammenhang sei noch einmal an zwei Märkte, die bereits im ersten Abschnitt erwähnt wurden, erinnert: die IEM-Börsen anlässlich der Präsidentschaftswahlen in Taiwan und Mexiko im Jahr 2000 (vgl. Kapitel 5.5 Prognosegenauigkeit). In Taiwan war das Rennen zwischen drei Kandidaten völlig offen, denn alle drei lagen in den Umfragen stets gleichauf. Auch an der Wahlbörse wurden sie zu fast identischen Kursen gehandelt. Doch als in der letzten Woche

vor der Wahl keine Umfragen mehr publiziert werden durften, kristallisierte sich am Markt der spätere Sieger als eindeutiger Favorit heraus.

Während der Markt im Falle Taiwans eine bemerkenswerte Prognoseleistung erbrachte, versagte er im Falle Mexikos ebenso deutlich. In diesem Fall lagen zwei Kandidaten in den Umfragen fast immer gleichauf. Die Mehrzahl der Händler glaubte kurz vor der Wahl an den Sieg des Kandidaten der Regierungspartei, dessen Kurs sich bei ca. 70 Prozent Siegeswahrscheinlichkeit einpendelte. Die Prognose erwies sich aber als falsch, denn der Oppositionskandidat konnte sich schließlich relativ deutlich mit einem Stimmenvorsprung von 6,5 Prozentpunkten durchsetzen. Es ist jedoch anzumerken, daß beide Märkte mit extrem niedrigen Händlerzahlen durchgeführt wurden, so daß aus diesen Einzelbeobachtungen keine zu drastischen Schlüsse gezogen werden sollten.

Wir haben gesehen, daß sowohl Umfragen als auch Wahlbörsen Trends manchmal sehr gut erkennen, in anderen Fällen aber auch falsch liegen. Dabei werden Befragungen oft wegen der teils enormen Schwankungen innerhalb kurzer Zeitspannen kritisiert. Widmen wir unsere Aufmerksamkeit kurz dieser Problematik.

8.4.3 Kurzfristige Schwankungen

Es wurde bereits gezeigt, daß die Kurse von Wahlbörsen v.a. längere Zeit vor der Wahl meist wesentlich weniger schwanken als die entsprechenden Umfragewerte. Insbesondere an VS-Märkten zeichnen sich die Preise durch geringe Volatilitäten aus und auch bei WTA-Märkten treten größere Schwankungen meist erst in den letzten zwei bis drei Monaten vor einer Wahl auf. Diese Stabilität dürfte in der Fragestellung einer Wahlbörse begründet sein – die Händler sind von Anfang an aufgefordert, langfristige Erwartungen in den Kursen zu berücksichtigen. Die Erwartungen unterliegen in der Regel geringeren Schwankungen als die Stimmungen, die mit Umfragen erhoben werden.

Österreich

In Österreich werden Börsen meist als VS-Märkte durchgeführt, so daß die Kurse fast immer stabiler sind als die entsprechenden Umfragewerte. Dabei ist auch auffallend, daß die Umfragewerte der einzelnen Institute bei gleichem Erhebungszeitraum teils deutlich voneinander abweichen. Dies sollte bei ähnlichen Erhebungsmethoden der einzelnen Meinungsforscher eigentlich nicht vorkommen. Die folgende Grafik zeigt beispielhaft den Kursverlauf der ÖVP bei der Nationalratswahl 1999 sowie die entsprechenden Umfragewerte, wobei leicht zu erkennen ist, daß die Umfragewerte mehr streuen als die Kurswerte.⁸⁸

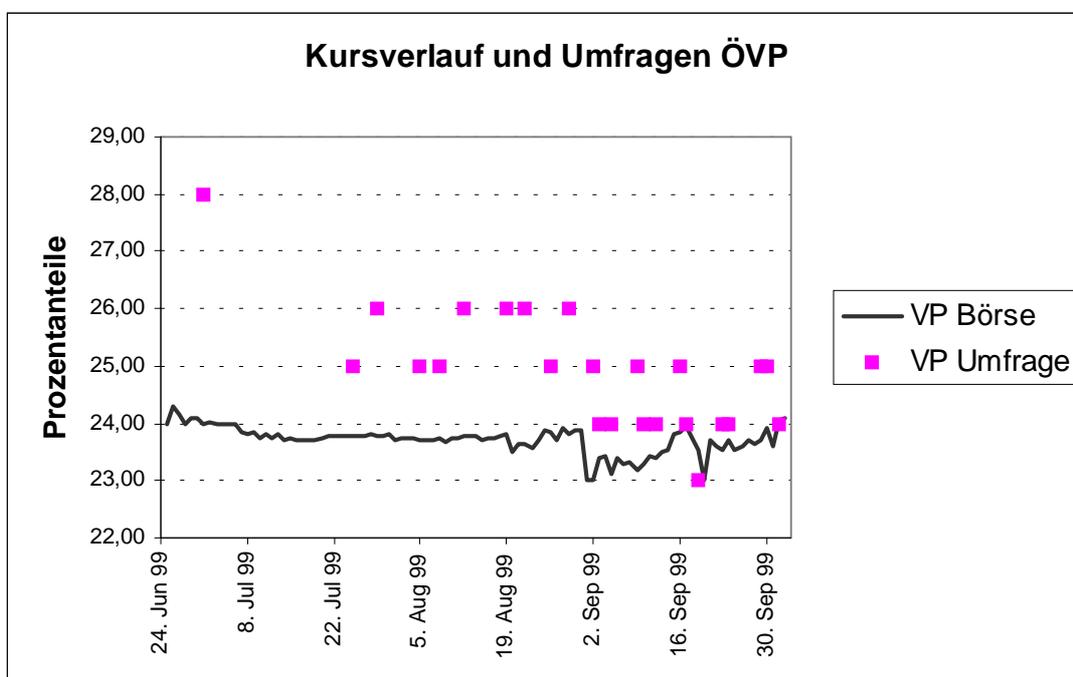


Abbildung 36: Umfragewerte und Kurse der ÖVP im Vergleich; Quelle: „Die Presse“, Umfrageinstitute

USA

In den Vereinigten Staaten dürfen nur ungeglättete Umfragen veröffentlicht werden, daher sind die Schwankungen noch größer als in Europa. Betrachten wir dies am besten anhand empirischer Daten. Die folgende Abbildung zeigt die

⁸⁸ die Standardabweichung der Kurse beträgt 0,23 Prozent, während die der Umfragen mit 1,06 Prozent mehr als vier mal so groß ist

Schwankungen in den Umfragewerten laut Gallup sowie am VS-Markt der Wahlbörse. Der einmalige „Ausreißer“ der Aktie von Gore ist dabei auf den teils illiquiden Markt zurückzuführen.

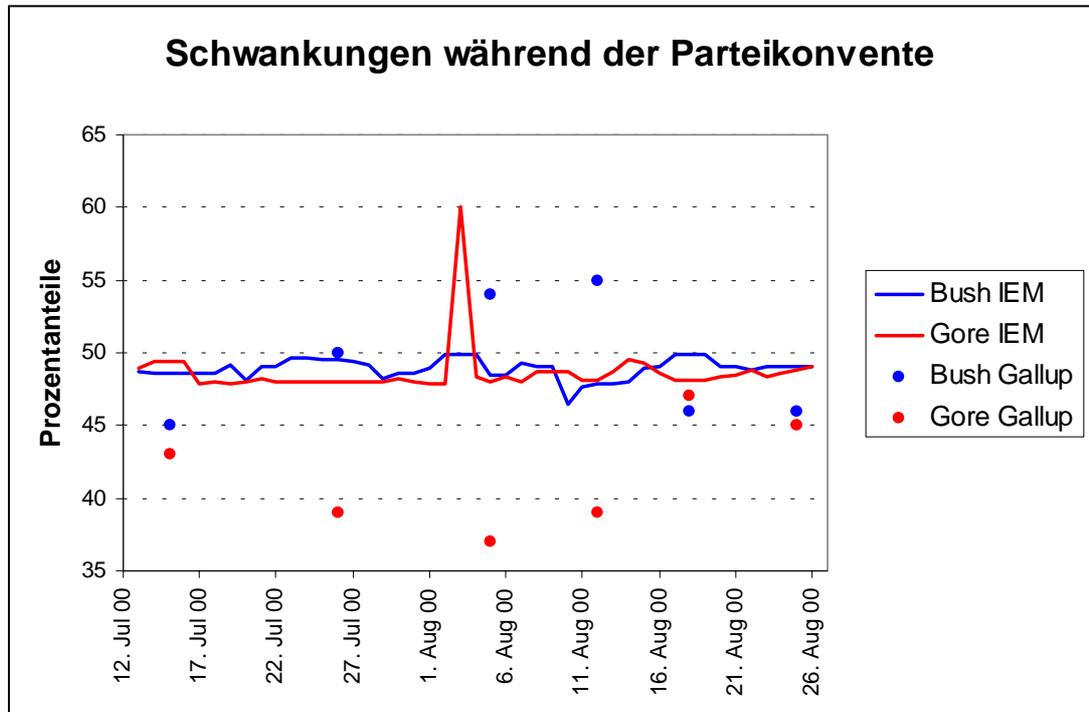


Abbildung 37: Schwankungen in der Zeit der Parteikonvente; Quelle: IEM, Gallup

Der Parteikonvent der Republikaner fand am 5. August seinen Abschluß, jener der Demokraten am 15. August. Man sieht, daß Bushs Vorsprung laut Gallup unmittelbar nach „seinem“ Konvent auf 17 Prozent angewachsen war. Nur zehn Tage später führte jedoch schon Gore – beflügelt vom Erfolg der demokratischen Versammlung. Diese kurzfristige Umfragestärke der Kandidaten nach deren jeweiligen Konventen sind in den USA schon lange bekannt (*convention boost*), daher reagierten die Märkte kaum auf die Parteiversammlungen.

Auch auf der folgenden Abbildung ist zu erkennen, daß die Umfragedaten stärker schwanken, als die Kurse an einem VS-Markt.⁸⁹ Gezeigt werden die täglichen Werte Gallups im Vergleich zu den Kursen an der IEM vom 1. Oktober bis zum Wahltag.

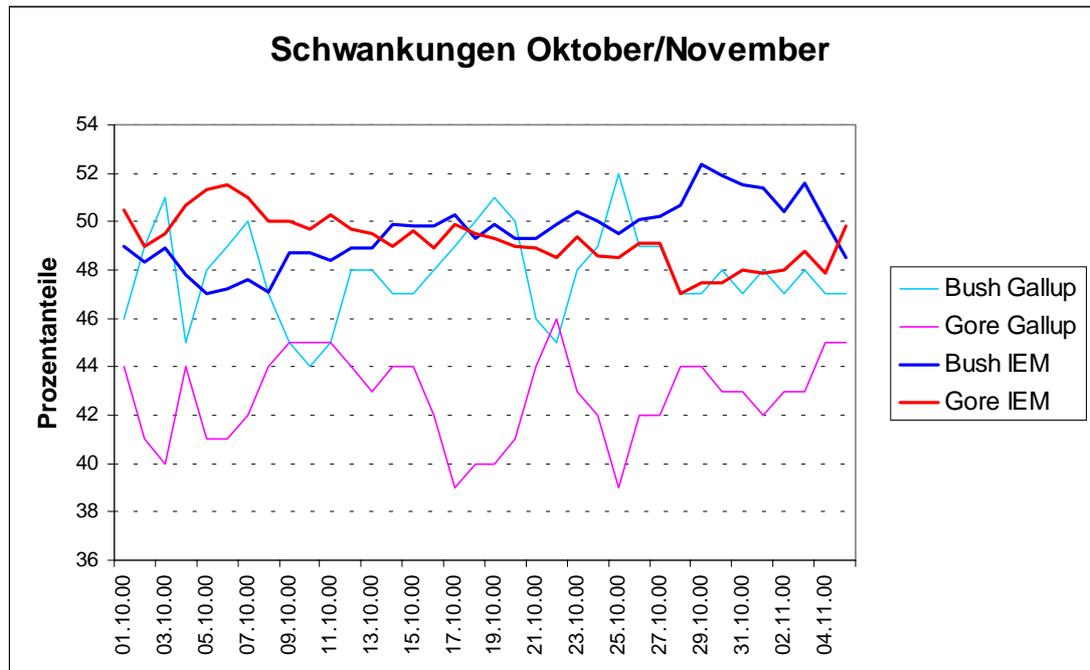


Abbildung 38: Schwankungen von VS-Kursen und Umfragen im letzten Monat vor der Wahl; Quelle: IEM, Gallup

In den USA schwanken selbst die Umfragewerte eines Instituts innerhalb einer Woche um bis zu 20 Prozent. Während solch extreme Ausschläge in Europa durch die Meinungsforscher „geglättet“ werden, sehen sich die amerikanischen Firmen mit der Frage konfrontiert, wie es zu solchen Schwankungen kommen kann. Die Antwort der bekanntesten Agentur ist ebenso überraschend wie entwaffnend: bei Gallup ist man überzeugt, daß die Stimmungen tatsächlich so stark variieren, wie dies erhoben wird. Damit steht das Institut zwar treu zu seinen

⁸⁹ die Standardabweichung ist bei den Umfragedaten in der Regel etwa doppelt so hoch wie bei den Kursen; die Schwankungen an WTA-Märkten sind hingegen meist größer als jene der Umfragen

Methoden und Prinzipien, jedoch wird die Möglichkeit von Erhebungsfehlern etwas zu kategorisch zurückgewiesen.

Wie wir aus Abschnitt 7.5 Statistische Grundlagen für den Einsatz von Stichproben wissen, liegt der „wahre Wert“ einer Merkmalsausprägung bei einer Erhebung mit einer Samplegröße von ca. 1000 zu 95 Prozent innerhalb von $\pm 3,2$ Prozent um den erhobenen Wert. Dies impliziert ein Spanne des Konfidenzintervalls von 6,4 Prozent, selbst da wird man in 5 Prozent der Fälle noch ein falsches Ergebnis erhalten. Um zu 99 Prozent sicher zu sein, müßte das Intervall auf 8,1 Prozent erhöht werden,⁹⁰ womit man nicht mehr von einem präzisen Ergebnis sprechen kann. Daher veröffentlichen Institute nur den wahrscheinlichsten Wert ihrer Erhebungen, was zu den erheblichen Schwankungen beitragen dürfte.

8.4.4 Prognosegenauigkeit

Der Wert einer Prognose läßt sich an zwei Dimensionen bemessen: an der Zeitigkeit ihrer Erstellung und an der Präzision der Vorhersage.

Es ist evident, daß eine korrekte Prognose eines Ereignisses umso höher einzuschätzen ist, je früher sie vorliegt. In dieser Hinsicht kann man Wahlbörsen nach den vorliegenden Daten ein tendenziell besseres Zeugnis ausstellen als Umfragen, denn oft liegen die Kurse schon Monate vor der Wahl nahe am Endergebnis, während Umfragewerte zu dieser Zeit teils noch erheblich vom späteren Wahlergebnis abweichen. Das hat seine Gründe v.a. in der unterschiedlichen Fragestellung, wodurch bei Umfragen die aktuelle Stimmung, bei Wahlbörsen hingegen längerfristige Erwartungen dominieren.

⁹⁰ der Standardfehler von 1,58 Prozent mal 2,56 (99% Wahrscheinlichkeit) ergibt eine Konfidenzintervallspanne von 8,09 Prozent

Die letztlich entscheidende Frage bleibt jedoch, welches Instrument das Wahlergebnis präziser vorhersagen kann. Sehen wir uns dazu zuerst eine Abbildung über die Schätzfehler einige Wahlbörsen und die entsprechenden Umfragen an.⁹¹

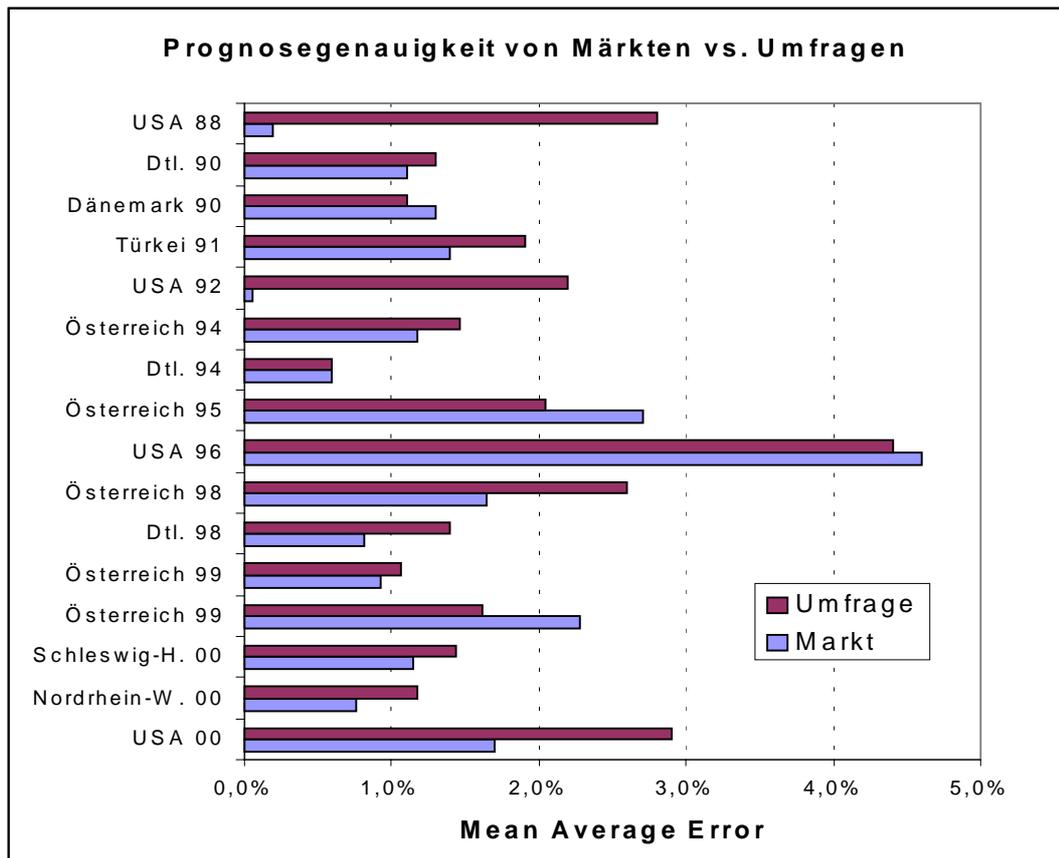


Abbildung 39: Wahlbörsen und Umfragen im Vergleich – international

Im Schnitt lagen Wahlbörsen in ihren Prognosen geringfügig besser als Umfragen, doch sollte man nicht übersehen, daß die Teilnehmer an PSMs auf Umfrageergebnisse zurückgreifen können um die Gesamtsituation zu beurteilen, was umgekehrt nicht der Fall ist. Betrachten wir einige spezielle Problemstellungen, um einen besseren Eindruck von Möglichkeiten und Grenzen der beiden Instrumente zu bekommen.

⁹¹ bei den angegebenen Märkten handelt es sich ausschließlich um VS-Märkte, bei den Umfragen handelt es sich meist um jene mit der jeweils besten Vorhersage

Prognosen bei erstmalig stattfindenden Wahlen

Solange Meinungsforscher bei ihren Vorhersagen auf Erfahrungswerte aus früheren Wahlgängen zurückgreifen können, sind halbwegs gute Prognosen (MAE um zwei Prozent) fast zu erwarten. Nur wenn sich die politische Landschaft erdrutschartig verändert oder Wahlen erstmalig stattfinden, verfügen wir über eine wirkliche Meßlatte der Prognosekunst.

Bedauerlicherweise haben sich hier beide Methoden nicht besonders bewährt: ein Testfall für die Umfrageforschung war die erste freie Wahl in Ostdeutschland 1990, wo die Institute desaströs versagten. Alle Umfragen überschätzten den Anteil der SPD und unterschätzten jenen der Union. So prognostizierte eine Erhebung den Anteil der Sozialdemokraten mit 34, jenen der CDU mit 23 Prozent. Das Wahlergebnis betrug schließlich nur 23 Prozent für die SPD und 42 Prozent für die CDU (vgl. Diekmann 1995, S. 389). Doch auch bei Urnengängen, die wiederholt durchgeführt wurden, konnten die Meinungsforscher selbst in jüngster Vergangenheit oftmals keine gute Vorhersage des Ergebnisses liefern.⁹²

Auch Wahlbörsen stellen bisweilen schlechte Prognosen, so etwa bei der EU-Volksabstimmung in Österreich 1995. Bei diesem einmaligen Ereignis konnte man auf keine Erfahrungswerte zurückgreifen, das Ergebnis wurde relativ deutlich verfehlt.⁹³ Zwar lagen Umfragen auch nicht besser, doch das Scheitern von Umfragen kann keine Rechtfertigung für das Versagen von PSMs sein.⁹⁴

Abweichungen bei kleinen Parteien

Fast bei jeder Wahl weichen die Kurse bzw. Umfragedaten vom Wahlergebnis ab. Erstaunlicherweise hängen die Abweichungen dabei kaum von der Größe der

⁹² Beispiele für Fehleinschätzungen durch die Institute waren etwa die Landtagswahlen 2000 in der Steiermark und im Burgenland sowie 2001 die Gemeinderatswahl in Wien

⁹³ die tatsächliche Zustimmung betrug 67 Prozent – gegenüber 55 Prozent in der Prognose

⁹⁴ ebensowenig wie die „Nicht-Repräsentativität“ der überwiegend studentischen Händlerschaft – das Ergebnis bei den Studenten betrug tatsächlich 55 Prozent, so daß der Markt zumindest diese Zustimmungsrates korrekt abbildete

Partei ab. Eigentlich sollte man annehmen, daß die Schätzfehler bei größeren Parteien eher höher wären als bei kleinen Parteien,⁹⁵ jedoch konnte dies weder bei Umfragen noch bei politischen Aktienmärkten beobachtet werden.

Kehren wir zu unserem Beispiel der letzten Nationalratswahl in Österreich zurück. In der nachfolgenden Tabelle sind jeweils die beiden Werte mit den größten Schätzfehlern rot, jener Wert mit dem geringsten Schätzfehler blau markiert.

	Wahlergebnis ⁹⁶	Wahlbörse	Gallup	OGM	Spectra
SPÖ	33,15	33,34	35	34	35
FPÖ	26,91	26,80	27	28	25
ÖVP	26,91	24,08	25	24	25
Grüne	7,40	7,19	7	6	7
LIF	3,65	4,40	2	4	5
DU	1,02	2,99	4	3	1
Andere	0,96	2,34	0	1	2
MAE⁹⁷	-	1,06	1,41	1,23	1,21

Tabelle 9: Schlußprognosen der Institute vs. Endergebnis der Wahlbörse bei der Österreichischen NRW 1999; Quellen: „Die Presse“, Umfrageinstitute

Es läßt sich kein eindeutiger Zusammenhang zwischen dem Prognosefehler und dem Stimmenanteil einer Partei erkennen, Schätzfehler treten sowohl bei Umfragen als auch an der Wahlbörse unabhängig von der Parteigröße auf.⁹⁸ So

⁹⁵ so wäre es möglich anzunehmen, daß die Abweichungen stets etwa ein Zehntel des Stimmenanteils betragen könnten – also 3 bis 4 Prozent bei großen Parteien und etwa ein halben Prozent bei kleineren Parteien

⁹⁶ offizielles Endergebnis inklusive Briefwähler und Auslandswahlkarten laut Innenministerium

⁹⁷ der MAE wird dabei immer bezogen auf sechs Parteien berechnet, um die Vergleichbarkeit zu gewährleisten, Umfragen wurde jedoch vereinzelt mit nur fünf Parteien, die Wahlbörse hingegen mit 7 Parteien durchgeführt (dies würde den MAE zugunsten der Wahlbörse verzerren)

⁹⁸ der MAE ist an der Wahlbörse mit 1,04 Prozent für die drei „Großen“ und 1,08 Prozent bei den anderen Parteien praktisch ident. Bei Gallup ist der Schätzfehler bei den kleineren Parteien im schnitt um 0,3 Prozent höher als bei den größeren, während insbesondere Spectra die Stimmenanteile der kleinen Parteien besser vorhersagen konnte als die von SPÖ, FPÖ und ÖVP

wurden am Markt SPÖ, FPÖ und Grüne sehr gut geschätzt, während bei ÖVP, DU und „Andere“ erhebliche Abweichungen festzustellen sind. Gallup überschätzte insbesondere die SPÖ und DU, während OGM bei Grünen und der ÖVP am weitesten daneben lag. Abgesehen von der Überschätzung des Stimmenanteils der SPÖ läßt sich somit selbst beim Vergleich der Umfrageinstitute untereinander keine eindeutige Tendenz feststellen.

Dieses Phänomen ist schon länger bekannt und doch kaum zu erklären. Während Abweichungen von 1,0 bis 1,5 Prozent bei größeren Parteien etwa einem Dreißigstel des Stimmenanteils entsprechen und damit innerhalb des 95-Prozent Konfidenzintervalls bleiben, sind derartige Abweichungen bei kleineren Parteien kaum zu verstehen. Wird die Gruppierung „DU“ bei einem Stimmenanteil von einem Prozent in Umfragen mit vier, an der Börse mit drei Prozent gehandelt, so wird die Wählerunterstützung damit um 200 bis 300 Prozent überschätzt. Bei einer Erhebung mit 1000 Befragten beträgt selbst das 99-Prozent Konfidenzintervall für eine 1-Prozent-Partei 0,2 bis 1,8 Prozent.⁹⁹ Nur in einem Prozent der Fälle sollte ein Wert außerhalb dieser Bandbreite gemessen werden. Trotzdem weichen zwei von drei Umfragen wesentlich weiter ab. Ebenso gelingt der Wahlbörse in diesem Fall keine gute Prognose.

Auch im Falle des Liberalen Forums liegen zwei der drei Umfrageinstitute mit ihrer Schätzung außerhalb des 99-Prozent Vertrauensintervalls. In der folgenden Tabelle zu Umfrage- und Marktdaten anlässlich der Präsidentenwahl 1998 in Österreich wird die Problematik etwas anders dargestellt: Neben dem Wahlergebnis und den Prognosen wird die relative Abweichung angegeben. Damit ist der Grad der Abweichung relativ zum Stimmenanteil gemeint.¹⁰⁰

⁹⁹ die Abweichungen bei DU und LIF betragen bis zu 14 Standardabweichungen, was laut Wahrscheinlichkeitstabellen praktisch nie vorkommen dürfte

¹⁰⁰ bei gleicher absoluter Abweichung nimmt diese bei kleineren Parteien zu: 2 Prozent Abweichung stellen bei 40 Prozent Stimmenanteil eine relative Abweichung von 5 Prozent dar, bei 8 Prozent Stimmenanteil beträgt der relative Wert bereits 25 Prozent

	Wahl	Börse	Abw.	Gallup	Abw.	Fessl	Abw.	market	Abw.
Klestil	63,49	60,83	4,19	66	3,95	60	5,50	61	3,92
Knoll	13,53	17,03	25,87	16	18,25	20	47,82	12	11,31
Schmidt	11,08	11,00	0,72	10	9,75	10	9,75	8	27,80
Lugner	9,94	7,99	19,62	6	39,64	8	19,52	4	55,73
Nowak	1,95	2,00	2,56	1	48,72	1	48,72	1	48,72
MAE	-	1,65	-	2,19	-	2,79	-	2,72	-

Tabelle 10: Grad der relativen Abweichung bei der österreichischen Präsidentenwahl 1998, alle Angaben in Prozent; Quelle: APSM-homepage

Anhand dieser Wahl läßt sich die Abhängigkeit des Schätzfehlers vom Stimmenanteil eines Kandidaten sehr gut zeigen, denn ein Kandidat bekam über 60 % der Stimmen, während sich nur 2 % der Wähler für Peter Nowak entschieden. Die übrigen Kandidaten erhielten zwischen 10 und 13,5 % der Stimmen. Während die relativen Abweichungen an der Wahlbörse streuen, läßt sich bei den Umfragen eine deutliche Tendenz feststellen. Die Abweichung fällt bei Klestil stets am geringsten aus und ist bei Nowak immer die höchste oder zweithöchste. Bei den übrigen drei Kandidaten liegen die Abweichungen meist zwischen diesen extremen, sind jedoch stets wesentlich höher als jene des Wahlsiegers.

Auch in der nächsten Grafik ist dieser charakteristische Zusammenhang dargestellt. Wieder wurden die relativen Abweichungen berechnet, wobei die Daten der Nationalratswahl 1999 als Grundlage dienten.

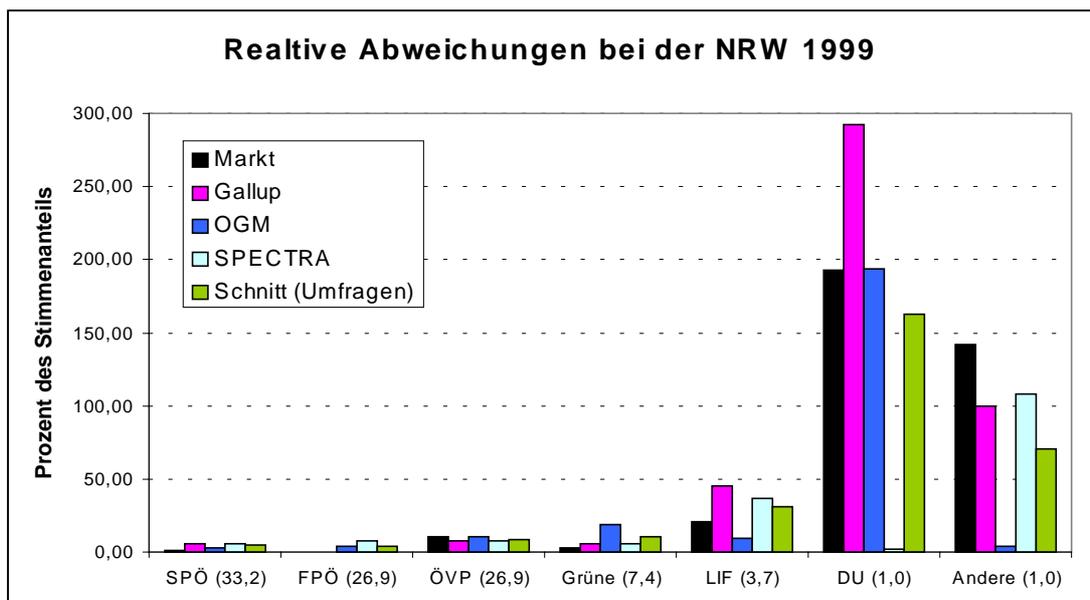


Abbildung 40: Relative Abweichungen bei der Nationalratswahl 1999; Quelle: APSM-homepage

Man erkennt, daß die relativen Schätzfehler mit sinkendem Stimmenanteil der Parteien enorm zunehmen.¹⁰¹ Bei den drei größeren Parteien beträgt der Fehler in der Regel rund 4 Prozent, bei den Grünen und dem Liberalen Forum 10 bis 30 Prozent, bei den kleinsten Parteien aber oft über 100 Prozent.¹⁰² Dies wird durch die vorhandenen Theorien in der Wahlforschung nicht erklärt und veranlaßt den Verfasser zu der Vermutung, daß bei beiden Instrumenten eine beträchtliche Portion „Raten“ in die Prognose einfließt.

Abweichungen der Umfrageinstitute und Börsen untereinander

Eines der überzeugendsten Argumente gegen die unbedingte Zuverlässigkeit von Umfragen stellen die teils enormen Abweichungen in den publizierten Daten der einzelnen Institute dar. Unter Einbeziehung der jeweiligen Konfidenzintervalle sind diese kaum erklärbar.

Ob dieses Phänomen auch beim Vergleich der Schlußkurse verschiedener Wahlbörsen zum selben Ergebnis auftreten würde, ist nicht verlässlich zu sagen, da dafür zu wenige Daten vorliegen. Die einzige Wahl, zu der dem Verfasser Daten von zwei Wahlbörsen zugänglich sind, ist die Nationalratswahl 1999. Die beiden Tageszeitungen „Die Presse“ und „Der Standard“ veranstalteten jeweils eine Börse deren Schlußkurse in der folgenden Tabellen dem Wahlergebnis gegenübergestellt werden.¹⁰³

¹⁰¹ der Korrelationskoeffizient zwischen Stimmenanteil und relativer Abweichung beträgt beim Markt $-0,67$, bei den Umfrage $-0,66$

¹⁰² dabei gelingt es sowohl dem Markt als auch einzelnen Umfragen bisweilen, auch bei kleinen Parteien sehr gute Schätzungen zu liefern (Markt bei den Grünen, Spectra bei „DU“, OGM bei „Anderen“), doch sollten diese im Lichte der oft enormen Abweichungen bei anderen Parteien nicht überbewertet werden

¹⁰³ dabei wurden aus Vergleichsgründen die Aktien „DU“, „KPÖ“ und „Andere“ der Presse-Wahlbörse unter der Rubrik „Andere“ zusammengefaßt, da es am Markt des Standard nur diese Aktie gab

	Wahlergebnis	„Presse“	„Standard“	Abweichung ¹⁰⁴
SPÖ	33,15	33,34	34,50	1,16
FPÖ	26,91	26,80	25,11	1,69
ÖVP	26,91	24,08	24,50	0,42
Grüne	7,40	7,19	7,70	0,51
LIF	3,65	4,40	5,20	0,80
Andere	1,98	5,33	4,64	0,69

Tabelle 11: Ergebnisse von Wahlbörsen im Vergleich; Quellen: „Die Presse“, „Der Standard“

Wir sehen, daß die Kurse durchwegs voneinander abweichen, wobei der Fehler im Vergleich zum Wahlergebnis interessanterweise fast immer in die selbe Richtung erfolgt (einzige Ausnahme sind die Grünen). Im Durchschnitt liegen die Werte um 0,88 Prozent auseinander, was verglichen zu üblichen Abweichungen von Umfragen eher gering ist.

Bei Umfragen wird wieder zwischen Beobachtungen in Europa und den USA unterscheiden, da die Bearbeitung der Daten in Europa die Vergleichbarkeit einschränkt. Zwar werden damit die Extreme verringert, was in der Regel zu geringeren Abweichungen führt. Allerdings trägt auch die größere Anzahl an Parteien zu (scheinbar) niedrigeren Schätzfehlern bei.

Österreich

Ein Vergleich von Umfragen im Zeitablauf ist nur dann wirklich seriös, wenn die Erhebungszeiträume fast ident sind. Dies trifft bei den Befragungen von Gallup und Spectra anlässlich der NRW 1999 in Österreich zu. sehen wir uns die Umfragewerte im Vergleich in der folgenden Tabelle an (Gallup vor dem Bindestrich, Spectra danach).

Man erkennt, daß die Differenzen zwischen den beiden Instituten im Zeitablauf annähernd konstant waren, wenn sich auch die Unterschiede bei den einzelnen Parteien leicht verschoben.

¹⁰⁴ Abweichung der Prognosen der beiden Wahlbörsen voneinander

Datum ¹⁰⁵	SPÖ	FPÖ	ÖVP	Grüne	LIF	DU	Andere	Differenz
2. Sept.	37-38	24-25	25-24	7-6	2-4	4-2	1-1	1,14
9. Sept.	35-37	26-25	25-24	7-6	2-5	4-2	1-1	1,43
16. Sept.	36-35	25-27	25-24	7-7	2-5	4-1	1-1	1,43
23. Sept.	35-36	28-27	24-24	7-6	2-5	4-1	0-1	1,43
29. Sept.	35-35	27-25	25-25	7-7	2-5	4-1	0-2	1,43
Differenz	1,00	1,40	0,60	0,60	2,80	2,60	0,60	1,37

Tabelle 12: Differenzen bei Umfragewerten im Zeitablauf; Quellen: Umfrageinstitute

Das Augenfälligste ist, daß die größten Abweichungen bei den kleinen Parteien „DU“ und „LIF“ auftraten – bei diesen tritt die „Expertenmeinung“ mehr in den Vordergrund, weil die Aussagekraft der Daten bei sehr kleinen Anteilen abnimmt. Dies demonstriert einmal mehr, daß selbst gutgemeinte Eingriffe der Meinungsforscher die Daten erheblich verzerren können.

Besonders interessant ist aber ein Vergleich der Schlußprognosen einzelner Institute vor einer Wahl. Sehen wir uns die entsprechenden Daten für die Nationalratswahlen 1995 und 1999 und die EU-Wahl 1996 an.

	SPÖ	FPÖ	ÖVP	Grüne	LIF	Andere	Differenz
NRW 95 ¹⁰⁶	32-35	25-24	30-27	7-6	6-6	0-2	1,67
EU 96 ¹⁰⁷	29-30	27-23	27-31	10-8	7-8	0-0	2,00
NRW 99 ¹⁰⁸	35-35	27-25	25-25	7-7	2-5	4-3	1,00

Tabelle 13: Differenzen bei Umfragewerten – Schlußprognose; Quellen: Umfrageinstitute

¹⁰⁵ Datum der Veröffentlichung, bei der letzten Umfrage war Gallup einen Tag später als Spectra

¹⁰⁶ Gallup vs. ISMA

¹⁰⁷ Gallup vs. market

¹⁰⁸ Gallup vs. Spectra

Die Abweichungen der Institute voneinander bewegen sich in Österreich meist zwischen ein und zwei Prozentpunkten, was etwa einer Standardabweichung entspricht. Dies ist im Einklang mit den theoretischen Grundlagen.

USA

Die bisher analysierten österreichischen Daten hatten den Nachteil, daß dabei nur wenige Umfragen miteinander verglichen werden konnten. Aussagekräftigere Ergebnisse aufgrund dichter Zeitreihen läßt die Analyse der Daten aus den USA erwarten. Da dem Verfasser verläßliche Daten nur vom Wahlkampf 1992 vorliegen, werden diese als Grundlage der folgenden Ausführungen verwendet. Die Abbildung zeigt die Differenzen der jeweiligen Umfragen von sechs verschiedenen Instituten von den beiden, die zeitlich jeweils am nächsten publiziert wurden.¹⁰⁹

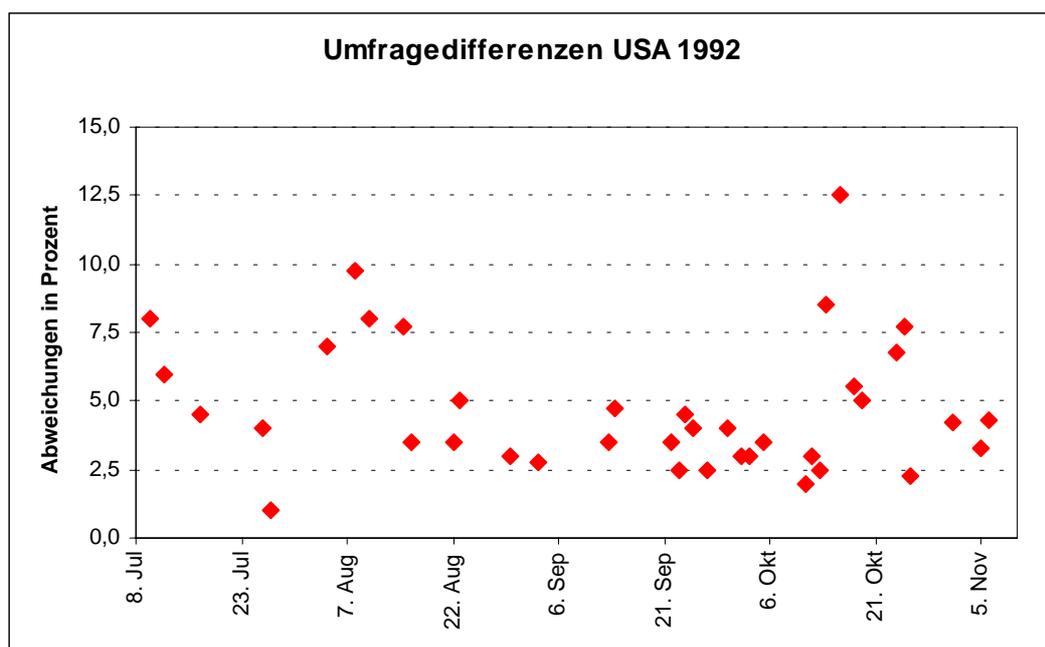


Abbildung 41: Umfragedifferenzen in den USA; Quelle: Forsythe et al. 1992, S. 1150

¹⁰⁹ wurden mehrere Umfragen am selben Tag veröffentlicht, so wurde deren Durchschnitt verwendet

Es ist zu erkennen, daß sich die Differenzen meist zwischen zweieinhalb und sechs Prozent bewegen, wobei kaum ein zeitlicher Trend auszumachen ist. Auch wenn die Zahlen aufgrund verschiedener Parteienanzahlen sowie unterschiedlicher Berechnungsmethoden nur sehr eingeschränkt mit denen aus Österreich zu vergleichen sind, sind doch zwei Dinge auffällig: erstens gibt es in beiden Ländern keinen deutlich erkennbaren zeitlichen Trend – Differenzen verlagern sich zwar unter den Parteien, bleiben in ihrem Ausmaß aber ziemlich konstant. Zweitens scheinen die Unterschiede in den USA tendenziell größer zu sein. In Österreich betragen sie ca. eine Standardabweichung, in den Vereinigten Staaten aber zwischen zwei und drei Standardabweichungen.¹¹⁰

Der Grund für die größeren Abweichungen in den USA ist dabei wohl in den, unter den Instituten variierenden Erhebungsmethoden aber auch in der geringeren Zahl der Parteien zu suchen. Betrachtet man nämlich die absolute Gesamtabweichung aller Parteien zusammen, so beträgt diese in Österreich zwischen 6 und 12 Prozent. In den USA liegt sie meist zwischen 5 und 12 Prozent, so daß hier keine Unterschiede auszumachen sind. Einzig „Ausreißer“ scheinen durch die Bearbeitung der Daten reduziert zu werden, die Gesamtprognose verbessert sich hingegen kaum.¹¹¹

¹¹⁰ solche Abweichungen sollten nach der Wahrscheinlichkeitstheorie kaum auftreten

¹¹¹ so weicht eine Umfrage in den USA um durchschnittlich 12,5 % - also gesamt um 25 % ab, was in Europa praktisch nie vorkommt

9 Designvorschläge und Ausblick für Wahlbörsen aus politikwissenschaftlicher Sicht

Wir haben gesehen, daß Wahlbörsen von Struktur und Konzept her einige entscheidende Unterschiede zu Umfragen aufweisen, wodurch sich verschiedene Einsatzmöglichkeiten ergeben. Auf den folgenden Seiten werden einige wichtige Faktoren für den Erfolg einer Wahlbörse behandelt, bevor ein Ausblick auf die zukünftigen Einsatzmöglichkeiten dieses Instrumentes gegeben wird.

9.1 Designvorschläge zur Sicherstellung einer guten Prognose

Für den Prognoseerfolg eines Marktes dürfte neben dem Engagement seiner Händler v.a. sein Design ausschlaggebend sein. Dabei sind vom Veranstalter bereits lange vor Handelsbeginn viele Entscheidungen zu treffen – von der Höhe des Einsatzes über die Marktform bis zur Dauer des Handels, wobei jeder dieser Faktoren einen erheblichen Einfluß auf die Effizienz des Marktes haben kann.

9.1.1 Marktdesign im engeren Sinne

Als erstes – und vom Veranstalter am besten zu steuern – sind Entscheidungen über die konkrete Gestaltung des Marktes in Bezug auf Einsatz, Dauer und Marktform zu treffen.

Geldeinsatz

Der Einfluß des Einsatzes auf den Erfolg einer Wahlbörse läßt sich nicht a priori ermitteln. So kann man argumentieren, daß höhere Einsätze zu einem vorsichtigeren und überlegteren Handel und damit auch zu einer rationaleren Preisbildung führen würden. Andererseits bedeuten hohe Einsätze in der Regel auch niedrigere Händlerzahlen, womit ein Markt anfälliger für Manipulationen wird. Umgekehrt sollten die hohen Händlerzahlen bei geringen Einsätzen bewußte Verzerrungen verhindern, während die effiziente Preisbildung durch die Motivation der Teilnehmer, die Spaß am Spiel haben und dabei erfolgreiche

Händler sein wollen, sichergestellt wird. Die empirischen Daten lassen keinen Zusammenhang zwischen Geldeinsatz und Prognosegenauigkeit erkennen. Lediglich jene Märkte, an denen einzelne Teilnehmer durch unterschiedliche Einsätze oder geringe Händlerzahlen eine große Marktmacht hatten, erzielten tendenziell schlechtere Ergebnisse als andere Börsen.

Die einzige Empfehlung lautet also, daß der Einsatz eher gering und unbedingt für alle Händler gleich sein sollte, um eine Konzentration von Marktmacht zu verhindern, da diese zu Manipulationen mißbraucht werden könnte.

Händleranzahl

Die Händlerzahlen hängen von einigen Faktoren ab. Neben der Höhe des Einsatzes sind insbesondere die Höhe des *event levels* sowie die Einbindung von Medien wichtige Faktoren. Während Märkte, die anläßlich von bundesweiten Wahlen durchgeführt wurden, meist sehr hohe Händlerzahlen aufwiesen, sind für Wahlen auf niedrigeren Ebenen oft nicht genug Teilnehmer zu finden.

Zur Akquirierung von Händlern hat sich die Einbeziehung von (Print)Medien bisher als sehr hilfreich erwiesen, jedoch waren solche Börsen aufgrund ihrer Publizität überdurchschnittlich oft Manipulationsversuchen ausgesetzt, so daß die höhere Händlerzahl unter Umständen mit einem Verlust an Prognosegenauigkeit „bezahlt“ werden muß.

Nach den empirischen Daten kann bei Märkten mit Händlerzahlen ab ca. 100 mit einer guten Prognose gerechnet werden. Dem Verfasser erscheint eine Marktgröße zwischen 300 und 600 Händlern optimal, um einen flüssigen und interessanten Handel zu ermöglichen. Bei geringeren Zahlen ist die Liquidität möglicherweise zu niedrig, was zu extremen Kurssprüngen und großen *spreads* führt. Bei extrem hohen Händlerzahlen wird der Markt u.U. durch eine Vielzahl von limitierten Orders fast gelähmt, so daß kaum noch Kursbewegungen möglich sind.

Um eine Übersättigung zu vermeiden und angemessene Händlerzahlen zu gewährleisten, sollten Märkte nur für „wichtige“ Wahlen durchgeführt werden. Damit sollten auch ohne die direkte Einbindung von Medien angemessene Teilnehmerzahlen erreichbar sein.

Marktform – Anzahl der Aktien

In den USA haben sich WTA-Märkte im Vergleich zu Stimmenanteilmärkten als wesentlich attraktiver erwiesen, da sie sich besser für das dort vorherrschende Zweiparteiensystem eignen. In Europa gibt es in der Regel mehr Parteien, so daß hier VS-Märkte die sinnvollere Form darstellen.

Ein Vergleich der Prognosegenauigkeit der beiden Marktformen ist nur bedingt möglich, weil das Abweichungsmaß MAE für WTA-Märkte nicht sinnvoll ist und die Aufgabenstellung an einem WTA-Markt (Prognose des Siegers) meist einfacher ist als die an einem VS-Markt (Abschätzung aller Stimmenanteile).

Zu klären ist auch die Anzahl der Aktien. Manchmal ist es nicht sinnvoll, jede Partei und Gruppierung durch eine Aktie abzubilden. So sollten nach Meinung des Verfassers nur die größeren Parteien separat ausgewiesen, kleinere indes in einer Sammelaktie „Andere“ berücksichtigt werden. Es hat sich nämlich gezeigt, daß die Prognosegenauigkeit insbesondere bei den die Aktien von Kleinstparteien (KPÖ, DU, CSA, etc.) bzw. bei der Aktie „Andere“ schlecht sind, denn aufgrund von Spekulationen sind diese regelmäßig massiv überbewertet.

Da alle Wahlmöglichkeiten am Markt abgebildet sein müssen, können solche Gruppierungen nicht ignoriert werden. Es erscheint hier aber sinnvoll, zumindest die Kleinstparteien in einer Sammelaktie zusammenzufassen.

Laufzeit

Aus den empirischen Daten läßt sich ablesen, daß die Kurse erst in den letzten Wochen vor einer Wahl eine sinnvolle Prognose des Wahlergebnisses darstellen, weil vorher der Ausgang noch zu unsicher ist oder aber Spekulationen bei Kleinparteien die Kurse verzerren. Ebenso bleiben die Kurse an WTA-Märkten meist sehr lange nahe der Ausgangsposition. Es konnte auch kein signifikanter Zusammenhang zwischen der Dauer eines Marktes und seiner Prognosegenauigkeit festgestellt werden, jedoch gibt es bei zu langen Laufzeiten immer wieder wochenlange Phasen, in denen der Handel praktisch zum Erliegen kommt. Dem Verfasser scheinen daher relativ kurze Laufzeiten von ca. zwei Monate optimal, da in dieser Zeitspanne eine seriöse Prognose möglich ist und gleichzeitig das Interesse der Händler aufrechterhalten bleibt.

9.1.2 Öffentlichkeit und Teilnehmerfeld

Teilnehmer

Eine der Stärken von Wahlbörsen liegt darin, daß ihr Händlerfeld nicht repräsentativ für die Wähler sein muß, um ein gutes Prognoseergebnis zu erzielen. Bisher dominierten bei den Teilnehmern stets Männer, Studenten und eher junge Bevölkerungsschichten, was jedoch der Qualität der Vorhersage keinen Abbruch tat. Dennoch sollten die Veranstalter darauf bedacht nehmen, daß die Händler bei einer Wahlbörse nicht aus einer zu einheitlichen Gruppe kommen, da ansonsten nur die Meinung in dieser Gruppe in den Kursen widerspiegelt sein könnte.¹¹² Abhilfe schafft hier schon die Einbeziehung einiger Händler aus irgendeiner anderen Gruppe bzw. Schicht.

Als ein wichtiger Erfolgsfaktor von Wahlbörsen hat sich das Vorhandensein von *marginal traders* herauskristallisiert. Durch die hohe Zahl an Transaktionen wirken diese aktiv auf die Preisbildung ein und erhöhen dabei die Effizienz eines Marktes. Natürlich kann ein Veranstalter nicht im vorhinein feststellen welche Händler besonders aktiv sein werden bzw. ob sich überhaupt *marginal traders* an einer Börse beteiligen werden, doch könnte hier die Schaffung von Händlerpools durch Institutionen wie IEM (USA) und Fundaq (Österreich) ein entscheidender Schritt sein. Die aktivsten dieser Händler könnten ermutigt werden, sich an neuen Märkten zu beteiligen, womit diese schneller zu aussagekräftigen Preisen gelangen könnten.

Öffentlichkeit

Der anfängliche Prognoseerfolg von Wahlbörsen hat sich mittlerweile als ein großes Handicap erwiesen, da die resultierende Berichterstattung in den Medien die Märkte für Manipulationen interessant gemacht haben. Auch haben die teilweise extrem hohen Händlerzahlen, die durch die Einbindung von Medien erreicht wurden, nicht immer zur Verbesserung der Prognosegenauigkeit von

¹¹² dies geschah etwa bei der Wahlbörse zur EU-Volksabstimmung, wo nur Studenten handelten, und das Ergebnis dann auch das Abstimmungsergebnis der Studenten widerspiegelte

PSMs geführt. Waren diese Medien anfangs hilfreich, um Teilnehmer zu gewinnen, so könnte die Schaffung von Händlerpools nun diese Funktion übernehmen. Die Veranstalter können nun direkt auf tausende potentielle Teilnehmer zurückgreifen und damit Börsen durchführen, die nicht in den Medien publiziert werden, wodurch sie für Manipulationen weniger interessant sind.

Politische Aktienmärkte konnten dann die besten Prognoseergebnisse erzielen, wenn der Handel möglichst unbeachtet von den Medien – und damit uninteressant für Spekulationen – stattfinden konnte, so daß dies die Maxime zukünftiger Märkte sein sollte. Die Zukunft liegt damit vielleicht bei eher kleiner Börsen, die vor der Wahl bewußt auf die Erwähnung in Medien verzichten.

9.1.3 Eingriffe durch den Veranstalter

In unserer Analyse haben wir festgestellt, daß es an Wahlbörsen zu regelmäßigen Verzerrungen kommt – so sind die Aktien kleinerer Parteien meist überbewertet und in der Folge jene größerer Parteien unterbewertet. Diese Fehler, die meist einen Großteil der Abweichung einer Börse vom Wahlergebnis ausmachen, sind bekannt und abschätzbar und könnten daher von den Veranstaltern in einer Schlußprognose korrigiert werden. Selbst wenn es auf diese Weise gelänge, meist bessere Vorhersagen zu erzielen, ist diese Vorgehensweise strikt abzulehnen!

Jede Bearbeitung der Kurse durch den Veranstalter brächte die selbe Subjektivität und die selben Möglichkeiten der Manipulation mit sich, die bereits bei der Erstellung von Prognosen aus Umfragedaten heftig kritisiert wurden. So wie Umfrageforscher durch den Versuch einer Prognose den Anspruch auf Repräsentativität verlieren, würde durch eine Korrektur der Kursdaten von Wahlbörsen deren Grundprinzip der Informationsaggregation durch den Markt aufgegeben.

9.2 Ausblick auf den Einsatz von Wahlbörsen

Wahlbörsen und Umfragen sind zwei unterschiedliche Instrumente um ein ähnliches Problem zu lösen. Dabei ist die Fragestellung unterschiedlich, so daß man durch PSMs zumindest eine „zweite Meinung“ erhält.

Umfragen erheben die aktuelle politische Stimmung sehr gut. Jedoch sind sie sehr stark auf den Erhebungszeitpunkt bezogen und erlauben damit keine Prognose im eigentlichen Sinn. Jeglicher Versuch einer Vorhersage macht eine repräsentative Erhebung zu einer unrepräsentativen Expertenschätzung. Das zentrale Problem der Umfrageforschung bleibt, daß viele Menschen erst sehr kurz vor der Wahl ihre Wahlentscheidung treffen, bzw. entscheiden, ob sie überhaupt zur Wahl gehen. Mit demselben Problem sind jedoch auch Wahlbörsen konfrontiert, denn noch weniger als die Experten der Institute können die einzelnen Händler den Wahlausgang mit Sicherheit vorhersagen.

Dennoch könnten Wahlbörsen einen wertvollen Dienst für die Wahlforschung leisten. Aufgrund seines Aufbaus ist ein politischer Aktienmarkt ein Instrument, mit dem nicht die aktuelle politischen Stimmung, sondern die langfristigen Erwartungen der Händler über das endgültigen Wahlergebnisses erfaßt und abgebildet werden. Damit stellen Wahlbörsen eine erfolgversprechende Möglichkeit dar, Wahlprognosen zu verbessern.

Wahlbörsen können auch dazu beitragen, den Blick für die grundlegenden Wahlaussichten zu schärfen und Abstand von nur kurzfristigen, von den Medien aufgebauchten, Ereignissen zu gewinnen (vgl. Frey/Götte/Kucher, 1996).

Insbesondere die Grundform der Wahlbörse – ein abgeschlossenes Experiment, das für Manipulationen uninteressant ist, weil es zu keiner regelmäßigen Publikation der Kurse in Massenmedien kommt – scheint relativ zuverlässig eine gute Vorhersage des Wahlergebnisses zu erlauben.

Doch auch ein gänzlich anderer Einsatz ist vorstellbar: Umfragen messen gerade zur Mitte einer Legislaturperiode oft massive Stimmenverluste für Regierungsparteien, die zwar real sein können, sich jedoch kaum in Wahlergebnissen folgender Jahre widerspiegeln. Hier könnten Wahlbörsen mit sehr langen Laufzeiten für mehr Kontinuität in der Messung politischer Stimmungen sorgen. Allerdings dürfte es schwierig sein, genügend aktive Händler zu finden, damit

über Jahre hinweg ein effizienter Handel gewährleistet werden kann.¹¹³ Somit ist ein solcher Markt nur sinnvoll, wenn sichergestellt ist, daß es genügend interessierte Händler gibt.

Eventuell könnten politische Aktienmärkte auch zu neuen Überlegungen in der Umfrageforschung Anlaß geben. So wäre es beispielsweise interessant, in einer Befragung nicht mehr die Wahlabsicht der Befragten, sondern deren Meinung über den Wahlausgang zu erheben. Man könnte also die Fragestellung von PSMs auf Umfragen übertragen, und prüfen, wie gut eine solche Prognose wäre.

Es war Ziel dieses Abschnitts, Stärken und Schwächen der beiden Instrumente aufzuzeigen, damit in Zukunft jeweils die Stärken genutzt werden können. Dabei soll der Siegeszug des einen Instruments nicht als Versagen oder gar Untergang des anderen gesehen werden. Beide Instrumente haben in der Wahlforschung ihre Berechtigung, denn je nach Problemstellung mag das eine oder andere ein präziseres oder aussagekräftigeres Ergebnis liefern.

Mit Umfragen können bei gleichbleibender Erhebungsmethodik Trends und Entwicklungen sehr gut aufgezeigt werden, es ergibt sich ein dynamisches Bild.

Die Kurse an Wahlbörsen sind in der Regel stabiler als Umfragewerte, wobei die Wahlchancen einzelner Parteien relativ unabhängig von kurzfristigen Schwankungen abgebildet werden. Andererseits können die Teilnehmer an Wahlbörsen dank des ununterbrochenen Handels jederzeit auf neue Entwicklungen reagieren, so daß stets die Möglichkeit zu Preisanpassungen besteht. Insgesamt ergibt sich ein relativ stabiles Bild der Erwartungen über den Wahlausgang

Die wohl größte Stärke von Wahlbörsen stellt ihr „demokratisches“ Grundprinzip dar. Bei gleichem Einsatz hat jeder Händler am Markt die selbe Macht, und alle können gleichermaßen auf die Preisbildung Einfluß nehmen. Über seine Aktivitäten am Markt kann jeder Teilnehmer darüber hinaus die Intensität seiner Präferenzen ausdrücken, während dies bei Umfragen oft nicht möglich ist.

¹¹³ an den IEM wurde vor der US-Präsidentenwahl 1996 ein Markt mit einer Laufzeit von zwei Jahren durchgeführt, doch gab es dabei immer wieder Phasen mit sehr geringer Handelstätigkeit

Der Verfasser kommt zum Schluß, daß trotz aller oft unvermeidlichen Mängel in der praktischen Anwendung bei Umfragen und PSMs die jeweiligen theoretischen Fundamente Gültigkeit haben: mittels Befragungen die auf Zufallsstichproben beruhen läßt sich die Intention der Wähler gut erfassen – dies wird am besten durch *exit polls* belegt. Allerdings unterliegen die Stimmungen erheblichen Schwankungen, weswegen eine Erhebung nicht als Prognose gesehen werden sollte. Auch Bearbeitungen durch Experten können daran nichts ändern, sondern führen zum Verlust des Anspruchs auf Repräsentativität.

Auch Wahlbörsen dürften leisten, was sie versprechen: die Meinungen und Informationen der Händler werden durch den Markt erfolgreich aggregiert und liefern oft schon frühzeitig eine gute und zuverlässige Vorhersage des Wahlergebnisses. Die Prognosekraft wird allerdings dann eingeschränkt, wenn die Kurse am Markt durch Manipulationen verzerrt werden, wie das insbesondere dann vorkommt, wenn eine Börse zu bekannt und damit für Eingriffe der Parteien interessant ist. Alles in allem sind Wahlbörsen ein sinnvolles neues Instrument, dessen sich die Wahlforschung in Zukunft freudig bedienen sollte.