

# Missgeschick oder Trend?

## Zur Prognosetauglichkeit von Wahlumfragen

von Dieter Roth und Andreas M. Wüst<sup>1</sup>

### I. Einführung

Wie vertrauenswürdig Wahlumfragen sind, wird oft am vermeintlichen Renommee eines Umfrageinstituts festgemacht. Dabei gibt es für die Qualität kaum eines anderen Instruments der empirischen Sozialforschung einen so eindeutigen Test wie für Wahlumfragen: das Ergebnis der Wahl. Wer auf der Grundlage von Umfragedaten den Wahlausgang richtig vorhersagt, hat nicht nur die zugrundeliegenden Daten adäquat erhoben, sondern beweist auch methodisch-technische Kompetenz in deren Analyse und Aufbereitung. Das ist plausibel und unter den Beobachtern publizierter Wahlumfragen (Wissenschaftlern, Journalisten und Politikern) vorherrschende Meinung. Deshalb ist die in den vergangenen Monaten geübte Kritik an den Demoskopien in Großbritannien nach ihrem gemeinhin als „Debakel“ charakterisierten Misserfolg bei der Unterhauswahl am 07. Mai 2015 für Wahl- und Umfrageforscher eine Frage, die einer genaueren Analyse bedarf und zumindest nach Antwortversuchen verlangt.

In der Tat haben bekannte und damit über weite Strecken auch erfolgreiche *Pollster*, wie *George Gallup* in den USA oder *Elisabeth Noelle-Neumann* in der frühen Bundesrepublik, die Qualität ihrer eigenen Umfragen immer wieder durch die Treffsicherheit ihrer Wahlprognosen beweisen wollen.<sup>2</sup> Selten aber lagen

<sup>1</sup> Die Autoren danken *Uwe Wagschal* und seinem Mitarbeitern für die Unterstützung bei der Datenaufbereitung.

<sup>2</sup> Eine Stichprobe *George Gallups* von weniger als 2000 Befragten übertraf in der Genauigkeit das Ergebnis einer Befragung von 2,4 Millionen Leser des „Literary Digest“: Dies markierte den ökonomischen Durchbruch für *Gallup* in den USA im Jahr 1936. Vgl. hierzu *Converse, J. M.*: *Survey Research in the United States. Roots and Emergence 1890-1960*, Berkeley, 1987, hier 121. *Elisabeth Noelle-Neumann*, die das Institut für Demoskopie Allensbach als „der Welt bestes Prognose-Institut“ bezeichnet hat (z.B. in *Süddeutsche Zeitung* vom 30.9.1998, 3) und auch behauptete, „die Allensbacher Wahlprognosen [gelten] als die genauesten der Welt“ (*Frankfurter Allgemeine Zeitung* vom 19.7.2000, 5), hat diese vermeintliche Genauigkeit immer wieder als ein zentrales Qualitätsmerkmal hervorgehoben. Trotz etlicher Fehlprognosen blieb das Renommee des Instituts lange Zeit unangetastet, bis eine zweifelhafte Veröffentlichung für die Republikaner vor der Landtagswahl in Baden-Württemberg 1996 und

Demoskopen so sehr daneben wie in Großbritannien vor der Wahl 2015; hier war der Wahlkampf in besonderer Weise durch Umfragen geprägt.

Wie so häufig vor Wahlen, ging es in den Medien dabei mehr um das *horse race*, die Zahlen und vermeintlich knappen Mehrheiten, als um einen Wettbewerb der Kandidaten und Parteien um Problemlösungen oder gar Wahlprogramme. Der Wahlkampf wurde in einer von den Umfragen geschaffenen Scheinwelt geführt.<sup>3</sup> Die Debatte sei verzerrt gewesen, wie verschiedene Kritiker schreiben und auch die BBC (*Andrew Marr*) konzidiert. Diskussionen über mögliche Koalitionen beherrschten das Feld. Noch am Wahltag verkündete der *Guardian* ein Patt: „Es könnte knapper nicht sein“.<sup>4</sup> Nach Auszählung der Stimmen lagen die Konservativen (36,9%) dann freilich 6,4 Prozentpunkte vor *Labour* (30,5%). Mindestens zehn Institute publizierten Umfragen, in denen *Labour* und die *Tories* höchstens einen Prozentpunkt auseinanderlagen.<sup>5</sup>

Angesichts dieser Diskrepanz spricht viel für einen oder auch mehrere systematische Fehler bei der Datenerhebung und -analyse, die beunruhigen müssten. Denn wenn dem so wäre, stellt sich die Frage, ob es sich dabei um ein „britisches Problem“ handelt oder aber um eine allgemein abnehmende Prognosefähigkeit, die auch andere Länder<sup>6</sup>, und damit auch Deutschland, beträfe.

## II. Der britische Fall

Blickt man auf das Wahlergebnis im Vereinigten Königreich, dann ist die Gewinn- und Verlustbilanz beider großer Parteien gegenüber der vorangegangenen Wahl 2010 gering. Die Konservativen gewannen 0,8% hinzu, *Labour* verbesserte sich um 1,5%. Größere und auch folgenswerere Veränderungen gab es bei den Liberaldemokraten (minus 12,6%) und bei der *United Kingdom Independence Party* (UKIP, plus 9,5%) – weniger im bevölkerungsreichsten Landesteil England, als vielmehr in Schottland, wo die *Scottish National Party* (SNP) 30% hinzugewann und *Labour* 17,7% verlor. Die Zugewinne von UKIP erwiesen sich

die nachfolgende Diskussion Zweifel an der Prognosepraxis aufkommen ließen (pointiert: „Kaffeessatz vom Bodensee“, in: *Süddeutsche Zeitung* vom 28.3.1996, 2).

<sup>3</sup> Die ZEIT (online), <http://www.zeit.de/ausland/2015-05/grossbritannien-wahl>.

<sup>4</sup> The Guardian Poll Projection (online), siehe <http://www.theguardian.com/politics/2015/may/07>.

<sup>5</sup> Der Standard (online) vom 18.5.2015, <http://derstandard.at/2000015996943/Wie-man-falsche-Prognosen-schönreden-kann>, vgl. auch Tabelle 1.

<sup>6</sup> Zukin, C.: What's the Matter With Polling?, in: The New York Times (online) vom 20.06.2015, <http://nyti.ms/1GyPSmU>. Der Autor verweist auf weitere „spectacular disasters“, wie falsche „Prognosen“ bei US-amerikanischen midterm elections, die starke Unterschätzung von Benjamin Netanyahu Likud in den Umfragen vor der Wahl in Israel 2015 und eben auch Großbritannien. Er folgert daraus: „How much can we trust the polls as we head toward the 2016 elections?“

in England als leicht überdurchschnittlich (plus 10,7%), ebenso die Verluste der *Liberal Democrats* (minus 16%).

Aufgrund der relativen Mehrheitswahl in 650 Einzelwahlkreisen wirkten sich die Veränderungen auf die Sitzverteilung viel stärker aus, als dies bei einem Verhältniswahlrecht der Fall wäre. In England konnten die *Tories* etwa nur 1,4% der Stimmenanteile, aber 21 Sitze dazu gewinnen, während *Labour* zwar einen Stimmenzuwachs von 3,6% verzeichnete, aber lediglich 15 zusätzliche Sitze gewann. UKIP verbuchte zwar beachtliche 10,7% für sich, aber nur ein einziges Mandat (Wahlkreis). In Schottland, als weiteres Beispiel, gewann die SNP die Hälfte aller Stimmen (50%) und 56 von 59 Wahlkreisen, ein Plus von 50 Sitzen. (Ein solches Ergebnis wurde Anfang März bereits einmal von der BBC prognostiziert, was zu intensiven Diskussionen über eine mögliche Koalition von *Labour* und SNP führte.)<sup>7</sup> *Labour*, als die in Schottland einst dominierende Partei, verlor dort lediglich ein Drittel der Stimmen, aber alle Wahlkreise bis auf einen.

Nun ist eine treffsichere Sitzprognose in einem Mehrheitswahlsystem stets eine besondere Herausforderung, weil Vorwahlumfragen für sämtliche Wahlkreise nicht vorliegen. Man konzentriert sich, zusätzlich zur Erhebung des landesweiten Trends, in der Regel auf Wahlkreise, die schon in der Vergangenheit hart umkämpft waren. Wenn es Meinungsforschern jedoch, wie vor dieser Wahl, nicht einmal gelingt, die landesweite Unterstützung (Stimmung) für die verschiedenen Parteien richtig zu „messen“, dann liegt es nahe, dass wir es hier mit einem oder mehreren methodischen Problemen zu tun haben.

Bei all dem ist zu konzedieren, dass sich die Rahmenbedingungen für Wahlumfragen durch Veränderungen der Beziehungen zwischen der Bevölkerung und den Parteien und in Folge dessen auch das Parteiensystem gegenüber der Zeit zweier dominanter Volksparteien stark gewandelt haben. Bis vor wenigen Jahren konnte man davon ausgehen, dass ein erkennbarer *swing* zwischen diesen Parteien auf die Verhältnisse in engen Wahlkreisen übertragbar war, wofür Umfragen ein zentrales, aber nicht das einzige Instrument darstellten. Jetzt steht man vor der Aufgabe, auch die Veränderungen der Stärke etlicher anderer Parteien, deren Summe und Relevanz für den Mandatsträger und die zweitstärkste Partei im jeweiligen Wahlkreis kalkulieren zu müssen. Hinzu kommt eine gewachsene Unsicherheit über die Höhe der Wahlbeteiligung, die wiederum ein Ergebnis des Wandels der Beziehungen und insbesondere der Bindungen zwischen Bürgern und Parteien ist. Schließlich ist auch eine ungewöhnlich hohe politisch-

<sup>7</sup> BBC News, 7. März 2015, <http://www.bbc.com/news/uk.politics-31780935>.

konjunkturelle Fluktuation durch Unabhängigkeitsbestrebungen oder Diskussionen über die Folgen von Mitgliedschaften (EU) oder Reaktionen auf internationale Sicherheitsprobleme (Kriege, Flüchtlingsströme etc.) feststellbar, die mittels Umfragen zwar gut zu erfassen sind, aber erst qualifiziert und mit Blick auf die daraus resultierende Effekte quantifiziert werden müssen. All dies trägt zur Erklärung möglicher Probleme von Wahlumfragen bei, doch kann es sicher nicht alle Unschärfen erklären.

Als ein weiterer Grund für die jetzige Fehlprognose wird, wie bereits früher (etwa nach der Unterhauswahl 1992), eine allgemeine Zurückhaltung konservativer Befragter bei Wahlumfragen angeführt („umfragescheue *Tories*“). Diese zeige sich, ganz in der Argumentation *Noelle-Neumanns* zur „Schweigespirale“<sup>8</sup> besonders dann, wenn in der öffentlichen Debatte eher linke Medien die Diskussion bestimmten. Eine überzeugende und vor allem hinreichende Erklärung bildet dies allerdings nicht.

Ex post melden sich Demoskopen zu Wort, die das Ergebnis vorausgesehen hätten, wie *James Morris* von GQR (*Greenberg, Quinlan, Rosner Research*), einem amerikanischen Institut. Er unterstellt den britischen Kollegen falsche Fragestellungen.<sup>9</sup> Ein früherer Berater *Barack Obamas*, *Jim Messina*, der in Großbritannien die *Tories* beriet, habe angeblich 319 Mandate für die Konservativen geschätzt und lag damit nur 12 Sitze vom tatsächlichen Ergebnis entfernt (allerdings unter der absoluten Mehrheit). *Peter Kellner* von *Yougov*, die vor allem internetbasierte Umfragen durchführen, weist als Reaktion auf das vermeintliche Prognosedesaster auf eigene, korrekte Prognosen bei der Londoner Bürgermeisterwahl 2012, bei der Europawahl 2014 und dem Referendum in Schottland 2014 hin. Dass all diese Wahlen nicht nach dem relativen Mehrheitswahlrecht durchgeführt wurden, erwähnt er nicht. Er verweist auf Ursachenanalysen, die ausstehen und die man abwarten müsse. „Barocke Erklärungen“ nennt das der Kolumnist der *Times*, *Philip Collins*.

Was sich zum Zeitpunkt der Abfassung dieses Beitrags sagen lässt, ist, dass die Kritiken an den „Prognosen“ vor der Wahl und am Wahltag selbst nur zum Teil zutreffend sind und das Instrument als solches noch nicht als fehlerverursachend identifiziert wurde. Im Gegensatz zu den Umfragen vor der Wahl hat der *exit poll* der britischen *Election Study Group* unter *John Curtice* (*University of Strathclyde*),

<sup>8</sup> *Noelle-Neumann, E.*: Die Schweigespirale. Öffentliche Meinung – unsere soziale Haut, München, 1980.

<sup>9</sup> Hier und nachfolgend: Der Standard (online), Fn. 5.

der in der BBC, ITV und Sky News in der Wahl Nacht (10 p.m.) vor Auszählung der Ergebnisse veröffentlicht wurde, mit „bemerkenswerter Genauigkeit“ (*Benjamin Lauderdale, London School of Economics*) das Endergebnis getroffen. Dies, obwohl die Anforderungen an die Wissenschaftler stark gestiegen sind und den Beteiligten auch sehr deutlich waren, wie aus einem Interview mit *John Curtice* vor der Wahl hervorgeht. Er weist auf Probleme und mögliche Fehlerquellen hin und beendet das Interview mit dem Satz: „Prepare for disappointment, it's almost guaranteed“ (Interview von *Robert Hutton*, durchgeführt am Tag vor der Wahl).

Man muss zumindest in diesem Fall also zwischen „Prognosen“, die auf der Basis von Befragungen von Wählern und Wählerinnen am Wahltag bei Verlassen des Wahllokals beruhen (*exit polls*), und den Befragungen von Wahlberechtigten vor der Wahl, die für ‚vermeintliche‘ Prognosen benutzt werden, unterscheiden. Letztere sind primär Momentaufnahmen, die die jeweilige politische Stimmung zum Zeitpunkt der Datenerhebung abbilden, sofern die Daten methodisch einwandfrei erhoben wurden. Umfrageforscher haben auch auf diesen Umstand hinzuweisen, doch wird allzu häufig *nolens volens* suggeriert, dass das Instrument der Befragung das Ergebnis der anstehenden Wahl erfassen könnte. Etliche Journalisten haben die Einschränkungen der Ergebnisse von Vorwahlumfragen verinnerlicht und wiesen entsprechend darauf hin.<sup>10</sup>

Blickt man auf die Veröffentlichungen vor der Wahl in Großbritannien, so deutete die weit überwiegende Zahl veröffentlichter Umfragen in den letzten zehn Tagen vor der Wahl auf ein Kopf-an-Kopf Rennen zwischen den Konservativen und *Labour* hin. Es gab unter 42 veröffentlichten Umfragen, im Zeitraum von zehn Tagen vor der Wahl, vier, die einen Vorsprung der Tories vor Labour von 6 bzw. 5 Prozentpunkten vermeldeten, zwei (*Ashcroft* bzw. *Ipsos-Mori*) zu einem relativ frühen Zeitpunkt, zwei am Tag vor der Wahl (*Survation* und *SurveyMonkey*).

Beachtet man die Fehlerbereiche, so liegen die Veröffentlichungen für die einzelnen Parteien im Vergleich zum späteren Wahlergebnis zum Teil außerhalb der statistischen Fehlermarge,<sup>11</sup> die mittlere Abweichung für alle Parteien liegt zwischen 1,2% und 2,8%. Wegen der relativen Einheitlichkeit der Zahlen war der sich in der Öffentlichkeit manifestierende Eindruck der einer sehr engen Wahl

<sup>10</sup> Beispiele: *Roll, E.*, in: Süddeutsche Zeitung vom 30.9.1998, hier 3; *Wirtz, C.*, in: Süddeutsche Zeitung vom 23.8.2002, hier 2.

<sup>11</sup> Die Fehlermarge bzw. das Fehlerintervall liegt bei einem Sicherheitsniveau von 95% und einer Stichprobengröße von 1000 bis 2000 Befragten, bei Parteien in der Stärke von 30-40%, im Bereich von +/- 3,1%, bei kleineren Parteien der Stärke 8-10% bei ca. 1,7-1,8%. Zur Berechnung siehe *Roth, D.*: Empirische Wahlforschung, Wiesbaden, 2008. Grundlage ist eine Zufallsauswahl der Befragten.

**Tabelle 1:** Veröffentlichungen von Umfrageinstituten im Vorfeld der Wahl zum Unterhaus im Vereinigten Königreich 2015

Umfragedatum	Institut / Auftraggeber	Con	Lab	LD	UKIP	Green	Other	Mittl. Abw.	Con-Lab
7 May	Wahlergebnis (ohne Nordirland)	37,8	31,2	8,1	12,9	3,8	6,3	0,0	6,6
5-7 May	Populus	34,0	34,0	9,0	13,0	5,0	6,0	1,8	=
6 May	Survation	37,0	31,0	10,0	11,0	5,0	6,0	1,2	6,0
30 Apr-6 May	SurveyMonkey	34,0	28,0	7,0	13,0	8,0	9,0	2,5	6,0
5-6 May	Lord Ashcroft	33,0	33,0	10,0	11,0	6,0	8,0	2,5	=
5-6 May	Ipsos MORI/Evening Standard	36,0	35,0	8,0	11,0	5,0	5,0	1,8	1,0
4-6 May	YouGov/The Sun	34,0	34,0	10,0	12,0	4,0	6,0	1,9	=
5-6 May	ComRes/Daily Mail, ITV News	35,0	34,0	9,0	12,0	4,0	6,0	1,5	1,0
4-6 May	Survation/Daily Mirror	31,0	31,0	10,0	16,0	5,0	7,0	2,6	=
3-6 May	ICM/The Guardian	34,0	35,0	9,0	11,0	4,0	7,0	2,1	-1,0
3-6 May	ICM/The Guardian	35,0	35,0	9,0	11,0	3,0	7,0	2,0	=
1-6 May	Panelbase	31,0	33,0	8,0	16,0	5,0	7,0	2,6	-2,0
4-5 May	Opinium	35,0	34,0	8,0	12,0	6,0	5,0	1,8	1,0
4-5 May	YouGov/The Sun	34,0	34,0	9,0	12,0	5,0	6,0	1,9	=
4-5 May	Survation/Daily Mirror	33,0	34,0	9,0	16,0	4,0	4,0	2,4	-1,0
3-5 May	ComRes/Daily Mail, ITV News	35,0	32,0	9,0	14,0	4,0	6,0	1,2	3,0
3-5 May	BMG/May2015.com, New Statesman	34,0	34,0	10,0	12,0	4,0	6,0	1,9	=
3-4 May	YouGov/The Sun	33,0	33,0	10,0	12,0	5,0	6,0	2,1	=
30 Apr-4 May	TNS	33,0	32,0	8,0	14,0	6,0	6,0	1,8	1,0
2-3 May	YouGov/The Sun	34,0	33,0	9,0	12,0	5,0	7,0	1,7	1,0
1-3 May	Lord Ashcroft	32,0	30,0	11,0	12,0	7,0	9,0	2,8	2,0
1-3 May	Populus	34,0	34,0	10,0	13,0	5,0	5,0	2,0	=
1-2 May	YouGov/Sunday Times	34,0	33,0	8,0	13,0	5,0	7,0	1,4	1,0
1-2 May	Survation/Mail on Sunday	31,0	34,0	8,0	17,0	4,0	6,0	2,8	-3,0
30 Apr-1 May	YouGov/The Sun	33,0	34,0	8,0	14,0	5,0	5,0	2,0	-1,0
30 Apr-1 May	Survation/Daily Mirror	33,0	34,0	9,0	16,0	3,0	5,0	2,5	-1,0
29-30 Apr	YouGov/Sun on Sunday	34,0	35,0	9,0	12,0	5,0	6,0	2,1	-1,0
29-30 Apr	Populus	33,0	33,0	9,0	15,0	4,0	5,0	2,0	=
29-30 Apr	YouGov/The Sun	34,0	35,0	8,0	12,0	5,0	6,0	2,0	-1,0
29-30 Apr	Panelbase	32,0	34,0	8,0	17,0	4,0	7,0	2,6	-2,0
28-30 Apr	ComRes/Independent on Sunday, Sunday Mirror	33,0	33,0	8,0	13,0	7,0	6,0	2,0	=
28-30 Apr	Opinium/The Observer	35,0	34,0	8,0	13,0	5,0	5,0	1,4	1,0
28-29 Apr	YouGov/The Sun	35,0	34,0	9,0	12,0	4,0	6,0	1,5	1,0
26-29 Apr	Ipsos MORI/Evening Standard	35,0	30,0	8,0	10,0	8,0	9,0	2,2	5,0
27-28 Apr	ComRes/ITV News, Daily Mail	35,0	35,0	7,0	11,0	6,0	6,0	2,4	=
27-28 Apr	YouGov/The Sun	34,0	35,0	9,0	12,0	4,0	6,0	1,9	-1,0
26-27 Apr	YouGov/The Sun	35,0	34,0	9,0	12,0	5,0	6,0	1,7	1,0
25-27 Apr	BMG/May2015.com	35,0	32,0	11,0	14,0	3,0	5,0	1,7	3,0
23-27 Apr	TNS	34,0	33,0	7,0	15,0	5,0	5,0	2,0	1,0
25-26 Apr	YouGov/The Sun	33,0	34,0	8,0	14,0	5,0	6,0	2,0	-1,0
24-26 Apr	Lord Ashcroft	36,0	30,0	9,0	11,0	7,0	7,0	1,8	6,0
24-26 Apr	ICM/The Guardian	35,0	32,0	9,0	13,0	5,0	6,0	1,2	3,0
24-26 Apr	Populus	33,0	36,0	8,0	14,0	5,0	4,0	2,4	-3,0

*Quelle:* [https://en.wikipedia.org/wiki/Opinion\\_polling\\_for\\_the\\_2015\\_United\\_Kingdom\\_general\\_election](https://en.wikipedia.org/wiki/Opinion_polling_for_the_2015_United_Kingdom_general_election); besucht am 15.7.2015; eigene Berechnungen (letzte zwei Spalten).

mit unterschiedlichen Koalitionsmöglichkeiten, die viele Spekulationen über die spätere Regierungsbildung erlaubte.

*Jon Mellon*, Mitglied der *British Election Study*, diskutierte am 11. Juni 2015 auf der Basis von Daten, die während des Wahlkampfes erhoben wurden (die Nachwahlerhebung war noch nicht abgeschlossen), fünf mögliche Ursachen, um *the great polling miss* zu erklären:<sup>12</sup>

1. *late swing*,
2. überhöhte Aussagen zur Wahlbeteiligung,
3. Stichproben- und Gewichtungsprobleme,
4. unklare Entscheidungen der „weiß-nicht“-Befragten,
5. *shy tory effect*.

Alle diese “Gründe” sind bereits bei früheren Wahlen in der einen oder anderen Form diskutiert worden. Neuere Erklärungsansätze erwähnte er nicht, was freilich auch daran liegen mag, dass er selbst zum „System“ der Wahl- und Umfrageforscher gehört. *Mellon* kommt jedenfalls zum Schluss, dass es Hinweise für Veränderungen der Wahlabsichten kurz vor der Wahl gab (*last minute swing*) und möglicherweise auch Hinweise auf Änderungen der Wahlbeteiligung. Er spricht von widersprüchlichen Hinweisen auf Wähler, die in Umfragen nicht ihre eigentlichen Präferenzen geäußert hätten, und von Veränderungen bei den vermeintlich Unentschlossenen („weiß nicht“-Kategorie). Er quantifiziert allerdings nicht. Befriedigen können Reflexionsversuche dieser Art kaum. Tatsächlichen Aufschluss könne man erst aus einer Nachwahlbefragung gewinnen, meint *Mellon*. Man muss abwarten, ob das wirklich gelingen wird. Ergebnisse von Nachwahlbefragungen sind häufig stark zugunsten des Wahlgewinners verzerrt. Bislang lassen sich eindeutige Hinweise auf systematische Gründe für die „schlechte Prognosequalität“ britischer Umfragen somit nicht erkennen.

### III. Die Situation in Deutschland

Auch in der Bundesrepublik Deutschland gab und gibt es Bundestags- und Landtagswahlen, in deren Vorfeld Ergebnisse von Umfragen veröffentlicht wurden, die vom Ergebnis der jeweiligen Wahl stark abwichen.<sup>13</sup> Solch vermeintlichen Fehlleistungen der Wahl- und Umfrageforschung werden in den Medien gerne themati-

<sup>12</sup> *Mellon J.*: Using the British Election Study to Understand the Great Polling Miss, 11. Juni 2015, in: <http://www.britishelectionstudy.com/bes-resources>.

<sup>13</sup> Auch die für das Fernsehen arbeitenden Umfrageinstitute lagen schon manches Mal „daneben“, vgl. *Hölscher, A.*: Orakel, Hexer, Meinungsmacher, in: *Süddeutsche Zeitung* vom 27.8.2004, hier 17.

siert, doch selten hinterfragt. Zwei Beispiele: So veröffentlichte Der Spiegel am 24.09.2002 einen Beitrag von *Norbert F. Pötzl* mit dem Titel: „Debakel für die Meinungsforscher“. Darin wurde zur BT-Wahl 2002 festgestellt: „Alle Umfragen vor der Wahl führten in die Irre: Die Erhebungen lagen weit neben dem tatsächlichen Ergebnis. Nicht einmal die grundsätzlichen Trends stimmten“. Im Artikel wurden zudem einzelne Institute und „falsche“ Ergebnissen dargestellt, im Grundtenor handelte es sich um eine Art Abrechnung mit der „Meinungsforscherzunft“ als Ganzer. Tatsache ist (siehe Tab. 3), dass die Abweichungen der veröffentlichten Umfragen vom Endergebnis bei allen Parteien (große und kleine) mit der Ausnahme eines Instituts bezüglich der FDP innerhalb des Fehlerbereichs liegen. Die mittlere Abweichung über alle Parteien liegt zwischen 0,7% und 1,4%. Bis auf ein Institut sagten alle eine Mehrheit für Rot-Grün und ein Scheitern der PDS voraus.

Das zweite Beispiel ist jüngerem Datums (13.05.2015) und wurde vom ZEIT-Herausgeber *Josef Joffe* verfasst. Der Titel war Programm: „Demoskopien-Desaster: Warum die Auguren den falschen Wahlausgang prophezeien“. *Joffe* bezieht sich zunächst auf Großbritannien und stellt fest: „Der größte Verlierer war nicht *Labour*, sondern die Umfrage-Industrie“. Dann stellt er eine Verbindung zu Deutschland her und wirft einem Institut vor, bei den Bundestagswahlen 2009, 2005 und 2002 die Union bzw. auch die SPD weit verfehlt zu haben. Der süffisante Zusatz lautet: „Solche krassen Abweichungen sind Routine“. Doch alle in diesem Zusammenhang aufgeführten Zahlen sind falsch. Wir erleben eine krude Mischung von Schlagworten und Halbwahrheiten; eine systematische Darstellung ist dem Autor entweder nicht möglich oder es wird hierauf bewusst verzichtet.

Wir müssen grundsätzlich differenzieren zwischen den veröffentlichten Umfragen vor der Wahl, die in der Regel nicht als Zeitpunkt-Messungen, sondern als „Prognosen“ wahrgenommen werden, und den 18 Uhr Prognosen am Wahltag (vgl. Tab. 9 im Anhang), die auf Daten aus *exit polls* beruhen, die viele Unsicherheiten der Vorwahlbefragungen, die das Ergebnis einer Umfrage beeinflussen können, ausschließen. So liegt ein zentraler Unterschied in der Messung der abhängigen Variable, denn in *exit polls* wird gerade vollzogenes Verhalten erhoben, während in Vorwahlumfragen die Wahlabsicht, also die (mögliche) Handlung, die in der Zukunft liegt, abgefragt wird. Selbst in Zeiten hoher Briefwahlanteile ist der Unterschied zwischen Vorwahlumfragen und Befragungen am Wahltag beträchtlich. Der zweite zentrale Unterschied liegt in der Größe der Stichprobe, die bei Vorwahlbefragungen in der Regel zwischen 1000 und 2000 Befragten liegt, bei *exit polls* jedoch das Zehn- bis Hundertfache dessen betragen kann und somit zu unterschiedlich großen Fehlerbereichen führt. Daher werden



*exit polls* hier nur kursorisch behandelt, der Schwerpunkt liegt auf Umfrageveröffentlichungen vor, doch in zeitlicher Nähe zum Wahltag.

Obwohl sich dieser Beitrag mit der „Prognosefähigkeit“ von Umfrageveröffentlichungen beschäftigt, werden nicht „Prognosen“ im Sinne eines wissenschaftlichen Modells zur Voraussage von künftigem Verhalten von Wählern behandelt,<sup>14</sup> sondern Veröffentlichungen der verschiedenen Institute, die man eher als „Projektionen“ auf ein zukünftiges Ereignis auf der Basis erhobener Daten bezeichnen könnte. Welche Daten einzelne Institute tatsächlich erhoben haben, um Parteistärken zu veröffentlichen, ist nicht immer klar. Bekannt sind die veröffentlichten Zahlen, in der Öffentlichkeit als „Prognosen“ verstanden, da sie häufig auch als Prognosen bezeichnet werden.<sup>15</sup>

## 1. Datengrundlage

Der zeitliche Rahmen der Betrachtung wird eingegrenzt. Erste Umfragen, in denen die Wahlabsicht erhoben wurde, gab es bereits in der Fünfziger Jahren.<sup>16</sup> Prognosen zum möglichen Wahlausgang kamen in größerem Umfang erst in den Sechziger Jahren auf, wobei Wahlsendungen im Fernsehen von 1965 an eine wichtige Rolle spielten.

Für die Analysen dieses Beitrags soll die Zeit nach der Vereinigung der beiden deutschen Staaten betrachtet werden, in der es ungleich mehr politische Veränderungen sowohl in der Parteienlandschaft (neben den etablierten Parteien behaupten sich die Grünen, es entstanden die PDS und später Die Linke, die AfD sowie etliche andere kleinere Parteien) als auch im Wahlverhalten der Bevölkerung gegeben hat, als in den Jahrzehnten zuvor.<sup>17</sup>

<sup>14</sup> Verdienstvolle Versuche machten immer wieder *Helmut Norpoth* und *Thomas Gschwend*. Vgl. *Norpoth, H./Gschwend, T.*: Politbarometer und Wahlprognosen: Die Kanzlerfrage, in *Wüst, A.M.* (Hg.): Politbarometer, Opladen, 2003, 109-123; *Norpoth, H./Gschwend, T.*: Mit Rot-Grün ins Schwarze getroffen: Prognosemodell besteht Feuertaufe, in *Falter, J.W.* et al. (Hg.): Wahlen und Wähler. Analysen aus Anlass der Bundestagswahl 2002, Wiesbaden, 2005, 371-387.

<sup>15</sup> *Wüst, A. M.*: Stimmung, Projektion, Prognose?, in *ders* (Hg.): Politbarometer, Opladen, 2003, 83-107.

<sup>16</sup> Eine ausführliche Beschreibung dieser Entwicklung findet sich bei *Jessen, A.*: Perspektiven der politischen Meinungsforschung, Wiesbaden, 2014.

<sup>17</sup> Seit Ende der 1980er Jahre zeigt sich zunächst auf der Landesebene, nach 1990 auch auf Bundesebene ein hohes politisches Veränderungspotential. Regierungswechsel, die nicht durch Parteientscheid, wie in der frühen Bundesrepublik, sondern auch durch die Entscheidung der Wähler herbeigeführt werden, häufen sich. Bei 100 Möglichkeiten, eine Regierung abzuwählen, gab es bis 1987 nur sieben Fälle des Regierungsverlustes mit anschließender Regierung unter einem Ministerpräsidenten, der einer anderen großen Partei angehörte (dabei fallen fünf Fälle in die Parteien-Konsolidierungsphase bis Ende der 1950er Jahre). Nach 1987 gab es bis 1998 bei 38 Möglichkeiten (die ersten Landtagswahlen 1990 in den neuen Bundesländern sind ausgenommen) bereits zehn vom Wähler herbeigeführte Wechsel (darunter ein Regierungswechsel auf Bundesebene). Seit 1998 bis 2013 gibt es bei 60 Wahlen 14 Regierungswechsel. Vgl. *Forschungsgruppe Wahlen e.V.*: Wahlergebnisse in Deutschland 1946-2013, Mannheim, 2014.

Bei der Darstellung der Umfragen werden wir uns auf veröffentlichte bundesweite Umfragen vor Bundestagswahlen seit 1998 (v.a. Emnid, Forsa, Forschungsgruppe Wahlen, Infratest dimap und IfD Allensbach) beschränken, die bis zu zehn Tage vor dem Wahltermin veröffentlicht worden sind. Wir greifen hier und nachfolgend auf die umfangreiche Dokumentation von [www.wahlrecht.de](http://www.wahlrecht.de) zurück. Leider beinhaltet die Datenbank erst Umfrageergebnisse ab 1995, so dass die Bundestagswahl 1994 ausgeklammert werden muss.<sup>18</sup> Für Landtagswahlen untersuchen wir eine Auswahl an Bundesländern (Nordrhein-Westfalen, Bayern, Baden-Württemberg, Hessen, Brandenburg, Schleswig-Holstein und den Stadtstaat Hamburg) seit 1995. Die Auswahl der Länder erfolgte nach regionaler und struktureller Varianz. Aufgrund der Tatsache, dass für die Länderebene etliche Veröffentlichungen erst 14 bis 16 Tage vor dem Wahltermin erfolgt sind und eine Begrenzung auf zehn Tage oftmals nur eine einzige Umfrage beinhaltet hätte, wurde in Bezug auf Landtagswahlen der Zeitraum der berücksichtigten Umfrageveröffentlichungen vor der Wahl auf bis zu 21 Tage erweitert.

Bei der kurzen Dokumentation der 18:00 Uhr Prognosen im Anhang zu diesem Artikel ziehen wir die von ARD und ZDF veröffentlichten Daten heran, zudem Fälle, in denen RTL Prognosen auf der Grundlage von *exit polls* erstellt hat. Für diese Prognosen am Wahltag erhebt seit 1998 Infratest dimap für die ARD entsprechende Daten, für das ZDF die Forschungsgruppe Wahlen, für RTL das Institut Forsa.

## 2. Fragestellung

Die folgenden Fragen versuchen wir im Rahmen dieses Beitrags zu beantworten:

1. Wie nahe kommen die veröffentlichten Umfragedaten der Parteistärken an das tatsächliche Wahlergebnis heran?
2. Gibt es einen Trend, der es rechtfertigen würde, von abnehmender Prognosequalität zu sprechen?
3. Finden sich Unterschiede zwischen einzelnen Umfrageinstituten hinsichtlich der Abweichung vom späteren Wahlergebnis?
4. Gibt es veränderte Rahmenbedingungen, die eine abnehmende Qualität von Umfragen erklären?

## 3. Analysemethoden

Um Antworten auf die ersten drei Fragen geben zu können, führen wir eine mehr-

<sup>18</sup> Die Daten wurden am Lehrstuhl für Vergleichende Regierungslehre an der Universität Freiburg aufbereitet und zur Verfügung gestellt. Die Autoren haben die Daten geprüft und weiter verarbeitet. Mögliche Fehler verantworten ausschließlich sie.

stufige Analyse der Daten durch. Für Bundestagswahlen bedienen wir uns zunächst einer deskriptiven Darstellung aller Umfrageveröffentlichungen in den letzten zehn Tagen vor den Bundestagswahlen 1998 bis 2013. Wir stellen die entsprechenden Umfrageveröffentlichungen nach Parteianteilen dem späteren Wahlergebnis gegenüber. Zudem berechnen wir die Beträge der Abweichungen je Partei (und der Sonstigen) sowie die Summe der Abweichungen pro Umfrageveröffentlichung. Die deskriptive Analyse ermöglicht uns eine erste Bewertung der Prognosequalität und damit eine vorläufige Antwort auf die Fragen 1 und 3. Für Landtagswahlen entfällt dieser Analyseschritt aufgrund der höheren Anzahl von Wahlen.

Sowohl für Bundestags- als auch für Landtagswahlen wird untersucht, ob sich die Prognosequalität im Zeitverlauf verschlechtert hat (Frage 2). Hierzu berechnen wir das arithmetische Mittel der Abweichungen jedes prognostizierten Parteianteils vom späteren Wahlergebnis für jede Umfrageveröffentlichung.<sup>19</sup> Wir nehmen keine Gewichtung der Abweichungen nach Parteianteilen vor,<sup>20</sup> da wir folgende Prämisse einführen: Jedes Institut prognostiziert denselben Parteienmarkt und hat damit das identische Risiko großer und kleiner Abweichungen, die von der jeweiligen Parteistärke abhängen. Wir respektieren darüber hinaus die individuelle Entscheidung jedes Instituts, Anteile für mehr oder weniger Parteien zu veröffentlichen (dies betrifft insbesondere REP, DVU, Piraten, FW, AfD, aber z.T. auch die PDS/Die Linke und die FDP) und errechnen unter Berücksichtigung der Anzahl der Parteien die durchschnittliche Abweichung je Parteianteil einer Umfrageveröffentlichung. Da es sich bei zusätzlich veröffentlichten Parteianteilen zumeist um kleine Parteien handelt, deren Stimmenanteil in der Regel einen kleineren Fehlerbereich aufweist und dadurch eine insgesamt bessere Projektion ermöglicht, nehmen wir in den nachfolgenden Regressionsmodellen eine Variable auf, die diese potenzielle „Wettbewerbsverzerrung“ zu korrigieren sucht.<sup>21</sup>

Wir stellen zunächst die mittleren Abweichungen für jede Bundestagswahl und alle Landtagswahlen des Ländersamples pro Jahr graphisch dar. Daran anschließend analysieren wir mittels eines linearen Regressionsmodells, ob es einen statistisch signifikanten Trend hin zu größeren Abweichungen vom späteren

<sup>19</sup> MAE (*mean absolute error*); vgl. Groß, J.: Die Prognosen von Wahlergebnissen: Ansätze und empirische Leistungsfähigkeit, Wiesbaden, 2010, hier 126-132.

<sup>20</sup> Wir teilen zwar grundsätzlich die Einschätzung Jessens (a.a.O., hier 32), dass „die Abweichung ins Verhältnis zur Größe des Stimmenanteils gesetzt werden [müsste]“ (MAPE = *mean absolute percentage error*), doch erscheinen uns die Parteiensysteme des Bundes und der Länder immer noch vergleichsweise stabil, um diese Varianz vernachlässigen zu können. Darüber hinaus haben sowohl MAE als auch MAPE Schwächen (Jessen, a.a.O., hier 31), so dass im Zweifel hier auf das einfachere Maß zurückgegriffen wird.

<sup>21</sup> Die Anzahl prognostizierter Parteien wird als potenziell erklärende Variable berücksichtigt.

Wahlergebnis gibt, wobei wir sowohl für den Zeitpunkt der Veröffentlichung als auch für die einzelnen Institute (Frage 3) kontrollieren. In der Landtagswahlanalyse wird darüber hinaus auch für die einzelnen Bundesländer kontrolliert.

Schließlich diskutieren wir Frage 4 angesichts unserer empirischen Befunde und des Forschungsstands.

4. Ergebnisse der empirischen Analyse

a) Bundestagswahlen

Wenn wir zunächst auf die Umfrageveröffentlichungen in den zehn Tagen vor den letzten fünf Bundestagswahlen blicken (Tab. 2 bis 6) und diese mit den späteren Wahlergebnissen vergleichen, dann lässt sich den Umfrageinstituten alles in allem eine passable Treffgenauigkeit attestieren. 1998 wichen die veröffentlichten Anteile für die im Bundestag vertretenen Parteien und alle anderen Parteien zusammengekommen pro Institut in der Summe 3,6% bis 7,9% vom Wahlergebnis ab. Vor allem das Ergebnis der CDU/CSU wurde überschätzt, das der PDS unterschätzt. In den Veröffentlichungen aller Institute lag das rot-grüne Lager vor dem schwarz-gelben, wenn auch der Abstand zwischen den Lagern unterschätzt wurde.

Tabelle 2: Veröffentlichungen von Umfrageinstituten im Vorfeld der Bundestagswahl 1998

		CDU/CSU			SPD			GRU			FDP			PDS			Andere			
	Datum	Δ			Δ			Δ			Δ			Δ			Δ			Σ Δ
Wahl- ergebnis	27.9.	35,1			40,9			6,7			6,2			5,1			5,9			
	Emnid	25.9.	39,0	3,9	40,0	0,9	7,0	0,3	5,0	0,6	4,0	1,1	5,0	0,9	7,7					
		18.9.	39,0	3,9	41,0	0,1	6,0	0,7	5,0	1,2	5,0	0,1	4,0	1,9	7,9					
Forsa	25.9.	38,0	2,9	42,0	1,1	6,0	0,7	5,0	1,2	4,0	1,1	5,0	0,9	7,9						
	23.9.	37,0	1,9	42,0	1,1	6,0	0,7	6,0	0,2	4,0	1,1	5,0	0,9	5,9						
FGW	18.9.	37,5	2,4	39,5	1,4	6,0	0,7	5,5	0,7	4,5	0,6	7,0	1,1	6,9						
Infratest	26.9.	38,0	2,9	40,0	0,9	6,0	0,7	6,0	0,2	6,0	0,9	5,0	0,9	6,5						
Dimap	19.9.	38,0	2,9	41,0	0,1	7,0	0,3	6,0	0,2	5,0	0,1	5,0	0,9	4,5						
IfD	26.9.	36,0	0,9	40,5	0,4	6,0	0,7	6,5	0,3	5,0	0,1	6,0	0,1	3,6						
Allensbach	23.9.	36,1	1,0	41,1	0,2	7,0	0,3	6,5	0,3	4,7	0,4	4,6	1,4	3,6						

Quelle: [www.wahlrecht.de](http://www.wahlrecht.de). Veröffentlichte Umfragen vor der Bundestagswahl 1998 im Zeitraum 10 Tage vor der Wahl. Forsa hat täglich Ergebnisse veröffentlicht; es sind hier nur die letzten beiden unterschiedlichen Veröffentlichungen dokumentiert. Aus Gründen der Vergleichbarkeit wurden die z.T., aber nicht konsistent veröffentlichten Anteile für REP und DVU unter „Andere“ summiert.

2002 betrug die Summe aller Abweichungen je Institut zwischen 4,2% und 8,2%. Mit Ausnahme einer Veröffentlichung blieb die PDS/Linke unter 5%. Dass das Ergebnis enger als 1998 würde, spiegelte sich in allen Umfrageveröffentlichungen wieder. Mit Ausnahme zweier Veröffentlichungen vom IfD Allensbach und einer von Emnid lag das rot-grüne Lager jedoch vor dem schwarz-gelben.

Tabelle 3: Veröffentlichungen von Umfrageinstituten im Vorfeld der Bundestagswahl 2002

		CDU/CSU		SPD		GRU		FDP		PDS/Linke		Andere		
	Datum	Δ		Δ		Δ		Δ		Δ		Δ		Σ Δ
<b>Wahl- ergebnis</b>	22.9.	38,5		38,5		8,6		7,4		4,0		3,0		
<b>Emnid</b>	14.9.	37,0	1,5	39,0	0,5	7,0	1,6	8,0	0,6	5,0	1,0	4,0	1,0	6,2
	12.9.	38,0	0,5	38,0	0,5	7,0	1,6	8,0	0,6	4,0	0,0	5,0	2,0	5,2
	20.9.	37,5	1,0	39,0	0,5	7,0	1,6	7,5	0,1	4,3	0,3	4,7	1,7	5,2
<b>Forsa</b>	18.9.	38,0	0,5	40,0	1,5	7,0	1,6	8,0	0,6	4,0	0,0	3,0	0,0	4,2
	13.9.	38,0	0,5	40,0	1,5	7,0	1,6	8,0	0,6	4,0	0,0	3,0	0,0	4,2
	13.9.	37,0	1,5	40,0	1,5	7,0	1,6	7,5	0,1	4,5	0,5	4,0	1,0	6,2
<b>FGW</b>	13.9.	36,0	2,5	38,5	0,0	8,0	0,6	8,5	1,1	4,7	0,7	4,3	1,3	6,2
<b>Infratest dimap</b>	20.9.	37,0	1,5	37,5	1,0	7,5	1,1	9,5	2,1	4,5	0,5	4,0	1,0	7,2
<b>IfD Allens- bach</b>	17.9.	37,3	1,2	37,0	1,5	7,2	1,4	10,1	2,7	4,4	0,4	4,0	1,0	8,2

Quelle: [www.wahlrecht.de](http://www.wahlrecht.de). Veröffentlichte Umfragen vor der Bundestagswahl 2002 im Zeitraum 10 Tage vor der Wahl.

Bei der Bundestagswahl 2005 sah es grundlegend anders aus. Die Summe aller Abweichungen ist für jedes Institut und alle Veröffentlichungen zweistellig (11,5% bis 15,3%). Grund dafür ist eine erhebliche Überschätzung der CDU/CSU sowie eine mäßige Unterschätzung der FDP. SPD, Grüne und die Linke wurden vergleichsweise gut getroffen. Dies legt nahe, dass es in den letzten Tagen vor der Wahl innerhalb des konservativen Lagers eine Verschiebung zugunsten der FDP gegeben hat, die auch von den Instituten, die kurz vor der Wahl veröffentlicht haben, nicht erfasst wurde.<sup>22</sup> Es gab also einen erheblichen Anteil taktischer Wähler im konservativen Lager, der sich ‚naturgemäß‘ sehr spät entschied.

<sup>22</sup> Die Analyse der Forschungsgruppe Wahlen (Umfrage in der Woche vor der Wahl) zeigt unter FDP-Wählern einen hohen Anteil, der parteipolitisch bei der Union zuhause ist: 41% der FDP-Wähler geben als ihre präferierte Partei die CDU/CSU an, 36% der FDP-Wähler identifizieren sich mit der Union, davon 22% sogar stark oder sehr stark.

Tabelle 4: Veröffentlichungen von Umfrageinstituten im Vorfeld der Bundestagswahl 2005

		CDU/CSU		SPD		GRU		FDP		Linke		Andere		
	Datum	Δ		Δ		Δ		Δ		Δ		Δ		Σ Δ
Wahl- ergebnis Emnid	17.9.	35,2		34,2		8,1		9,8		8,7		3,9		
	13.9.	42,0	6,8	33,5	0,7	7,0	1,1	6,5	3,3	8,0	0,7	3,0	0,9	12,4
	12.9.	40,5	5,3	34,5	0,3	7,0	1,1	7,0	2,8	8,0	0,7	3,0	0,9	13,5
	8.9.	42,0	6,8	33,0	1,2	7,0	1,1	7,0	2,8	9,0	0,3	2,0	1,9	14,1
Forsa	16.9.	42,0	6,8	33,0	1,2	6,5	1,6	7,5	2,3	7,5	1,2	3,5	0,4	13,5
	12.9.	42,0	6,8	35,0	0,8	7,0	1,1	6,0	3,8	7,0	1,7	3,0	0,9	15,3
FGW	9.9.	41,0	5,8	34,0	0,2	7,0	1,1	7,0	2,8	8,0	0,7	3,0	0,9	11,5
Infratest dimap	8.9.	41,0	5,8	34,0	0,2	7,0	1,1	6,5	3,3	8,5	0,2	3,0	0,9	11,5
InfD	16.9.	41,5	6,3	32,5	1,7	7,0	1,1	8,0	1,8	8,5	0,2	2,5	1,4	12,5
Allensbach	13.9.	41,7	6,5	32,9	1,3	7,2	0,9	7,0	2,8	8,5	0,2	2,7	1,2	12,9
	9.9.	41,5	6,3	32,7	1,5	7,2	0,9	7,0	2,8	8,9	0,2	2,7	1,2	12,9
GMS	12.9	42,0	6,8	33,0	1,2	8,0	0,1	7,0	2,8	7,0	1,7	3,0	0,9	13,5

Quelle: [www.wahlrecht.de](http://www.wahlrecht.de). Veröffentlichte Umfragen vor der Bundestagswahl 2005 im Zeitraum 10 Tage vor der Wahl.

Dies lag zum Teil an der sich zuspitzenden Diskussion um mögliche Koalitionen. Über eine Große Koalition wurde zwar im Vorfeld der Wahl spekuliert, die Umfragedaten gaben dazu jedoch keinen Anlass, alle Institute sagten eine Regierung von Schwarz/Gelb voraus. Der Abstand zwischen beiden Lagern war am Wahltag (2,7%) dann allerdings ungleich enger als in den Umfragen vor der Wahl (zwischen 6% und 10%). Die Abweichungen der zeitlich späteren Veröffentlichungen erwiesen sich bei dieser Wahl den früheren gegenüber als höher (FGW und Infratest dimap).

Verglichen mit 2005 war 2009 für die Umfrageinstitute wieder eine erfolgreichere Wahl. Die Abweichungen zum späteren Wahlergebnis betrugen in der Summe 4,2% bis 9,0%. Vor allem SPD und FDP wurden unterschätzt. Am Sieg für Schwarz-Gelb bestand jedoch bei keiner der Vorwahlveröffentlichungen ein Zweifel. Auch der Abstand zwischen den beiden Lagern wurde relativ gut getroffen.

Und auch bei der Bundestagswahl 2013 konnten die Umfrageinstitute zeigen, dass ihre Instrumente und Projektionsmodelle funktionstüchtig sind. Die Differenz zwischen Veröffentlichung und Wahlergebnis betrug 4,9% bis 9,9%. Die Grünen wurden bei allen Instituten überschätzt, die AfD unterschätzt. Kein Insti-

tut hatte das Scheitern der FDP an der 5%-Hürde projiziert, was zur Folge hatte, dass alle Institute von einer Mehrheit für Schwarz-Gelb ausgingen, denn alle (außer INSA) ermittelten für die AfD Werte von unter 5%.

Tabelle 5: Veröffentlichungen von Umfrageinstituten im Vorfeld der Bundestagswahl 2009

		CDU/CSU		SPD		GRU		FDP		Linke		Andere		
	Datum	Δ		Δ		Δ		Δ		Δ		Δ		Σ Δ
<b>Wahl- ergebnis</b>	27.9.	33,8		23,0		10,7		14,6		11,9		6,0		
<b>Emnid</b>	17.9.	35,0	1,2	25,0	2,0	11,0	0,3	13,0	1,6	12,0	0,1	4,0	2,0	6,5
<b>Forsa</b>	25.9.	33,0	0,8	25,0	2,0	10,0	0,7	14,0	0,6	12,0	0,1	6,0	0,0	4,2
	23.9.	35,0	1,2	26,0	3,0	11,0	0,3	13,0	1,6	10,0	1,9	5,0	1,0	8,0
<b>FGW</b>	18.9.	36,0	2,2	25,0	2,0	10,0	0,7	13,0	1,6	11,0	0,9	5,0	1,0	8,4
<b>Infratest dimap</b>	17.9.	35,0	1,2	26,0	3,0	10,0	0,7	14,0	0,6	11,0	0,9	4,0	2,0	8,4
<b>IfD Allens- bach</b>	22.9.	35,0	1,2	24,0	1,0	11,0	0,3	13,5	1,1	11,5	0,4	5,0	1,0	5,0
<b>GMS</b>	18.9	36,0	2,2	25,0	2,0	11,0	0,3	13,0	1,6	11,0	0,9	4,0	2,0	9,0

Quelle: www.wahlrecht.de. Veröffentlichte Umfragen vor der Bundestagswahl 2009 im Zeitraum 10 Tage vor der Wahl.

Tabelle 6: Veröffentlichungen von Umfrageinstituten im Vorfeld der Bundestagswahl 2013

		CDU/CSU		SPD		GRU		FDP		Linke		AfD		Andere		
	Datum	Δ		Δ		Δ		Δ		Δ		Δ		Δ		Σ Δ
<b>Wahl- ergebnis</b>	22.9.	41,5		25,7		8,4		4,8		8,6		4,7		6,2		
<b>Emnid</b>	20.9.	39,0	2,5	26,0	0,3	9,0	0,3	6,0	1,2	9,0	0,4	4,0	0,7	7,0	0,8	6,5
	15.9.	39,0	2,5	26,0	0,3	10,0	1,6	5,0	0,2	9,0	0,4	4,0	0,7	7,0	0,8	5,5
<b>Forsa</b>	20.9.	40,0	1,5	26,0	0,3	10,0	1,6	5,0	0,2	9,0	0,4	4,0	0,7	6,0	0,2	4,9
	17.9.	39,0	2,5	25,0	0,7	9,0	0,6	5,0	0,2	10,0	1,4	3,0	1,7	9,0	2,8	9,9
<b>FGW</b>	19.9.	40,0	1,5	27,0	1,3	9,0	0,6	5,5	0,7	8,5	0,1	4,0	0,7	6,0	0,2	5,1
	13.9.	40,0	1,5	26,0	0,3	11,0	2,6	6,0	1,2	8,0	0,6	4,0	0,7	5,0	1,2	7,9
<b>Infratest dimap</b>	12.9.	40,0	1,5	28,0	2,3	10,0	1,6	5,0	0,2	8,0	0,6	2,5	2,2	6,5	0,3	8,8
<b>IfD Allens- bach</b>	20.9	39,5	2,0	27,0	1,3	9,0	0,6	5,5	0,7	9,0	0,4	4,5	0,2	5,5	1,2	6,4
	18.9.	39,0	2,5	26,0	0,3	11,0	2,6	6,0	1,2	9,0	0,4	3,5	1,2	5,5	1,2	9,4
<b>INSA</b>	19.9.	38,0	3,5	28,0	2,3	8,0	0,4	6,0	1,2	9,0	0,4	5,0	0,3	6,0	0,2	8,3

Quelle: www.wahlrecht.de. Veröffentlichte Umfragen vor der Bundestagswahl 2013 im Zeitraum 10 Tage vor der Wahl. Aus Gründen der Vergleichbarkeit wurden die z.T., aber nicht konsistent veröffentlichten Anteile für Piraten und FW unter „Andere“ summiert.

Der tatsächliche Vorsprung des bürgerlichen Lagers vor Rot-Grün beträgt 12,2%. Kein Institut hat Ergebnisse aus Umfragen in dieser Größenordnung veröffentlicht. Die Summe der Abweichungen ist über alle Institute am höchsten bei der CDU/CSU; kein Institut taxierte die Unionsparteien über 40%. Wenn Institute mehr als ein Umfrageergebnis innerhalb der letzten zehn Tage vor der Wahl veröffentlicht haben, so ist die Summe der Abweichungen bei der wahltagnahen Veröffentlichung kleiner als bei der wahltagfernen Veröffentlichung.

Bereits die Vergleiche zwischen veröffentlichten Parteianteilen und Wahlergebnissen dokumentierten eine insgesamt moderate Varianz der Abweichungen nach Wahlen und Umfrageinstituten. Abbildung 1, in der die mittlere Abweichung zwischen veröffentlichten Parteianteilen (ohne andere Parteien) und dem späteren Wahlergebnis nach Wahljahren dargestellt ist, unterstreicht diesen Befund. Ein Trend ist kaum feststellbar, aber ein erheblicher Unterschied zwischen der Treffgenauigkeit der Umfrageveröffentlichungen im Jahr 2005 und den anderen betrachteten Wahljahren. Im Jahr 2005 gab es offensichtlich Faktoren, die eine akurate Messung und Projektion der Parteianteile erschwert haben.

Die erste Regressionsanalyse in Tabelle 7 bestätigt, dass es auch unter Berücksichtigung des Zeitpunkts jeder Veröffentlichung keinen Trend hin zu Umfrageveröffentlichungen gibt, die das Wahlergebnis schlechter treffen (Modell 1). Wird die Anzahl der projizierten Parteianteile berücksichtigt (Modell 2), zeigt sich eine tendenzielle Verschlechterung im Zeitverlauf ( $p < 0,1$ ). Die zusätzliche Berücksichtigung der einzelnen Institute erbringt keine Modellverbesserung (Modell 3).

Anders verhält es sich, wenn die Wahl 2005 als Besonderheit Berücksichtigung findet (Modell 4). Sobald das Jahr 2005 als Dummy-Variable in das Regressionsmodell eingeführt wird, wird es zum mit Abstand erklärungskräftigsten Faktor, der einen sehr hohen Anteil der Abweichungen zwischen Umfrageveröffentlichungen und Wahlergebnissen statistisch erklärt. Ein Institut, dem allerdings auch die meisten Veröffentlichungen zuzurechnen sind, weicht nun in der Treffgenauigkeit signifikant von den anderen Instituten ab.<sup>23</sup> Da dieses Institut insgesamt Werte für eine unterdurchschnittlichen Anzahl von Parteien veröffentlicht

<sup>23</sup> 18 von insgesamt 54 Umfrageveröffentlichungen stammen von diesem Institut. 1998 und 2005 weist der Durchschnitt der veröffentlichten Parteianteile des Instituts die im deskriptiven Institutsvergleich größte Abweichung vom späteren Wahlergebnis auf.



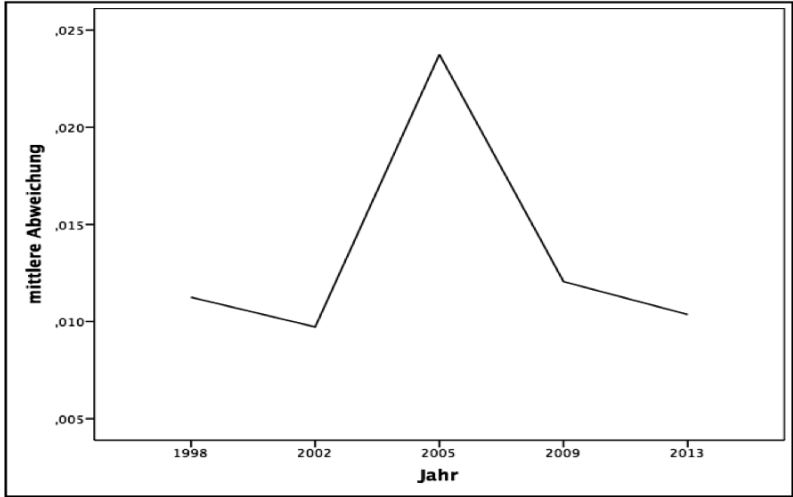
hat (5,2 im Vergleich zu 5,4 im Durchschnitt), trägt die Integration des Instituts-Dummies in das Modell dazu bei, dass die Anzahl prognostizierter Parteien nun keinen signifikanten Einfluss mehr auf die Abweichungen vom späteren Wahlergebnis hat. Zentrales Ergebnis des letzten Modells (4), dass die tendenzielle Signifikanz eines Trends nun nicht mehr feststellbar ist. Stattdessen hat nun – wie erwartet – der Zeitpunkt der Veröffentlichung einen tendenziellen Einfluss auf die Treffgenauigkeit einer Umfrageveröffentlichung.

Tabelle 7: Erklärungsfaktoren für Abweichung eines Umfrageergebnisses vom Wahlergebnis (Bund; lineare Regressionen)

	Modell 1 b (SE)	beta	Modell 2 b (SE)	beta	Modell 3 b (SE)	beta	Modell 4 b (SE)	beta
<i>Schritt 1</i>								
Konstante	-0,042 (0,294)		-0,672 (0,352)		-0,672 (0,352)		-0,062 (0,143)	
Trend (Jahr)	0,000 (0,000)	0,025	+0,000 (0,000)	0,330	+0,000 (0,000)	0,330	0,000 (0,001)	0,033
Abstand zur Wahl (Tage)	0,000 (0,000)	0,173	0,000 (0,000)	0,098	0,000 (0,000)	0,098	+0,000 (0,000)	0,117
<i>Schritt 2 bzw. 5 (schrittweise)</i>								
N Parteien			** -0,004 (0,001)	-0,486	** -0,004 (0,001)	-0,486		
<i>Schritt 3 bzw. 6 (schrittweise)</i>								
Emnid								
Forsa							*0,002 (0,001)	0,155
FGW								
GMS								
Infratest								
IfD Allensbach								
INSA								
<i>Schritt 4 (schrittweise)</i>								
1998								
2002								
2005							***0,013 (0,001)	0,891
2009								
2013								
R <sup>2</sup> <sub>adj.</sub>	-0,008		0,117		0,117		0,785	
N	54		54		54		54	

+ p < 0,1; \* p < 0,05; \*\*p < 0,01; \*\*\*p < 0,001.

Abbildung 1: Mittlere Abweichung der Umfrageveröffentlichungen (Parteianteile) vom späteren Wahlergebnis bei Bundestagswahlen seit 1998 im Zeitverlauf (Skala von 0 bis 1)



b) Landtagswahlen

Abbildung 2: Mittlere Abweichung der Umfrageveröffentlichungen (Parteianteile) vom späteren Wahlergebnis bei Landtagswahlen seit 1995 im Zeitverlauf (Skala von 0 bis 1)

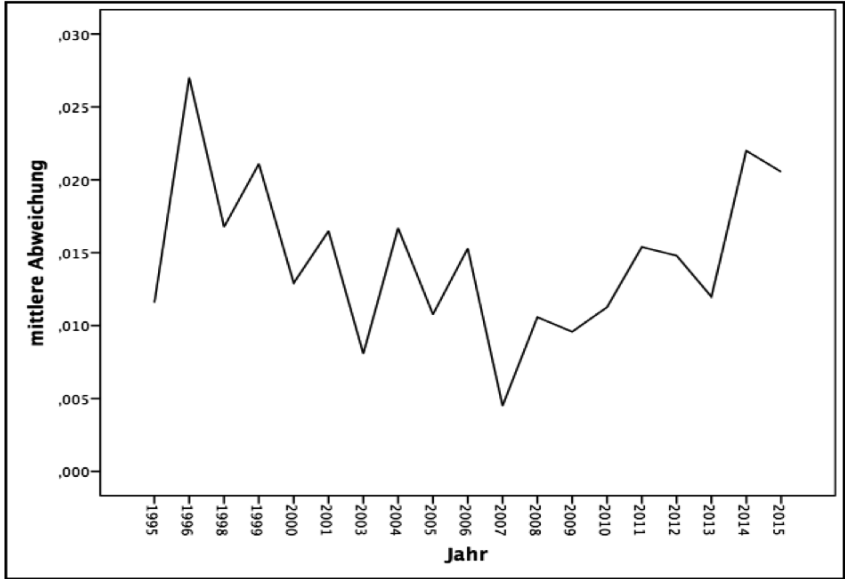


Tabelle 8: Erklärungsfaktoren für Abweichung eines Umfrageergebnisses vom Wahlergebnis (Länder; lineare Regressionen)

	Modell 1		Modell 2		Modell 3		Modell 4	
	b (SE)	beta	b (SE)	Beta	b (SE)	beta	b (SE)	beta
<i>Schritt 1</i>								
<b>Konstante</b>	-0,011 (0,198)		-0,011 (0,198)		-0,011 (0,198)		-0,011 (0,198)	
<b>Trend (Jahr)</b>	0,000 (0,000)	0,009	0,000 (0,000)	0,009	0,000 (0,000)	0,009	-0,000 (0,000)	-0,025
<b>Abstand zur Wahl (Tage)</b>	*0,000 (0,000)	0,186	*0,000 (0,000)	0,186	*0,000 (0,000)	0,186	*0,000 (0,000)	0,163
<i>Schritt 2 (schrittweise)</i>								
<b>N Parteien</b>								
<i>Schritt 3 (schrittweise)</i>								
<b>Emnid</b>								
<b>Forsa</b>								
<b>FGW</b>								
<b>GMS</b>								
<b>Infratest</b>								
<b>YouGov</b>								
<i>Schritt 4 (schrittweise)</i>								
<b>Bayern</b>								
<b>Brandenburg</b>							***0,006 (0,002)	0,255
<b>Baden- Württemberg</b>								
<b>Hessen</b>								
<b>Hamburg</b>							**0,003 (0,001)	0,215
<b>NRW</b>								
<b>Schleswig- Holstein</b>							*-0,002 (0,001)	-0,172
<b>R<sup>2</sup><sub>adj.</sub></b>	0,022		0,022		0,022		0,153	
<b>N</b>	160		160		160		160	

+ p < 0,1; \* p < 0,05; \*\*p < 0,01; \*\*\*p < 0,001.

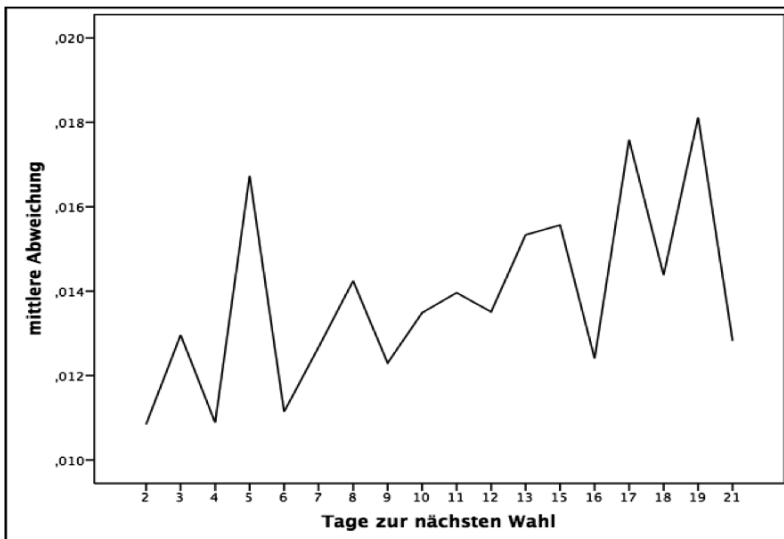
Die Entwicklung der Prognosegenauigkeit bei Landtagswahlen der sieben ausgewählten Bundesländer (Abb. 2) zeigt ein etwas anderes Bild als im Bund (Abb. 1). Bis 2007 nehmen die mittleren Abweichungen ab, seitdem wieder zu.<sup>24</sup> Wie alle Regressionsmodelle (Tab. 8) ausweisen, ist für den Gesamtzeitraum kein Trend zu größeren Abweichungen vom Wahlergebnis feststellbar. Erwartungsgemäß treffen die Umfrageergebnisse, die in geringerem Abstand zur Wahl veröffentlicht wurden, das Wahlergebnis besser als Umfragen in größerem Ab-

<sup>24</sup> Der vergleichsweise glatte Verlauf zwischen 2008 und 2012 ist v.a. durch die höhere Zahl publizierter Umfragen in diesem Zeitraum (69 der insgesamt 160) begründet, so dass einzelne „Ausreißer“ weniger ins Gewicht fallen als vor oder nach diesem Zeitraum.

stand zur Wahl (Abb. 3). Dies ist nachweislich ein Ergebnis des längeren Beobachtungszeitraums von 21 anstatt zehn Tagen.<sup>25</sup> Weder die Anzahl veröffentlichter Parteienteile noch die Kontrolle der Institute<sup>26</sup> kann die Varianz der Abweichungen vom Wahlergebnis signifikant reduzieren.

Nicht unerwartet sind Kontexteffekte nachweisbar. Brandenburg und Hamburg, wo das arithmetische Mittel der Abweichungen mit 1,9% bzw. 1,6% je Parteianteil am höchsten ist, unterscheiden sich vom allgemeinen Abweichungsmuster genauso signifikant wie Schleswig-Holstein, das mit durchschnittlich 0,4% Abweichung von jedem Parteianteil bislang die treffsichersten Umfragen aufweist.<sup>27</sup> Die großen Flächenländer Nordrhein-Westfalen, Bayern, Baden-Württemberg und Hessen weisen dagegen keine signifikanten Abweichungen vom Abweichungsmuster der sieben Länder insgesamt auf.

*Abbildung 3:* Mittlere Abweichung der Umfrageveröffentlichungen (Parteianteile) vom späteren Wahlergebnis bei Landtagswahlen seit 1995 nach Abstand zur Wahl (max. 21 Tage)



<sup>25</sup> Wird ein Zeitraum von zehn Tagen gewählt, werden beide Variablen (Modell 1) insignifikant.

<sup>26</sup> Es wurde nur für Institute mit mindestens zehn veröffentlichten Umfragen im Beobachtungszeitraum kontrolliert.

<sup>27</sup> Wird ein Zeitraum von zehn Tagen gewählt, wird der Dummy für Schleswig-Holstein insignifikant (Modell 4), die Effekte für Brandenburg und Hamburg bleiben erhalten.

## 5. Veränderte Rahmenbedingungen für repräsentative Wahlumfragen

Die Bedingungen, mit Hilfe von Stichproben repräsentative Erhebungen durchzuführen, ändern sich ständig, weil sich die Gesellschaft, die Kommunikationsstruktur und die technischen Möglichkeiten verändern, Menschen durch das Instrument der Umfrage zu erreichen.<sup>28</sup> Die zentrale Aufgabe der Umfrageinstitute ist es, diese Veränderungen genau zu beobachten, zu bewerten und sich ihnen anzupassen bzw. neue Wege zu finden, um ihr Hauptziel zu erreichen: die zügige Erhebung valider und verlässlicher Daten unter ökonomisch vertretbarem Aufwand.

Neben den strukturellen Veränderungen der Gesellschaft in den letzten 30 Jahren, der deutlichen Ausdifferenzierung des Parteiensystems und der steigenden Distanz der Bürger zu den Parteien, die auch in der sinkenden Wahlbeteiligung ihren Ausdruck findet, sind es vor allem die technischen Veränderungen in der Kommunikation, die es den Instituten erschweren, aus der jeweils definierten Grundgesamtheit eine repräsentative Stichprobe zu realisieren.

Als erstes ist hier die Nutzung von Handys (*cellphones*), und in den letzten Jahren die vermehrte ausschließliche Nutzung von Mobilfunktelefonen (*cellphones only*) bei der Kommunikation zwischen Menschen zu nennen. Für die USA berichtet Zukin<sup>29</sup>, ehemaliger Präsident der *American Association for Public Opinion Research*, dass der Anteil der ausschließlichen Handy-Nutzer von 6% im Jahre 2005 auf 43% in der ersten Hälfte des Jahres 2014 gestiegen sei. Weitere 17% benutzten „meistens“ ein Handy. Die Zahlen in Deutschland sind nicht so hoch, haben sich jedoch in den letzten Jahren vervielfacht (2003: 4%, 2008: 9%, 2014: 14%).<sup>30</sup>

Das bedeutet für die Institute, die ihre Daten über Telefoninterviews erheben, einen stark erhöhten Aufwand, denn sie müssen eine eigene Stichprobe aus dem Mobilfunknetz ziehen bzw. ihre Telefonstichprobe mit Handynummern ergänzen (*dual sampling frames*). Das Handynetz, also die Grundgesamtheit aus dem die Stichprobe gezogen wird, ist jedoch ungleich größer als das Telefonnetz. Verkürzt: die Kosten der Erhebung steigen beträchtlich bzw. die Qualität der erhobenen Daten ist in Gefahr. Unter diesem Kostendruck wird nach neuen Verfah-

<sup>28</sup> Etliche Hinweise zu den verschiedenen Herausforderungen für die Umfrageforschung finden sich in einigen Beiträgen zu *Aus Politik und Zeitgeschichte*, 2015, 43-45.

<sup>29</sup> Zukin, C.: What's the Matter With Polling?, New York Times, 20. Juni 2015, <http://nyti.ms/1GyPSmU>.

<sup>30</sup> Vgl. hierzu Hunsicker, S./Schroth, Y.: Dual-Frame Ansatz in politischen Umfragen, Arbeitspapiere der Forschungsgruppe Wahlen e.V., 2 (2014), hier 9-10. Das Autorenteam kommt zum Schluss, dass die Datenqualität (Repräsentativität) reiner Festnetzstichproben in Deutschland (noch) so hoch sei, dass man auf den Dual-Frame Ansatz, der bundesweit problemlos umsetzbar sei, bisher verzichten könne.

ren gesucht und da bieten sich Internet-Umfragen an. Es besteht allerdings kaum ein Zweifel daran, dass diese Verfahren bisher nicht in der Lage sind, die geforderte Repräsentativität zu gewährleisten und deshalb kaum geeignet sein dürften, repräsentative Umfragen im Bereich der empirischen Wahlforschung durchzuführen.<sup>31</sup> Trotzdem haben sich bei den Wahlen in Großbritannien, Israel und auch den *midterm elections* 2014 in den USA, die alle wegen ihrer „Fehlprognosen“ besonders in der Kritik stehen, nach den Beobachtungen von Zukin viele neue Institute mit derartigen nichtrepräsentativen, telefonischen Befragungen am Wettkampf der Institute vor den Wahlen beteiligt, was eine Erklärung für die vermeintlich sinkende Prognosequalität von Umfragen darstellen könnte.

Ein zweites Problem der Umfrageinstitute sind die weiter steigenden Verweigerungs-raten (*unit nonresponse*). Auch hier berichtet Zukin von dramatisch niedrigen Realisierungsquoten bei seriösen Instituten, wie etwa *Pew Research*,<sup>32</sup> die für die Repräsentativität einer Umfrage existentiell sein können. In Deutschland ist der Rückgang der Antwortquoten nicht ganz so dramatisch, doch inzwischen liegen die Antwortraten auch hier unter 30%.<sup>33</sup> Das eigentlich Erstaunliche dabei ist, dass selbst unter diesen widrigen Umständen immer noch Umfragen realisiert werden können, die repräsentativen Charakter haben und verlässliche Aussagen zulassen. Vergleiche mit den Daten des Zensus oder Mikrozensus ermöglichen eine solche Einschätzung. Trotzdem geht von niedrigen Antwortraten eine große Gefahr für die Treffsicherheit von Umfragen in der besonders labilen Zeitspanne unmittelbar vor Wahlen aus. Ein Trend zu weniger genauen Wahlvoraussagen auf der Basis von Umfragen wäre deshalb nicht erstaunlich.

Ein weiteres Problem für die Genauigkeit von Umfragen vor dem Wahlgang ist die überschätzte Wahlbeteiligung. Dieses Phänomen wird schon länger beobachtet, kann aber selbst durch Gewichtungsprozeduren, die Erfahrungen aus früheren Abweichungen berücksichtigen, nicht hinreichend korrigiert werden, da die Beteiligungsquoten weiter abnehmen. Durch erweiterte Fragenkataloge lässt sich das Problem zwar eingrenzen, aber diese Vorgehensweise kostet Fragebogenplatz, den nicht alle Institute oder Auftraggeber bereit sind zu investieren.

<sup>31</sup> Dies bedeutet freilich nicht, dass sich das Internet nicht für Hypothesentests der empirischen Wahlforschung eignet.

<sup>32</sup> In den 1970er Jahren lag die Realisierungsquote bei Telefoninterviews in den USA bei ca. 80%, 1997 war sie auf 36%, 2014 auf 8% (sic!) gesunken.

<sup>33</sup> In Deutschland gibt es wenige Veröffentlichungen zur *Response Rate*, der Realisierungsquote bei Stichproben, weil die Institute diese Zahlen als Betriebsgeheimnisse betrachten. Eine Orientierung erhält man in den Beiträgen zur Gesundheitsberichterstattung des Bundes durch das Robert Koch Institut in Berlin. Dort wird berichtet, dass die Studie GEDA 2010 (Gesundheit in Deutschland aktuell) noch eine Realisierungsquote von 29% hatte, die von 2014 nur noch von 22%.

Dies führt uns zu einem letzten Aspekt, der die Qualität von Umfragen zunehmend beeinflusst. Es ist die geringer werdende Bereitschaft der Medien, die inzwischen ja die Hauptauftraggeber politischer Umfrageforschung sind, seriöse Umfragen, die auch entsprechendes Geld kosten, zu finanzieren. Der Wunsch nach der schnellen und überraschenden Schlagzeile im Konkurrenzkampf der Medien untereinander ist oft wichtiger als die Datenqualität. Nur noch selten wird in den Medien eine Studie zu einer politischen Entwicklung oder einem politischen Problem, das die Menschen bewegt, dargestellt, vielmehr werden Einzelergebnisse und insbesondere solche der sogenannten Sonntagsfrage benutzt, um eine Argumentation zu stützen oder eine Strategie zu erläutern.<sup>34</sup> Diese Entwicklung in den Medien begünstigt die Erhebung von online-Befragungen, weil sie billig und schnell sind. Das Publikum wird den Unterschied zu seriösen Umfragen aber kaum feststellen können. Es liegt in der Verantwortung der qualitätsbewussten Medien, diese Entwicklung zu korrigieren.

## V. Schlussbetrachtung

Gegenstand dieses Beitrags war die Prognosefähigkeit deutscher Umfrageinstitute im Lichte der jüngsten Ereignisse vor den Wahlen zum britischen Unterhaus. Hatten die meisten Veröffentlichungen zumindest einen Gleichstand zwischen den Konservativen und *Labour* projiziert, so konnte der konservative Premierminister am Wahltag mit einem Vorsprung von 6,5 Prozentpunkten und einer absoluten Sitzmehrheit (331 von 650) die Wahl gewinnen. War diese Diskrepanz zwischen den letzten Umfrageveröffentlichungen und dem späteren Wahlergebnis eine Singularität oder Ausdruck eines Trends?

Nun gibt es zweifellos Faktoren, die eine möglichst exakte Wahlprognose in einem Land mit relativer Mehrheitswahl erschweren. Dennoch waren die Abweichungen bei dieser Wahl besonders groß. Auch in Deutschland gibt es immer wieder Wahlen, in deren Vorfeld die Projektion des späteren Wahlergebnisses weniger gut gelingt. Die Bundestagswahl 2005 war eine solche Wahl, charakterisiert durch eine nahezu einmalige Ausgangslage, in der eine Wiederwahl der Regierungskoalition von vorneherein aussichtslos erschien, doch am Ende die führende Regierungspartei (SPD) noch profitierte und sich aufgrund eines cleveren Wahlkampfes mit dem amtierenden Kanzler als Zugpferd noch in eine Große

<sup>34</sup> Vgl. Pörzgen, G.: Medien lieben Zahlen. Ein Report zum Verhältnis von Journalismus und Demoskopie, in: *Aus Politik und Zeitgeschichte*, 43-45 (2014), 32-38.

Koalition mit der CDU/CSU „retten“ konnte.<sup>35</sup> Es waren wie im Fall *Cameron* 2015, *Netanjahu* 2015, *Schröder* 2002 oder auch *Kohl* 1994 die amtierenden Regierungschefs, die auf der Zielgerade noch einen Sieg ihrer Regierungen sicherstellen konnten. Dieser Effekt zugunsten von Regierungsparteien ist bekannt,<sup>36</sup> nur greift er selten erst in den letzten Tagen vor einer Wahl. Es war ein regierungsspezifischer *last minute swing*, der offenbar auch die Unterhauswahl in Großbritannien 2015 entschied.

Eine derartige Deutung kann als ad hoc-Erklärung und damit auch als eine Art Ausrede verstanden werden, um vermeintliche Fehlleistungen der Umfrageinstitute zu kaschieren. In jedem Fall sind solche Wahlen für Umfrageinstitute und Wahlforscher unbefriedigend. Wie unsere systematische Analyse jedoch zeigt, sind große Abweichungen der veröffentlichten Parteianteile vom späteren Wahlergebnis selten: In der Regel treffen die Umfragen unmittelbar vor Wahlen die späteren Parteistärken gut. Dies ist angesichts der Herausforderungen, die Umfrageinstitute aufgrund der immer schwerer erreichbaren Grundgesamtheit, ihrer Zurückhaltung im Antwortverhalten, abnehmender Parteibindungen und einer schwindenden Wahlnorm (Wahlbeteiligung) ein bemerkenswertes Ergebnis unserer Analysen.

Wir haben im Zeitraum 1995 bis 2015 auch keinen Trend zu schlechteren Wahlprojektionen feststellen können.<sup>37</sup> Wie sich bei der Analyse der Landtagswahlen zeigte, spielen spezifische Kontexte eine gewisse Rolle für die Treffsicherheit. Darüber hinaus sind Umfragen, die näher am Wahltermin liegen, treffsicherer als Umfragen, die in größerem zeitlichen Abstand zur Wahl durchgeführt werden. Es konnten auch keine besonders „guten“ oder „schlechten“ Institute identifiziert werden, auch wenn es unter den Veröffentlichungen einige Ausreißer gab, die zumindest zu Nachfragen einladen. Mitunter dürfte es sich schlicht um Umfragen handeln, deren Datengrundlage schlechter war als die der anderen Institute.

Kann man also Entwarnung geben und „den“ Umfragen einfach „vertrauen“? Das wäre ein voreiliges Fazit. Umfragen basieren auf Stichproben und mit Stich-

<sup>35</sup> Wüst, A.M./Roth, D.: Schröder's Last Campaign: An Analysis of the 2005 Bundestag Election in Context, in: *German Politics*, 15/4 (2006), 439-459; Hilmer, R.: Vorwahlumfragen 2005 – ein Debakel? Ist der Ausgang von Wahlen noch vorhersehbar?, in: Gabriel, O.W. et al.: *Wahlen und Wähler. Analysen aus Anlass der Bundestagswahl 2005*, Wiesbaden, 2009, 103-116; Wüst, A.M.: Ein verschossener Elfmeter und die Verpflichtung zum Erfolg, in: *Einsichten und Perspektiven*, 3, 2005, 24-31.

<sup>36</sup> Roth, D.: Die Bundestagswahl 2005 ... und niemand ist zufrieden, in: Oberreuter, H. (Hg.): *Unentschieden: Die erzwungene Koalition*, München, 2009, 219-236, hier 225; Wüst, A.M./Roth, D., a.a.O., hier 457-458.

<sup>37</sup> Dies konstatiert auch die Meta-Analyse *Anne Jessens*: „Ein Trend ist nur bedingt auszumachen, vieles spricht indes für eine Phasenbildung, bei der unklar ist, in welcher sich die politische Meinungsforschung zur Zeit befindet“ (*Jessen, A.*, a.a.O., hier 52).



proben verbunden sind statistische Fehlerbereiche, die auch zu größeren Abweichungen vom späteren Wahlergebnis führen können. Zum anderen gibt es Wahlen, bei denen das Instrument der Umfrage offenbar schlechter funktioniert. Es ist zumindest nicht bekannt, dass die Umfrageinstitute ihre Projektionsmodelle den Erfahrungen – zum Beispiel des Jahres 2005 in Deutschland – angepasst hätten. Sicher, es werden heute mehr Umfragen in unmittelbarer Nähe zur Wahl durchgeführt und veröffentlicht als noch vor 20 Jahren. Dadurch erhöht sich die Chance, das Einsetzen später Dynamiken vor einer Wahl noch zu erfassen. Wünschenswert wäre indes, dass sich auch die akademische Wahlforschung stärker an Projektionen und deren Modellierung beteiligt. „Erklärung und Prognose sind strukturell identisch, d.h. ihre (deduktive) Logik unterscheidet sich nicht“, und daher ist auch die Prognose ein Gütekriterium sozialwissenschaftlicher Forschung.<sup>38</sup> Viel zu häufig werden Forscher, die sich diesem Thema widmen, insgeheim belächelt. Manche Kollegen hoffen dann eher, dass „das Modell“ daneben liegt. Sie sollten sich eher darum bemühen, einen eigenen Beitrag zur Prognoseleistung von Wahlumfragen zu liefern, anstatt nach der Wahl unter vorgehaltener Hand zu beteuern, das Ergebnis eigentlich schon vor der Wahl „gewusst“ zu haben. Es ist noch Raum für neue Prognosemodelle. Aber Vorsicht: Risiko! Man kann auch falsifiziert werden.

Anhang

Tabelle 9: 18-Uhr Prognosen an Bundeswahltagen im Fernsehen

Jahr	ARD	RTL	ZDF
1998	0,43 (1,9)	0,43 (1,2)	0,13 (0,3)
2002	0,63 (1,5)	0,17 (0,5)	0,33 (0,5)
2005	0,47 (1,2)	0,63 (1,0)	0,83 (1,8)
2009	0,33 (0,6)	0,20 (0,5)	0,53 (1,1)
2013	0,29 (0,5)	0,23 (0,5)	0,54 (1,1)

Mittelwert der Abweichungen über alle Parteien (einschl. „Sonstige“); in Klammern: jeweils höchste Abweichung einer Partei

Quelle: Programmaufzeichnungen der Forschungsgruppe Wahlen e.V., Mannheim.

<sup>38</sup> Schnell, R./Hill, P.B./Esser, E.: Methoden der empirischen Sozialforschung, München, 2008, hier 65.

Wie in Großbritannien, sind die von den deutschen Fernsehanstalten veröffentlichten Prognosen bei Schließung der Wahllokale (18:00 Uhr) von einer hohen Präzision. Die Mittelwerte der Abweichungen sind deutlich niedriger als die bei Umfragen vor der Wahl. Die 18:00 Uhr Prognosen werden auf der Basis von *exit polls* erstellt. Die Stichproben hierzu werden auf ähnliche oder dieselbe Weise gezogen wie die für die Hochrechnungen am Wahlabend. Bei den Daten handelt sich um Ergebnisse von Befragungen der Wähler in ausgewählten Wahllokalen unmittelbar nach der Stimmabgabe.

Die erhobenen Daten sind – wie dargestellt – qualitativ hochwertiger als Umfragedaten und eignen sich in begrenztem Umfang auch für Analysen der Wahl, also der Beantwortung der Fragen: Welche soziodemografischen Gruppen haben wie gewählt? Und warum wurde eine bestimmte Partei gewählt? Wahlverhalten nach Alter, Geschlecht und der Zugehörigkeit zu sozialen Gruppen kann dargestellt werden. Bei engen Wahlen sind die Vergleiche zwischen den Veröffentlichungen der Fernsehsender zumindest für kurze Zeit von hohem Publikumsinteresse.

In der öffentlichen Debatte tauchen diese Prognosen kaum auf, wohl aber sind sie im Wettkampf der Sender untereinander wichtige „Qualitätskriterien“. Sie sind ein Spannungselement des Wahlabends.<sup>39</sup> Wegen der hohen Treffsicherheit ermöglichen die Prognosen frühe Interviews von Politikern durch die Fernsehjournalisten.

<sup>39</sup> Seibert, S.: Diesmal wirklich schöne Bilder ... ein Wahlwochenende mit Dieter Roth, in: Wüst, A.M. (Hg.): Politbarometer, Opladen, 2003, 345-348.

## Autoren / Authors

Prof. Dr. h.c. *Klaus-Dieter Lehmann*  
Präsident des Goethe-Instituts  
Dachauer Straße 122  
80637 München  
praesident@goethe.de

Prof. Dr. *Dirk Meyer*  
Institut für Volkswirtschaftslehre  
Lehrstuhl für Ordnungsökonomik  
Helmut-Schmidt-Universität  
der Bundeswehr Hamburg  
Holstenhofweg 85  
22043 Hamburg  
dirk.meyer@hsu-hh.de

Prof. Dr. *Werner J. Patzelt*  
Institut für Politikwissenschaft  
Lehrstuhl für Politische Systeme und  
Systemvergleich  
Bergstraße 53  
01069 Dresden  
werner.patzelt@tu-dresden.de

Prof. *Bert A. Rockman*, PhD  
Department of Political Science  
Purdue University  
2116B Beering Hall  
100 N. University Street  
West Lafayette, IN 47907-2098  
barockma@purdue.edu

Prof. Dr. *Dieter Roth*  
Institut für Politische Wissenschaft  
Universität Heidelberg  
Bergheimer Straße 58  
69115 Heidelberg  
dieter.roth@urz.uni-heidelberg.de

Dr. *Henrik Scheller*  
Politik und Regieren in Deutschland  
und Europa  
Universität Potsdam  
August-Bebel-Straße 89  
14482 Potsdam  
henrik.scheller@uni-potsdam.de

Dr. *Andreas M. Wüst*  
External Fellow  
Mannheimer Zentrum für Europäische  
Sozialforschung  
Universität Mannheim  
A5, 6, 68159 Mannheim  
andreas.wuest@mzes.uni-mannheim.de

# Autorenhinweise

Manuskripte von Aufsätzen und Berichten können in deutscher, englischer oder französischer Sprache eingereicht werden. Sie sollten der Redaktion *in dreifacher Ausfertigung* (ergänzt um eine Datei im Word-Format) zugehen und, zur Gewährleistung einer konzentrierten Diskussion, nicht mehr als 20-25 Manuskriptseiten (etwa 50.000 Zeichen einschließlich Fußnoten und Leerzeichen) umfassen. Jeder der ZSE angebotene Beitrag unterliegt einem zweifachen externen Begutachtungsprozess (Review-Verfahren).

Verlag und Redaktion übernehmen keinerlei Haftung für Manuskripte, die unverlangt eingereicht werden; sie werden zurückgeschickt, wenn Rückporto beigelegt ist. Die Annahme von Manuskripten setzt voraus, dass diese nicht gleichzeitig bei anderen Zeitschriften zur Begutachtung eingereicht oder anderwärtig publiziert werden. Bei Annahme eines Beitrags überträgt der Autor dem Verlag sämtliche Rechte zur Veröffentlichung.

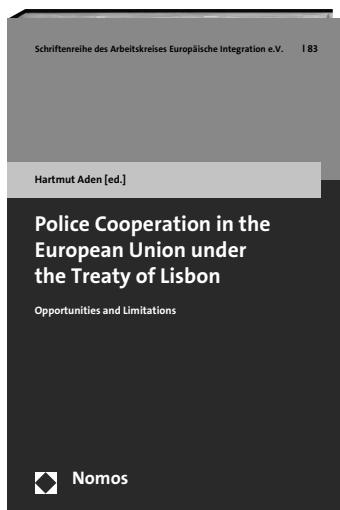
Hinweise zur formalen Gestaltung der Manuskripte:

- Die Autoren werden gebeten, durchgängig die *neue Rechtschreibung* zu verwenden.
- Für alle Beiträge wird eine *deutsch- und englischsprachige Zusammenfassung (Abstract)* von in der Regel nicht mehr als 12 Zeilen (à 60 Anschläge) erbeten, die dem Beitrag vorangestellt und auch im Internet veröffentlicht wird.
- Die *Gliederung* eines Manuskripts sollte durch römische Ziffern (I., II., III., usw.), arabische Ziffern (1., 2., 3., usw.) sowie durch Kleinbuchstaben (mit halber Klammer) bis höchstens aa), bb) erfolgen. Die beiden erstgenannten Kategorien erhalten eine eigene Überschrift in Fettdruck.
- *Hervorhebungen* im Text werden nur durch *Kursivschrift* gekennzeichnet. *Autoren-namen* sind durchgängig kursiv zu schreiben, sowohl im Text als auch in Fußnoten. Mehrautorenwerke werden durch Schrägstrich (ohne Leerzeichen) getrennt; Herausbernamen erscheinen in Normalschrift.
- *Fundstellennachweise* sind ausschließlich in *fortlaufend durchnummerierten Fußnoten* auszuweisen. Mehrere Literaturangaben bzw. Quellen in einer Fußnote werden durch Strichpunkt getrennt; am Ende jeder Fußnote steht ein Punkt. Gesetze und vergleichbare Rechtsnormen werden im *Fließtext in runden Klammern* nachgewiesen.
- *Literaturangaben* werden folgendermaßen zitiert: *Autor (Nachname, abgekürzter Vorname)*: Titel, Erscheinungsort, Erscheinungsjahr. Beispiel: *Milward, A.S.*: The European Rescue of the Nation-State, 2<sup>nd</sup> ed., London u.a., 2000.  
Bei *Beiträgen in Zeitschriften und Sammelwerken* sind die entsprechenden Quellenangaben (Herausgeber, Titel, Seitenangaben, ggf. Seitenzahl bei direktem Zitat) in Normalschrift wie folgt hinzuzufügen: *Grimm, D.*: Vertrag oder Verfassung, in: Staatswissenschaften und Staatspraxis, 6/4 (1995), 509–531; *Schmidt, M.G.*: Die Europäisierung öffentlicher Aufgaben, in: Ellwein, T./Holtmann, E. (Hg.): 50 Jahre Bundesrepublik Deutschland, Opladen, 1999, 385–394, hier 390.  
Wird ein *Titel mehrfach* zitiert, so gilt ab dem zweiten Beleg: *Schmidt, M.G.*, a.a.O., 392. Werden *mehrere Titel desselben Autors* zitiert, so ist der jeweilige Titel in *Kurzform* wie folgt mit anzugeben: *Milward, A.S.*: The European Rescue, a.a.O., 24.  
*Untertitel* werden nur angegeben, wenn andernfalls nicht auf den Inhalt der Quelle geschlossen werden kann.  
*Absätze von Paragraphen* sind mit „Abs.“ und arabischer Ziffer zu bezeichnen: Art. 53 Abs. 3 GG.
- *Tabellen und Schaubilder* sind mit durchnummerierten *Überschriften (in Kursivschrift)* zu versehen. Quellenangaben werden am Fuß der Tabelle angefügt.

An English version is available on special request.

# Polizei Kooperation in der EU:

Kontinuität und Wandel seit dem Vertrag von Lissabon



## Police Cooperation in the European Union under the Treaty of Lisbon

Opportunities and Limitations

Herausgegeben von Prof. Dr. Hartmut Aden

2015, 266 S., brosch., 54,- €

ISBN 978-3-8487-0843-7

(Schriftenreihe des Arbeitskreises Europäische Integration e.V., Bd. 83)

[www.nomos-shop.de/21676](http://www.nomos-shop.de/21676)

Die Polizei Kooperation in der EU war lange Zeit intergouvernemental organisiert. Durch den Vertrag von Lissabon (2009) ist sie mit der gemeinsamen Justiz- und Innenpolitik voll in die Europäische Union integriert worden.

Was hat sich seither verändert, welche Probleme bestehen fort? Welche praktischen Auswirkungen haben die neuen Regeln des Ver-

trages von Lissabon? Wie nutzt das Europäische Parlament seinen formal gewachsenen Einfluss? Welche Rolle spielen Menschenrechte und Datenschutz? Die Beiträge aus Wissenschaft und Praxis analysieren Kontinuität und Wandel der Polizei Kooperation in der EU aus den Perspektiven von Politik, Recht und Praxis.

Bestellen Sie jetzt telefonisch unter 07221/2104-37.

Portofreie Buch-Bestellungen unter [www.nomos-shop.de](http://www.nomos-shop.de)

Alle Preise inkl. Mehrwertsteuer



# Nomos

# „More Economic Approach“ im europäischen Wettbewerbsrecht



## Die rechtsstaatlichen Grenzen des 'more economic approach' im Lichte der europäischen Rechtsprechung

Eine Untersuchung des Handelns der Kommission auf dem Gebiet des europäischen Wettbewerbsrechts am Maßstab übergeordneter Vertrags- und Verfassungsgrundsätze

Von Dr. Jan Wolters

2015, 371 S., brosch., 99,- €

ISBN 978-3-8487-1736-1

(Schriftenreihe Europäisches Recht, Politik und Wirtschaft, Bd. 383)

[www.nomos-shop.de/24222](http://www.nomos-shop.de/24222)

Die Arbeit liefert eine detaillierte Untersuchung der EuGH/EuG-Rechtsprechung zu den Anforderungen, die die Gerichte an die Vereinbarkeit des more economic approach im europäischen Wettbewerbsrecht mit grundlegenden rechtsstaatlichen Prinzipien der Europäischen Union stellen.

Zahlreiche aktuelle Entscheidungen illustrieren, in welchem Umfang die Gerichte der Kommission in ihrem Bestreben nach einer verstärkten ökonomischen Betrachtung in der Entscheidungspraxis folgen. Die Untersuchung bietet wertvolle Orientierung bei der Einschätzung des künftigen Spielraums eines derartigen Ansatzes.

Bestellen Sie jetzt telefonisch unter 07221/2104-37.  
Portofreie Buch-Bestellungen unter [www.nomos-shop.de](http://www.nomos-shop.de)

Alle Preise inkl. Mehrwertsteuer



# Nomos

# Agrarhandelspolitik der EU



## Agrarinteressen als Verhandlungsmasse

Die Handelspolitik der Europäischen Union zwischen nationalen Präferenzen und internationalen Zwängen

Von Dr. Dieter Konold

2015, 382 S., brosch., 69,– €

ISBN 978-3-8487-1726-2

(Internationale Politische Ökonomie, Bd. 14)

[www.nomos-shop.de/24214](http://www.nomos-shop.de/24214)

Die Handelspolitik der EU ist – oder scheint – widersprüchlich. Einerseits setzt sich die Gemeinschaft für eine liberale Weltwirtschaftsordnung ein, andererseits schützt sie ihren Agrarsektor. Gemeinhin wird der Sonderstatus der Landwirtschaft auf die Stärke der „Bauernlobby“ zurückgeführt. Eine eingehende Untersuchung der agrarpolitischen Entwicklungen auf europäischer Ebene sowie in Frankreich und Deutschland zeigt jedoch, dass die Bauernverbände nicht (mehr) über die Einflussressourcen verfügen, um das Politikfeld zu bestimmen.

Dieser Einflussverlust ging nicht mit einer entsprechenden Liberalisierung des Agrarsektors einher, weil der Fortbestand etablierter Strukturen nicht nur landwirtschaftlichen Interessen dient. Institutionelle Faktoren, budgetäre Erwägungen und gesellschaftliche Forderungen wirken Status quo-fördernd. In internationalen Handelsgesprächen bietet diese Konstellation den europäischen Unterhändlern die Möglichkeit, Agrarinteressen verhandlungstaktisch einzusetzen.

---

Bestellen Sie jetzt telefonisch unter 07221/2104-37.

Portofreie Buch-Bestellungen unter [www.nomos-shop.de](http://www.nomos-shop.de)

Alle Preise inkl. Mehrwertsteuer



## Nomos

# Friedenskonsolidierung als Aufgabe der Verfassungsgerichtsbarkeit



## Friedenskonsolidierung durch Verfassungsgerichtsbarkeit in Bosnien und Herzegowina

Lehren aus einem Versuchslabor der Internationalen Gemeinschaft

Von RA Dr. Christian Steiner

2015, 557 S., brosch., 89,— €

ISBN 978-3-8329-6305-7

(Schriftenreihe der Europäischen Akademie Bozen, Bereich »Minderheiten und Autonomien«, Bd. 28)

[www.nomos-shop.de/13280](http://www.nomos-shop.de/13280)

Friedenskonsolidierung ist auch Aufgabe der Verfassungsgerichtsbarkeit. Im Nachkriegsbosnien kommen Verfassungsgericht und Menschenrechtskammer dabei eine Schlüsselrolle zu. Auch 15 Jahre nach Ende des bewaffneten Konflikts ist der Fortbestand des 1995 in Dayton geschaffenen komplexen Staatsgebildes nicht gesichert. Die Innen- wie Außenpolitik bleibt – befördert von den alten Eliten – geprägt von der fortbestehenden ethnischen Spaltung.

Schlichtend, ausgleichend und legitimierend, mit einem Auge für die übergeordneten Ziele von Staat und Gesellschaft, wirken in diesem Kontext die gemischt national-international besetzten obersten Gerichte: Sie gewähren effektiven Grundrechtsschutz, fördern die nationale Integration im Vielvölkerstaat, schlichten Kompetenzstreitigkeiten in der Ethnokratie und legitimieren sowie kontrollieren internationale Intervention.

---

Bestellen Sie jetzt telefonisch unter 07221/2104-37.  
Portofreie Buch-Bestellungen unter [www.nomos-shop.de](http://www.nomos-shop.de)

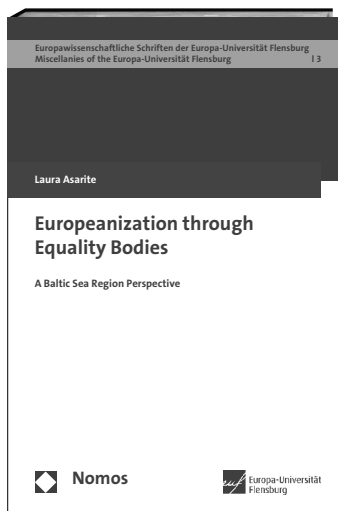
Alle Preise inkl. Mehrwertsteuer



**Nomos**



# Europäisierung durch Gleichstellung



## Europeanization through Equality Bodies

A Baltic Sea Region Perspective

Von Dr. Laura Asarite

2015, ca. 245 S., brosch., ca. 49,– €

ISBN 978-3-8487-2148-1

*(Europawissenschaftliche Schriften der Europa-Universität Flensburg | Miscellanies of the Europa-Universität Flensburg, Bd. 3)*

Erscheint ca. August 2015

[www.nomos-shop.de/24465](http://www.nomos-shop.de/24465)

Die Arbeit ist eine interdisziplinäre Studie über die Rolle der nationalen Gleichbehandlungsstellen (GBS) und des Europäischen Netzes der GBS in Europa. Sie gibt Aufschluss über den Zusammenhang zwischen der Vernetzung der GBS und der Europäisierung im Bereich der Förderung der Gleichstellung. Die Entwicklung eines gemeinsamen Ansatzes zur Förderung der Gleichstellung in der EU ist auf zwei Arten möglich – durch Rechtsharmonisierung und Europäisierung. Mit Hilfe der Untersuchung der Ostseeregion kann gezeigt werden, dass die

nationalen GBS und deren Vernetzung zur Europäisierung in der Gleichstellungsförderung führen können. Die Forschungsarbeit trägt dazu bei, eine bedeutende Lücke in der wissenschaftlichen Debatte über die GBS zu schließen und treibt die Debatte über Politiknetzwerke in der EU voran. Darüber hinaus wird belegt, dass der einzigartige Charakter der GBS und deren Vernetzung einen erheblichen Beitrag zu der Entwicklung fragmentierter, nationaler Ansätze hin zu einem gemeinsamen europäischen Konzept für die Gleichstellungsförderung liefern.

Bestellen Sie jetzt telefonisch unter 07221/2104-37.

**Portofreie Buch-Bestellungen unter [www.nomos-shop.de](http://www.nomos-shop.de)**

Alle Preise inkl. Mehrwertsteuer



# Nomos