

Zur Analyse von Wahlergebnissen in Parteihochburgen unter Berücksichtigung von Regressionsphänomenen

Methodological Contribution to the Analysis of Election Results in Party Strongholds

Thomas Ostermann und Rainer Lüdtke

Zusammenfassung

Regelmäßige Wahlen bilden ein Kernstück jeder demokratischen Verfassungsordnung. Insbesondere in den letzten drei Jahrzehnten hat sich die Wahlforschung im Bereich der politischen Wissenschaften mehr und mehr etabliert und spielt bei den Berichterstattungen sowohl im Vorfeld von Wahlen als auch in der Analyse von Wahlergebnissen eine entscheidende Rolle. Bei der Wahlanalyse sind vor allem die Änderungen in den sogenannten „Partei-Hochburgen“ von besonderem Interesse. Dieser Artikel soll anhand von Daten der Landtagswahl in Hessen 2008 sowie der Bundestagswahl 2005 klären helfen, ob es sich bei den Verlusten, die alle großen demokratischen Parteien in ihren Hochburgen erlitten haben, tatsächlich um beunruhigende Entwicklungen oder um ein Regressionsphänomen zur Mitte handelt. Entsprechende statistische Verfahren und Modelle werden in diesem Zusammenhang vorgestellt, angewandt und diskutiert.

Abstract

Regular elections are at the heart of every democratic constitutional order. Particularly in the last three decades electoral studies have increasingly established themselves within the field of political sciences and now play an important role in media and press coverage prior to elections and in the analysis of election results. In this analysis the results in party strongholds are of particular interest to researchers. Based on data of the election to the German State Parliament of Hesse in 2008 and of the election to the German Bundestag in 2005 this article aims at answering the question whether the losses of all big democratic parties in their strongholds are a serious and alarming matter or whether to a certain extent these losses can be explained by regression-to-the-mean. Relevant statistical methods and models are introduced, applied and discussed.

1 Einleitung

Wahlen stellen in der demokratischen Ordnung ein wichtiges und öffentlich hoch diskutiertes Instrument der politischen Willensbekundung der Bevölkerung dar. Freie Wahlen gelten als Qualitätsmerkmal demokratischer Verfahren in der Politik, mit denen dem Bürger ein Recht auf Mitbestimmung und Partizipation zur Bestimmung der Repräsentanten des Volkes geboten wird (Derichs et al. 2006). Regelmäßige Wahlen bilden daher ein Kernstück jeder demokratischen Verfassungsordnung. In der Bundesrepublik Deutschland wird dies in Artikel 28, Absatz 1 und Artikel 38, Absatz 1 des Grundgesetzes für die Wahlen zum Bundestag und zu den Vertretungen in Ländern, Kreisen und Gemeinden nach dem Prinzip der allgemeinen, gleichen, unmittelbaren, geheimen und freien Wahl geregelt, welche in regelmäßigen Abständen stattfinden.

Insbesondere in den letzten drei Jahrzehnten hat sich die Wahlforschung im Bereich der demoskopischen Wissenschaften mehr und mehr etabliert und spielt bei den Berichterstattungen sowohl im Vorfeld von Wahlen als auch in der Analyse von Wahlergebnissen eine entscheidende Rolle (Brehm 1999). Neben den obligatorischen Umfrage- und Hochrechnungsszenarien, die mittlerweile bereits bei kommunalen Wahlen eingesetzt werden, sind Wahlanalysen immer dann von besonderem parteiinternen, aber auch öffentlichen Interesse, wenn sich Ergebnisse der vorhergegangenen Wahl nicht wiederholen. Neben globalen Wählerwanderungen wird in diesem Zusammenhang auch analysiert, in welchen Bezirken die deutlichsten Verluste aufgetreten sind.

Hier sind die sogenannten „Partei-Hochburgen“ von besonderem Interesse, also diejenigen Gebiete, in denen eine Partei über einen längeren Zeitraum besonders hohe Stimmenanteile erhält. Nach Kirschey (2006) verfolgt die Analyse solcher Hochburgen das primäre Ziel aufzuzeigen, „wie das aktuelle Wahlergebnis in diesen Gebieten ausgefallen ist und ob bzw. welche Abweichungen zum Landesergebnis eingetreten sind.“ Dieses spielt auch in der öffentlichen Diskussion eine große Rolle. So hört man immer wieder die Behauptung, dass die eine oder andere Partei besonders in ihren Hochburgen „herbe Verluste“ hinnehmen musste, zuletzt z. B. in einem Bericht der Südhüringischen Zeitung (2008) zur Landtagswahl 2008 in Hessen.

Auf den ersten Blick erscheint der Sinn einer solchen Analyse unmittelbar evident. Wählerverluste in Gebieten, in denen eine Partei bisher besonders viele Stimmen sammeln konnte, müssen per se beunruhigend sein. Aus wissenschaftlich-statistischer Sicht gilt dies nicht uneingeschränkt, es ist durchaus möglich, dass es sich bei den besonders herben Verlusten in Hochburgen um ein rein technisches Phänomen handelt, das als „Regression zur Mitte“ bekannt ist.

Dieser Artikel soll anhand von Daten der Landtagswahlen in Hessen 2008 sowie der Bundestagswahl 2005 klären helfen, ob es sich bei den Verlusten, die alle großen demokratischen Parteien in ihren Hochburgen erlitten haben, tatsächlich um beunruhigende Entwicklungen oder um ein Regressionsphänomen handelt.

Ziel ist es dabei nicht, sich fokussierend auf diese beiden Wahlen zu beschränken. Vielmehr soll exemplarisch dargelegt werden, wie eine Analyse von Wahlhochburgen adäquat unter Berücksichtigung statistischer Sondereffekte erfolgen kann.

2 Regression zur Mitte

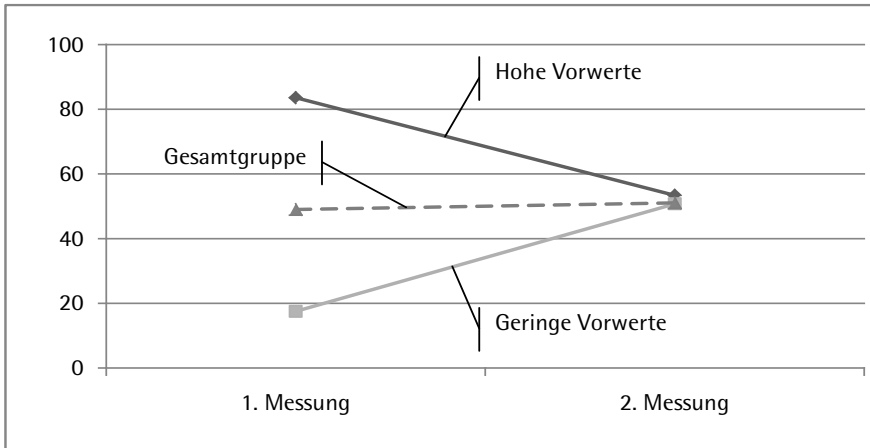
2.1 Beschreibung des Phänomens

Regression-zur-Mitte (regression-to-the-mean, RTM) ist ein rein statistisches Phänomen, das bei Wiederholungsmessungen auftritt, die an selektierten Populationen vorgenommen werden. Das Phänomen wurde erstmals von Galton (1887) bei der Analyse von Körpergrößen von Nachkommen auffällig großer bzw. kleiner Väter beschrieben. Demzufolge haben besonders große Väter tendenziell kleinere Söhne, während besonders kleine Väter Söhne bekommen, die größer sind als sie selbst.

Wie man heutzutage weiß, liegt dieser Beobachtung ein einfaches Gesetz zugrunde, das völlig unabhängig vom inhaltlichen Zusammenhang auftritt. Entscheidend ist der selektive Blick, mit dem die Daten analysiert werden. Man behält nicht mehr das Ganze im Auge sondern konzentriert sich ausschließlich auf eine Gruppe von Beobachtungen, die in ihrer Ausprägung besonders extrem sind, also besonders klein oder besonders groß. Ist man nun an der zeitlichen Fortentwicklung dieser extremen Gruppe interessiert (und blendet daher die „durchschnittlichen“ Beobachtungen aus), so werden die Folgemessungen in dieser Subgruppe im Durchschnitt näher am Mittelwert der Gesamtgruppe liegen.

Dieses lässt sich einfach mithilfe von zwei Reihen von Zufallszahlen zwischen 0 und 100 erzeugen. Abbildung 1 zeigt bspw. ein Szenario, in dem auf zwei Subgruppen selektiert wurde, zum einen auf die Gruppe von Beobachtungen, bei denen bei der ersten Messung Werte über 70 gemessen wurden (extrem hohe Vorwerte), zum anderen die Gruppe mit Erstmessung unter 30 (extrem geringe Vorwerte). Dargestellt sind die Mittelwerte der beiden Zufallszahlenreihen für diese Subgruppen sowie die Mittelwerte der Gesamtgruppe aller Zahlen. Es ist deutlich zu sehen, dass die Mittelwerte der Wiederholungsmessung näher am Mittelwert der Gesamtpopulation liegen als die der Erstmessung (Lüdtke/Ostermann 2005).

Abbildung 1 Regression zur Mitte am Beispiel zweier Zufallszahlenreihen zwischen 0 und 100



2.2 Mathematischer Hintergrund

Mathematisch lässt sich dieser Sachverhalt über eine bivariate Verteilung beschreiben. Im obigen, künstlichen Beispiel wäre eine bivariate Gleichverteilung angebracht, aus mathematischen Gründen wird üblicherweise aber eine bivariate Normalverteilung angenommen. Bezeichnet man mit X die Erstmessung eines Merkmals, so lässt sich die Zweitmessung Y über eine lineare Regressionsgleichung auf die Erstmessung zurückführen:

$$Y - \mu = \rho(X - \mu) + \varepsilon, \quad \varepsilon \approx N(0, \sigma^2(1 - \rho^2)) \quad (1)$$

Dabei bezeichnet μ den Mittelwert der Verteilung (der bei beiden Messungen ja gleich ist), σ^2 deren Varianz und ρ die Korrelation zwischen beiden Messungen, die üblicherweise zwischen 0 und 1 liegt; ε ist ein normalverteilter Fehler mit Erwartungswert 0 und Varianz. Damit ergibt sich für den bedingten Erwartungswert von Y bei gegebenem $X = x$

$$Y - \mu = \rho(X - \mu) + \varepsilon, \quad \varepsilon \approx N(0, \sigma^2(1 - \rho^2)) \quad (2)$$

Dieser unterscheidet sich von x um

$$\Delta = x - E(Y | X = x) = x - [(1 - \rho)\mu + \rho x] = (1 - \rho)(x - \mu) \quad (3)$$

woraus sich unmittelbar der RTM-Effekt ablesen lässt: Ist x größer als μ , so ist Δ größer als 0, also x größer als der zu erwartende Wert bei der Zweitmessung; falls $x < \mu$ folgt genau das Gegenteil. Dieser Effekt ist um so größer, erstens, je größer der tatsächlich gemessene Wert x ist und zweitens, je geringer die Korrelation ρ zwischen X und Y ist.

Dieses gilt auch dann noch, wenn man die Voraussetzung aufgibt, dass die Verteilung zu beiden Messzeitpunkten stabil ist, d. h. der Mittelwert μ bei Erst- und Zweitmessung gleich ist. Bezeichnet man mit τ die Veränderung, die allen Messwerten gleichartig widerfährt (also etwa landesweite Verluste bzw. Gewinne einer Partei), so folgt aus

$$Y - \mu = \tau + \rho(X - \mu) + \varepsilon, \quad \varepsilon \approx N(0, \sigma^2(1 - \rho^2)) \quad (4)$$

$$E(Y | X = x) = (1 - \rho)\mu + \tau + \rho x \quad (5)$$

dass für besonders große x der erwartete Gewinn kleiner ist als der durchschnittliche Gewinn, bzw. der erwartete Verlust größer ausfällt. Umgekehrt wird für kleine x ein größerer Gewinn bzw. kleinerer Verlust erwartet.

Gleichung (4) kann dazu benutzt werden, einen statistischen Test zu konstruieren, der den Effekt einer Intervention bewertet, die nur in einer selektierten Gruppe von besonders auffälligen Merkmalsträgern (z. B. Wahlhochburgen) durchgeführt wird. Dazu ist es lediglich notwendig, den wahren Wert von μ zu kennen. Dieses ist letztendlich der t-Test innerhalb des linearen Regressionsmodells von $Y - \mu$ auf $X - \mu$, ob der Achsenabschnitt (hier τ) von 0 verschieden ist oder nicht. Die Teststatistik t ist gegeben durch

$$t = \frac{\hat{\tau}}{\sqrt{s^2 \left(\frac{1}{n} + \frac{\bar{X}^2}{\sum_{i=1}^n (X_i - \bar{X})^2} \right)}} \quad (6)$$

wobei $\hat{\tau}$ den Schätzer des Achsenabschnitts und s^2 den Mean Squared Error (MSE) aus der Regressionsanalyse bezeichnet. X_i sind die Werte der Erstmessung des i -ten Merkmalsträgers aus der selektierten Gruppe und \bar{X} ihr Mittelwert, $i=1, \dots, n$. t ist gemäß einer t-Verteilung mit $(n-2)$ Freiheitsgraden verteilt.

Eine ausführliche Darstellung und Herleitung des Tests findet sich bei Mee und Chua (1991). Dort findet sich auch ein illustratives Beispiel zu acht Studenten, die bei einer Prüfung für einen Highschool-Abschluss mit einem durchschnittlichen Punktwert von 57,4 durchfielen. Nach einem speziell ausgerichteten Auffrischkurs stieg die Prüfungsleistung auf 60,4 Punkte an. Auf den ersten Blick ist dies ein klares Indiz dafür, dass der Auffrischkurs erfolgreich war; eine Interpretation, die vor allem auch durch das Ergebnis eines verbundenen t-Tests mit einem einseitigen p-Wert von $p=0,0428$ gestützt wird. Diese Sichtweise berücksichtigt aber nicht, dass es sich bei den Prüflingen ausschließlich um diejenigen Studenten handelte, die in der ersten Prüfung besonders schlecht abgeschnitten hatten und damit einer selektierten Gruppe angehörten. RTM-Effekte, und damit verbesserte Prüfungsergebnisse in der Wiederholungsprüfung, waren daher zu erwarten. Statt eines verbundenen t-Tests wäre daher ein korrigierter t-Test gemäß Gleichung (6) angebracht gewesen. Tabelle 1 zeigt die Einzeldaten der acht Studenten. Nimmt man nun wie Mee und Chua an, dass das mittlere Prüfungsergebnis aller Studenten bei der Erstprüfung bei $\mu=75$ Punkten liegt, dann lässt sich der korrigierte t-Test mit jedem statistischen Programmpaket leicht berechnen, indem man zunächst von allen Einzelwerten (bei Erst- und Wiederholungsprüfung) 75 abzieht und dann eine einfache lineare Regressionsanalyse durchführt. Im Falle der obigen acht Studenten ergibt sich ein t-Wert von $t=1,08$ und demzufolge ein (einseitiger) p-Wert von $p=0,16$, also kein statistisch signifikantes Ergebnis, und damit kein Indiz dafür, dass der Auffrischkurs den gewünschten Effekt hatte.

Tabelle 1 Prüfungsleistungen von acht Studenten, die bei einer Prüfung durchgefallen waren, und nach einem Ergänzungskurs eine Wiederholungsprüfung absolvierten

Student	Leistung bei Erstprüfung (Punkte)	Leistung bei Wiederholungsprüfung (Punkte)
1	45	49
2	52	50
3	63	70
4	68	71
5	57	53
6	55	61
7	60	62
8	59	67

2.3 Regressionseffekte im sozialwissenschaftlichen Kontext

Wie das in Kapitel 2.1 beschriebene Beispiel der Zufallszahlen zeigt, bedarf es keinerlei inhaltlich begründeter Mechanismen, um einen Regressionseffekt auszulösen. Entscheidend ist lediglich der selektive Blick auf die Daten. Damit ist das Phänomen universell und natürlich auf jede andere Datenerhebung im sozial- oder politikwissenschaftlichen Kontext anwendbar. Nachtigall und Suhl (2002) geben hierzu ein eindrucksvolles Beispiel: Nach Furby (1973) findet sich bei Kindern von besonders intelligenten Eltern eine durchschnittlich geringere Intelligenz. Da die Intelligenz eines Menschen positiv mit seinem sozioökonomischen Status korreliert, wurden diese Daten verschiedentlich dahingehend interpretiert, dass ein höher sozioökonomischer Status kausal eine (relative) Minderbegabung in der Folgegeneration bedingt; eine Interpretation, die den dahinter liegenden Regressionseffekt vollständig vernachlässigt und daher als nicht zwingend bewertet werden muss.

Bei der Analyse von Wahlergebnissen stellt eine Fokussierung auf Wahlhochburgen, zumindest wenn man diese nur auf Basis der Ergebnisse der vorhergehenden Wahl definiert, ebenfalls einen Selektionsprozess im obigen Sinne dar: Betrachtet werden dabei (ausschließlich) Messungen am oberen Rand der Verteilung, Messwerte aus der Mitte oder gar dem unteren Rand der Verteilung werden ausgeblendet oder zumindest weniger stark berücksichtigt. Dementsprechend kommt es zwangsläufig zu Regressionseffekten und damit möglicherweise zu Fehlinterpretationen der Daten.

3 Daten

Nach diesen Vorbemerkungen ist also zu erwarten, dass eine Partei in einem Gebiet, in dem sie ein besonders gutes Wahlergebnis erzielt hat, bei der Folgewahl überdurchschnittlich verliert (bzw. falls sie gewonnen hat, dass sie unterdurchschnittlich gewinnt). Nur wenn diese überdurchschnittlichen Verluste (bzw. unterdurchschnittlichen Gewinne) nicht durch Regressionseffekte zu erklären sind, hat dieses eine Bedeutung, die einer detaillierteren Ursachenanalyse bedarf.

Wir haben daher für CDU, SPD, FDP und Bündnis 90/die Grünen die Wahlergebnisse der hessischen Landtagswahlen vom 27.1.2008 mit denen vom 2.2.2003 verglichen. Datengrundlage war jeweils der Anteil der Landesstimmen in allen

55 hessischen Wahlkreisen, wie sie auf den entsprechenden Webseiten des Hessischen Statistischen Landesamtes veröffentlicht wurden.¹

Analoge Analysen können für die Bundestagswahlen vom 22.9.2002 und 18.9.2005 durchgeführt werden. Hier beziehen wir uns auf die Anteile der Zweitstimmen je Wahlkreis, entsprechend den Veröffentlichungen des Bundeswahlleiters.²

3.1 Analyse der hessischen Landtagswahlen

Abbildung 2 zeigt die Verluste bzw. Gewinne der vier großen Parteien je Wahlkreis bei den Landtagswahlen in Hessen 2008 im Vergleich zu den Stimmanteilen bei den Wahlen 2003. Für alle vier Parteien wird der RTM-Effekt durch die eingezeichnete Regressionsgerade eindrucksvoll belegt: je größer der Stimmanteil 2003 war desto geringer waren die Gewinne 2008 (SPD, FDP) bzw. desto größer die Verluste (CDU, Grüne).

Besonders deutlich wird das bei den Grünen. Sie hatten 2003 im Wahlkreis Frankfurt/Main V einen Stimmanteil von 26,8 % und verloren hier mit 9,8 Prozentpunkten überdurchschnittlich; landesweit betrug der Verlust nur 2,6 Prozentpunkte. Abbildung 2 (c) belegt nun, dass es sich dabei nicht etwa um (im statistischen Sinn) unerwartete Verluste handelte: genau dieser Wert war aufgrund des RTM-Effekts prognostizierbar. In ähnlicher Weise ist der unterdurchschnittliche Gewinn der FDP von 1,4 Prozentpunkten in ihrer Hochburg Hochtaunus II zu interpretieren (landesweite Gewinne 1,5 Prozentpunkte): Nicht der geringere Gewinn ist überraschend, sondern eher die Tatsache, dass der tatsächliche Gewinn nur so wenig unterhalb der landesweiten Ergebnisse lag (Abbildung 2 (d)).

Anders bei der CDU, die in Hessen landesweit durchschnittlich zwölf Prozentpunkte verlor. Ihre größten Verluste (18,7 % und 17,6 %) hatte sie genau in den Wahlkreisen Fulda I und Fulda II zu verzeichnen, bei denen sie 2003 die höchsten Stimmanteile hatte (jeweils 67,9 %). Abbildung 2 (a) bestätigt den RTM-Effekt zwar grundsätzlich, zeigt aber auch, dass aufgrund dieses Effekts in beiden Kreisen lediglich ein Verlust von etwa 15 Prozentpunkten zu erwarten war.

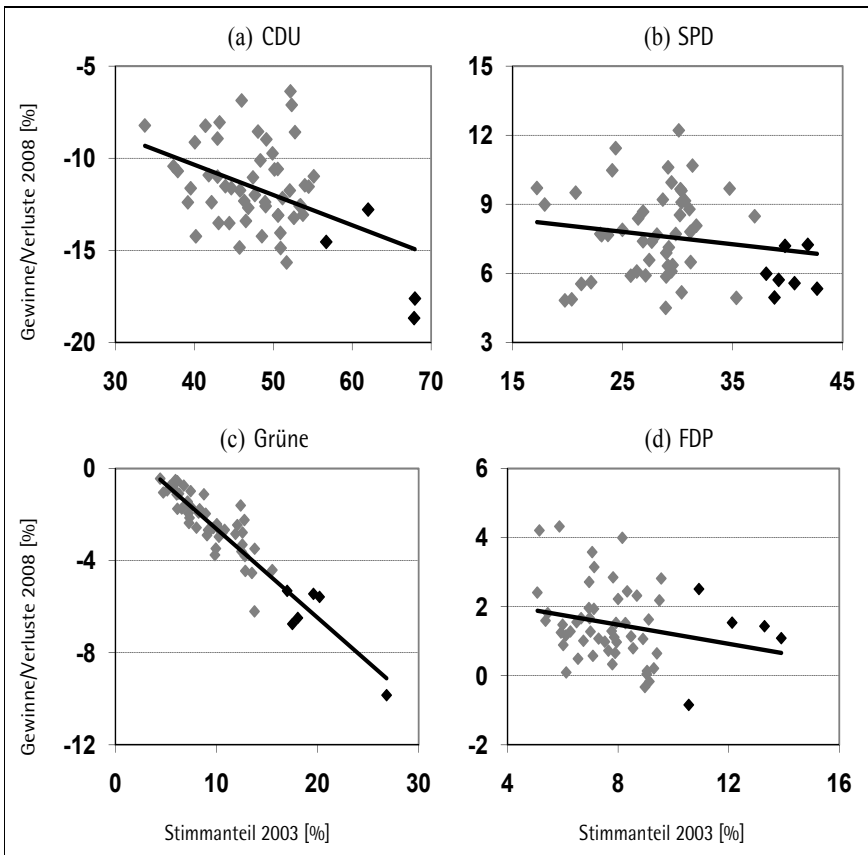
Nimmt man nun an, dass Parteien in ihren Wahlhochburgen einen anderen Wahlkampf führen als in den restlichen Wahlkreisen (z. B. weil man die dortigen Wähler besonders motivieren möchte oder man sich umgekehrt deren Stimmen sehr sicher ist und daher nur wenig Mittel einsetzt), so stellt sich unmittelbar die Frage, ob die gewählte Strategie erfolgreich war oder nicht. Mit anderen Worten: es ist zu über-

1 <http://www.statistik-hessen.de/subweb/ltw2003/endg/tabz999001.htm>;
<http://www.statistik-hessen.de/subweb/ltw2008/index.htm>.

2 http://www.bundeswahlleiter.de/de/bundestagswahlen/BTW_BUND_05/downloads/.

prüfen, ob sich das Ergebnis in den Wahlhochburgen signifikant vom Gesamttrend der Partei unterscheidet. Statistisch lässt sich das durch das Einfügen eines Interventionseffekts τ lösen, wie dieses in Gleichung (4) geschehen ist. Mit Gleichung (6) steht dann ein Test zur Verfügung, der genau diese Frage überprüft. Völlig falsch wäre es dagegen, einen einfachen verbundenen t-Test zu verwenden, da dieser den Selektionsmechanismus, nur extreme Werte zu analysieren, nicht berücksichtigt und daher Veränderungen zur Mitte ausschließlich auf den Interventionseffekt zurückführt.

Abbildung 2 Gewinne und Verluste der einzelnen Parteien je Wahlkreis bei der Landtagswahl in Hessen 2008 im Vergleich zur Wahl 2003



„Hochburgen“ schwarz gekennzeichnet. Zur Illustration haben wir alle diejenigen Wahlkreise als Hochburgen definiert, in denen die Partei 2003 deutlich über dem mittleren Wahlergebnis lag (in Abbildung 2 hervorgehoben), wobei die Definition von „deutlich“ allein auf der Basis des grafischen Eindrucks getroffen wurde.

Tabelle 2 Tests auf überdurchschnittliche Verluste (bzw. unterdurchschnittliche Gewinne) in auf der Basis von Stimmanteilen im Wahljahr 2003 willkürlich definierten Hochburgen im Vergleich der hessischen Landtagswahlen von 2003 und 2008

Partei	Definition Hochburg %	N	Wahl 2003 %	Wahl 2008 %	Test auf Interventionseffekt *
CDU	> 56	4	63,6	47,7	$t_{\text{verb.}}: p = 0,027$ $t_{\text{korr.}}: p = 0,345$
SPD	> 38	7	40,1	46,1	$t_{\text{verb.}}: p = 0,002$ $t_{\text{korr.}}: p = 0,812$
FDP	> 10	5	12,2	13,3	$t_{\text{verb.}}: p = 0,282$ $t_{\text{korr.}}: p = 0,697$
Grüne	> 16	6	19,9	13,3	$t_{\text{verb.}}: p = 0,001$ $t_{\text{korr.}}: p = 0,975$

* *p*-Werte sind einseitig.

$t_{\text{verb.}}$ – verbundener *t*-Test, $t_{\text{korr.}}$ – nach Mee & Chua korrigierter *t*-Test gemäß Gleichung (6).

Tabelle 2 zeigt, dass für keine Partei überdurchschnittliche Verluste in ihren jeweiligen Hochburgen reklamiert werden können, zumindest wenn man den um den RTM-Effekt korrigierten *t*-Test (Gleichung (6)) zugrunde legt. Mit anderen Worten: ein eventuell in diesen Hochburgen anders geführter Wahlkampf hätte keinen nachweisbaren Effekt gehabt. Ein Urteil, das völlig anders ausgesehen hätte, wenn der einfache verbundene *t*-Test als Entscheidungsgrundlage verwendet worden wäre. Hier hätten CDU, SPD und Grüne die in ihren Hochburgen erzielten Ergebnisse als auffällig beurteilt.

3.2 Analyse der Bundestagswahl 2005

Bei der Bundestagswahl 2005 zeigt sich ein ähnliches Ergebnis (Abbildung 3), was nach dem bisher Gesagten auch kaum anders zu erwarten war: In ihren jeweiligen Hochburgen gewinnen alle Parteien unterdurchschnittlich hinzu oder verlieren überdurchschnittlich. Aufgrund eines Sondereffekts muss man bei der Analyse allerdings eine Einschränkung machen: sie gilt für Westdeutschland mit Ausnahme des Saarlands. Dieser Sondereffekt liegt im Wahlergebnis der PDS begründet, die 2005 besonders zugewinnen konnte, ihre Gewinne aber überwiegend in Ostdeutschland und dem Saarland erzielte (hier trat der ehemalige SPD-Ministerpräsident Oskar Lafontaine für die PDS an). Diese Gewinne gingen fast ausschließlich zu Lasten der SPD, so dass deren Verluste hier besonders groß waren und aus dem Gesamtbild herausfallen.

Abbildung 3 Gewinne und Verluste der einzelnen Parteien je Wahlkreis in Westdeutschland (außer Saarland) bei der Bundestagswahl 2005 im Vergleich zur Wahl 2002

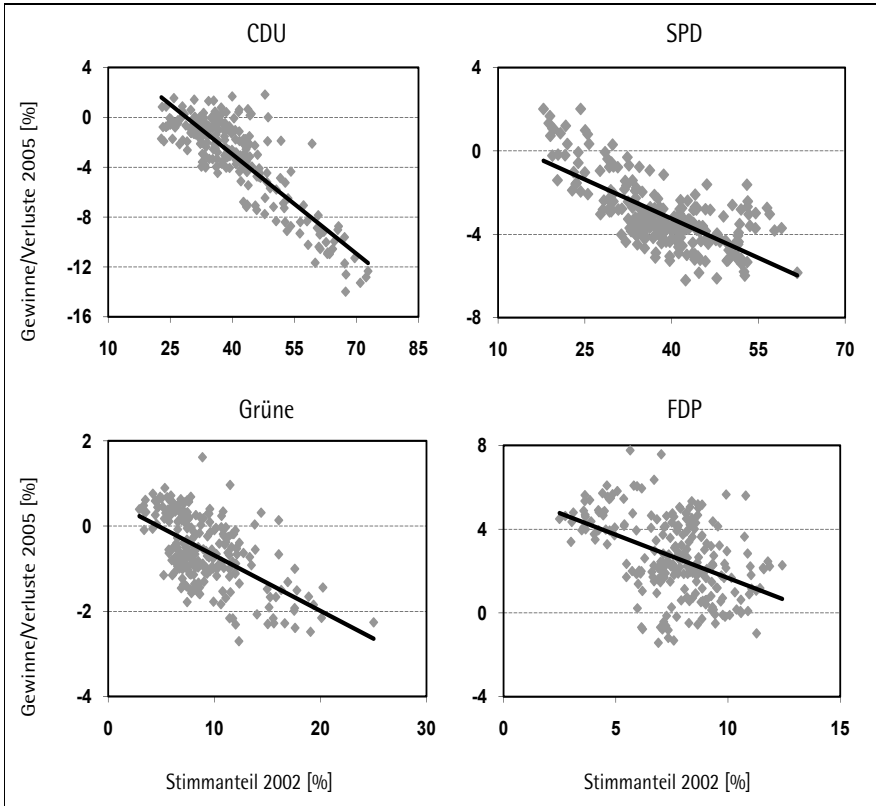


Tabelle 3 zeigt zudem, dass weder bei der CDU noch bei FDP oder Grünen Zusatz-effekte in den Hochburgen nachweisbar sind. Bei der SPD ist ein solcher allerdings mit dem korrigierten t-Test auch statistisch nachweisbar: die Partei verlor in ihren Hochburgen zwar leicht überdurchschnittlich, aber deutlich weniger als aufgrund des RTM-Effekts zu erwarten gewesen wäre.

Tabelle 3 Tests auf überdurchschnittliche Verluste (bzw. unterdurchschnittliche Gewinne) in willkürlich definierten Hochburgen im Vergleich der Bundestagswahlen von 2005 und 2002

Partei	Definition Hochburg %	N	Wahl 2002 %	Wahl 2005 %	Test auf Interventionseffekt*
CDU	> 67,0	8	69,4	57,1	$t_{\text{korr.}}: p = 0,244$
SPD ¹	> 55,0	5	58,4	54,6	$t_{\text{korr.}}: p = 0,004$
FDP	> 11,5	5	11,8	14,1	$t_{\text{korr.}}: p = 0,767$
Grüne	> 20,0	5	22,2	20,2	$t_{\text{korr.}}: p = 0,593$

* *p*-Werte sind einseitig, 1 nur Westdeutschland außer Saarland.

$t_{\text{korr.}}$ – nach Mee & Chua korrigierter *t*-Test.

4 Diskussion

In der Analyse von Wahlergebnissen spielen viele Effekte eine Rolle, bspw. der von Lazarsfeld et al. (1969) beschriebene Bandwagon-Effekt oder Mitläufer-Effekt (Anschluss an die Mehrheitsmeinung, z. B. um schließlich als Sieger dazustehen) und der Underdog-Effekt (Anschluss an die Minderheitsmeinung, z. B. aus Trotz), die speziell für Wahlhochburgen von Interesse sind (Schoen 2002). Aber auch Interventionen auf Seiten der Partei sind zu berücksichtigen: So ist z. B. denkbar, dass Parteien in Wahlkreisen, in denen sie in der Wahl zuvor sehr gut abgeschnitten haben, einen weniger intensiven Wahlkampf betreiben, bzw. der politische Gegner – auch wegen des Anreizes der Überhangmandate – besondere Anstrengungen unternimmt. Andererseits ist es auch möglich, dass Anhänger der Parteien, die zuvor sehr gut abgeschnitten haben, bei der kommenden Wahl aufgrund einer eher moderaten Werbung der Partei weniger mobilisiert werden.

Überlagert werden diese Effekte aber in jedem Fall vom reinen Regressionseffekt, auf den wir uns in der hier dargestellten Analyse beschränkt haben. Dieser Regressionseffekt ist ein wenig beachtetes statistisches Artefakt, das allein auf Selektionsbedingungen zurückzuführen ist und zunächst keine inhaltliche Entsprechung hat. Erst Abweichungen vom erwarteten Regressionseffekt können die Grundlage für eine inhaltliche Diskussion und Interpretation von Wahlergebnissen aus Wahlhochburgen sein. Auf diesem Wege ist es durchaus möglich, die oben genannten Effekte als „Interventionseffekte“ zu modellieren (obwohl ja im engeren Sinne keine Intervention vorliegt) und die von uns vorgeschlagene Technik, den korrigierten *t*-Test, anzuwenden.

Dabei ist zu beachten, dass es sich in unseren Beispielen um reine prä-post Szenarien ohne Baseline-Absicherung handelt. Bezogen auf die Parteihochburgen ist es daher unwahrscheinlich, dass der korrigierte Test auch dann die entsprechenden Resultate ergibt, wenn er auf Hochburgen angewandt wird, die konstant über mehrere Jahre von einer Partei dominiert wurden und plötzlich einbrechen. Für eine solche Entwicklung sind die Gründe eher inhaltlich, bspw. im Wechsel/Rücktritt eines Spitzenkandidaten zu suchen.

Weiterhin ist in der Interpretation unserer Ergebnisse zu berücksichtigen, dass es sich bei den verwandten Daten bereits um Aggregatdaten handelte. Diese sind sicher nicht ausreichend, um ein individuelles Wahlmodell zu testen, was aber auch nicht Intention der vorliegenden Arbeit war. Trotzdem könnte mit entsprechenden Daten möglicherweise noch deutlicher gezeigt werden, dass die größeren Verluste der Parteien in den Hochburgen der Wahl 2005 auf das Phänomen der Regression zur Mitte zurückzuführen sind und nicht auf tatsächlich andere Kalküle der Wähler. Ein Individualmodell könnte dann Aufschluss über die verbleibenden Effekte geben, wie diese in der Bundestagswahl 2005 bei der SPD auch bei Anwendung des korrigierten Tests deutlich hervortreten (siehe Tabelle 3).

Da der Status einer Wahlhochburg multifaktoriell begründet ist und je nach Partei unterschiedlich ausfällt, ist sowohl der zugrunde liegende Selektionsmechanismus für den Regressionseffekt als auch der Einfluss anderer Parameter nur schwer modellierbar. Dementsprechend und vor dem Hintergrund des Vorliegens von Daten aus der Grundgesamtheit der Wähler (einschließlich eines tatsächlichen Mittelwerts) bietet sich für dieses Szenario das Verfahren von Mee und Chua in der hier vorgestellten Form optimalerweise an.

Am Beispiel von Wahlergebnissen in Hochburgen konnte hier mit dem letztgenannten Verfahren eine solche Analyse dargestellt werden. Auch wenn in den vorgestellten Beispielen die Art der Wahlkampfstrategie in Hochburgen nicht explizit bekannt ist, so sprechen Arbeiten zur Wahlforschung dafür, dass Hochburgen von Parteien sowohl in der Phase des Wahlkampfes (Hoecker 2005) als auch in der Analyse des Wählerverhaltens (Schoon 2006; Probst 2007) eine besondere Bedeutung für die Parteien besitzen. Gerade in diesen Regionen sollen Stammwähler der Partei mobilisiert werden. Umgekehrt wird es in den entsprechenden Analysen immer wieder als bedeutsam hervorgehoben, wenn andere Parteien eine Hochburg für sich erobern.

Insgesamt hat die Auseinandersetzung mit diesem Phänomen der Regression zur Mitte und dessen Modellierung seit seiner ersten Formulierung durch Galton zu einer umfangreichen Bibliographie in allen Bereichen der quantitativen Wissenschaften geführt. In der Medizin z. B. stellt sich die zentrale Frage, wie groß ein solcher Regressionseffekt ist und wie er in epidemiologischen Längsschnit-

hebungen von einem Effekt einer gesundheitsrelevanten Intervention abzugrenzen ist. So wird z. B. die Frage nach dem Erfolg einer rehabilitationsmedizinischen Behandlung in der Reduktion von Arbeitsunfähigkeitstagen (AU-Tagen) gemessen. Seit den Längsschnittuntersuchungen von Wagner (1977), in denen sich Personen mit vielen AU-Tagen in den nächsten Jahren deutlich in ihrem Gesundheitszustand verbesserten wird dieser Effekt unter dem Stichwort „AU-Trend“ kontrovers diskutiert (Zwingmann/Wirtz 2005).

Aber auch bei anderen Szenarien der empirischen Sozialforschung taucht dieses Phänomen auf: So berichten Tversky und Kahnemann (1974), dass sich Piloten nach einer guten (besonders sanften) Landung auf einem Flugzeugträger in den Folgelandungen deutlich verschlechtern. In einer anderen Untersuchung stellte Secrist (1933) fest, dass große Firmen mit der Zeit an Bedeutung (gemessen an deren Umsatz) verlieren. Lee und Smith (2002) konnten in einem ähnlichen gelagerten Szenario zeigen, dass die Performance von erfolgreichen Football-Vereinen sich in den Folgejahren verschlechtert und daß diese Beobachtung auf einen Regressionseffekt zurückzuführen ist. Tatsächlich lassen sich solche Szenarien mit den mittlerweile zur Verfügung stehenden statistischen Methoden recht gut voraussagen, wenn entweder der Selektionsprozess modelliert werden kann (Varghese 1997) oder der tatsächliche Mittelwert der Grundpopulation bekannt ist (Mee/Chua 1991). Aber auch wenn der tatsächliche Mittelwert in einem bestimmten Intervall vermutet wird, so kann mit einer Erweiterung des Algorithmus von Mee und Chua eine Aussage über die Wahrscheinlichkeit des Vorliegens eines Regressionseffekts gemacht werden (Ostermann et al. 2008).

Literatur

- Brehm, J., 1999: Alternative corrections for sample truncation: Applications to the 1988, 1990, and 1992 senate election studies. *Political Analysis* 8(2): 183-199.
- Derichs, C., T. Heberer und J. Hippler, 2006: Wahlen und Regierbarkeit im globalen Rahmen. S. 11-24 in: T. Heberer und C. Derichs (Hg.): *Wahlsysteme und Wahltypen – Politische Systeme und regionale Kontexte im Vergleich*, Wiesbaden: VS Verlag für Sozialwissenschaften.
- Furby, L., 1973: Interpreting regression toward the mean in developmental research. *Developmental Psychology* 8(2): 172-179.
- Galton, F., 1886: Regression towards mediocrity in hereditary stature. *Journal of the Anthropological Institute* 15: 246-263.
- Hoecker, M., 2005. Die Oberbürgermeisterwahl in Stuttgart 1996 – Parteipolitik und Wahlkampfstrategie: die kommunale Persönlichkeitwahl im Spannungsfeld der modernen Parteiendemokratie. Eine Einzelfallstudie. Dissertation Universität Stuttgart.
- Kirschey, T., 2005: Bundestagswahl 2005 – regionale Parteihochburgen und ihre Strukturen. *Statistische Monatshefte Rheinland-Pfalz* 10: 593-604.
- Lazarsfeld, P. F., B. R. Berelson und H. Gaudet, 1969: *Wahlen und Wähler*. Neuwied und Berlin: Verlag Luchterhand.

- Lee, M. und G. Smith, 2002: Regression to the mean and football wagers. *Journal of Behavioral Decision Making* 15(4): 329-342.
- Lüdtke, R. und T. Ostermann, 2005: Regression zur Mitte – ein Thema in der Krebsforschung? *Deutsche Zeitschrift für Onkologie* 37(3): 169-175.
- Mee, R. und T. Chua, 1991: Regression Toward the mean and the paired sample t-test. *American Statistician* 45(1): 39-42.
- Nachtigall, C. und U. Suhl, 2002: Der Regressionseffekt – Mythos und Wirklichkeit. *Methevalreport* 4(2). Jena: Lehrstuhl für Psychologische Methodenlehre und Evaluationsforschung am Institut für Psychologie der Friedrich-Schiller-Universität Jena: http://www.metheval.uni-jena.de/materialien/reports/report_2002_02.pdf (24.7.2009).
- Ostermann, T., S. N. Willich und R. Lüdtke, 2008: Regression toward the mean – a detection method for unknown population mean based on Mee and Chua's algorithm. *BMC Medical Research Methodology* 8: 42.
- Probst, L., 2007: Vorwahlenanalyse zur Bürgerschaftswahl 2007. Institut für Politikwissenschaft Universität Bremen.
- Schoen, H., 2002: Wirkungen von Wahlprognosen auf Wahlen. S. 171-191 in: T. Berg (Hg.): *Moderner Wahlkampf*, Opladen: Leske und Budrich.
- Schoon, S., 2006: Wählerverhalten und Strukturmuster des Parteienwettbewerbs in Mecklenburg-Vorpommern nach der Landtagswahl 2006. *Rostocker Informationen zu Politik und Verwaltung* 27: 9-20.
- Secrist, H., 1933: *Triumph of mediocrity in business*. Chicago: Bureau of Business Research, Northwestern University.
- Südhüringische Zeitung 29.01.2008: Wahl in Hessen: Herbe Verluste in CDU-Hochburgen. Rhön im Landestrend.
- Tversky, A. und D. Kahneman, 1974: Judgment under uncertainty: Heuristics and Biases. *Science* 185: 1124 – 1131.
- Varghese, G., W. D. Johnson, A. Shahane und T. G. Nick, 1997: Testing for treatment effect in the presence of regression toward the mean. *Biometrics* 53 (1): 49-59.
- Wagner, H., 1977: Fehlerquellen bei Kurerfolgsbeurteilungen mittels Arbeitsausfallzeiten wegen Krankheit. *Zeitschrift für Physiotherapie* 29: 313-338.
- Zwingmann, C. und M. Wirtz, 2005: Regression zur Mitte. *Die Rehabilitation* 44(4): 244-251.

Anschrift der Autoren

PD Dr. Thomas Ostermann
Lehrstuhl für Medizintheorie, Integrative
und Anthroposophische Medizin
Universität Witten/Herdecke
Gerhard-Kienle-Weg 4
58313 Herdecke
thomaso@uni-wh.de

Rainer Lüdtke
Karl and Veronica Carstens Stiftung
Am Deimelsberg 36
45276 Essen
r.luedtke@carstens-stiftung.de