



**Institut für Statistik
Ludwig-Maximilians-Universität**

Bachelorarbeit

über das Thema

**Modellanalysen zur Schätzung von Wählerwanderung mit
Schwerpunkt auf das multinomiale Logit-Modell**

Autor: Robert Fuchs

Matrikelnummer:

Betreuer: Prof. Dr. Kauermann

Prüfer: Prof. Dr. Kauermann

Abgabedatum: 10. Juli 2014

I Abstract

Diese Bachelorarbeit beschäftigt sich mit dem multinomialen Logit-Modell zur Schätzung der Wählerwanderung. Es wird hierbei versucht, mit Hilfe von Aggregatdaten auf Individualdaten zu schließen und das Wahlverhalten Einzelner zu schätzen. So kann zum Beispiel die Wahrscheinlichkeit eines Wählers, dass er parteitreu wählt, festgestellt werden. Doch wie gut sind diese Schätzungen tatsächlich?

Das Ziel dieser Arbeit ist es, die Probleme des Modells vorzustellen und mögliche Lösungen zu finden. Um dies zu erreichen, wurden verschiedene Simulationen durchgeführt, zudem wurden verschiedene Datensätze von den Bundestagswahlen 2009 und 2013 sowie den Landtagswahlen 2008 und 2013 begutachtet.

II Inhaltsverzeichnis

I Abstract	I
II Inhaltsverzeichnis	II
1 Einleitung	1
2 Daten	1
3 Das multinomiale Modell zur Schätzung der Wählerwanderung	2
3.1 Ergebnisse und Probleme des Modells	3
3.2 Lösungsversuch	6
4 Simulationen zur Modellanalyse	9
4.1 Einfluss von geringen Wähleranteilen in der ersten Wahl	10
4.1.1 2x2 Tabelle	10
4.1.2 3x3 Tabellen	14
4.2 Einfluss der Bezirksgrößen	17
4.2.1 2x2 Tabellen	17
4.2.2 3x3 Tabellen	18
4.3 Einfluss von geringen Übergangswahrscheinlichkeiten	19
4.3.1 2x2 Tabellen	19
4.3.2 3x3 Tabellen	21
4.4 Lösungsversuch mit Bootstrap	22
4.5 Bootstrap bei den Bundestagswahlen	24
5 Fazit	26
6 Literatur	27
7 Abbildungsverzeichnis	28
8 Tabellenverzeichnis	31
9 Inhalt der CD-ROM	32

1 Einleitung

Regelmäßig haben deutsche Staatsbürger die Möglichkeit, an Wahlen teilzunehmen. Es gibt Bundestagswahlen, Landtagswahlen, Europawahlen und noch viele mehr. Mit Verbesserung der statistischen Methoden werden die Fragen nach dem Verhalten von Individuen immer wichtiger. Wie verhält sich beispielsweise ein Wähler bei der Bundestagswahl 2013, der 2009 SPD gewählt hat? Wie hoch ist die Wahrscheinlichkeit, dass er der SPD treu bleibt und diese erneut wählt? Oder wie wahrscheinlich ist es, dass er 2013 die FDP oder überhaupt nicht wählt? Solche Fragen werden bei den Wahlanalysen immer mehr begutachtet und durch verschiedene Modellierungen geschätzt. Die Problematik hierbei ist, dass die Erhebung von Individualdaten, beispielsweise durch Nachwahlbefragungen, aufwändig und kostspielig ist. Daher wird versucht, mit Hilfe von verschiedenen Modellen einen Schluss von Aggregatdaten auf Individualdaten zu ermöglichen. Eines davon ist das multinomiale Logit-Modell, das durch die Wahlergebnisse der beiden Wahlen in verschiedenen Bezirken eine Schätzung für Individuen abgibt. Es werden Wahrscheinlichkeiten abgegeben, welche Partei ein Wähler bei der Wahl 2013 wählt, wenn er 2009 eine bestimmte Partei gewählt hat. Doch wie zuverlässig sind solche Modellangaben? In dieser Bachelorarbeit soll diese Frage mit Hilfe von Simulationen und verschiedenen Wahldaten beantwortet werden.

2 Daten

Den Schwerpunkt der Daten bilden die Datensätze der Landtagswahlen 2008 und 2013 sowie der Bundestagswahlen 2009 und 2013 in den Münchner Bezirken, welche auf der Homepage der Stadt München für die Wahlen von 2013 frei verfügbar sind [1] [2]. Die Daten der Wahlen 2008 und 2009 wurden von der Stadt München bereit gestellt. Die Datensätze wurden zudem im Rahmen der universitären Lehrveranstaltung ‘Statistisches Praktikum’ aufbereitet. Da zwischen den Bundestagswahlen und Landtagswahlen die Stimmbezirke neu geordnet wurden, mussten die Bezirke in Identifikationsnummern zusammengefügt werden, um einen Vergleich zwischen den Wahlen zu ermöglichen. Somit ergeben sich für die Bundestagswahlen 650 und für die Landtagswahlen 649 neu geordnete Bezirke mit folgenden Informationen: Anzahl an Wahlberechtigten, Wählern, beantragte Wahlscheine, ungültige Stimmen sowie die absoluten Stimmzahlen aller in München angetretenen Parteien.

3 Das multinomiale Modell zur Schätzung der Wählerwanderung

Das multinomiale Logit-Modell ist eine Verallgemeinerung des einfachen logistischen Modells auf eine Zielgröße mit mehr als zwei Ausprägungen. Es wird hierbei davon aus gegangen, dass die Zielvariable $Y_i \in \{1, \dots, q, c\}$ keine Ordnungsstruktur besitzen und somit nominalskaliert sind. Zudem sind Kovariablen $x_i = (x_{i1}, \dots, x_{ip})$ und Regressionskoeffizienten β gegeben. Y_i stellt hierbei die, in der zweiten Wahl, gewählt Partei im Bezirk i und x_i die Wahlanteile der Partei $j \in 1, \dots, p$, in der ersten Wahl, im Bezirk i dar. [3]

Die Wahrscheinlichkeit für die Kategorie r ergibt sich wie folgt:

$$P(Y_i = r|x_i) = \pi_{ir} = \frac{\exp(x_i^T \beta_r)}{1 + \sum_{s=1}^q \exp(x_i^T \beta_s)} \quad r = 1, \dots, q \quad [3]$$

Für die Referenzkategorie c gilt:

$$\pi_{ic} = \frac{1}{1 + \sum_{s=1}^q \exp(x_i^T \beta_s)} \quad [3]$$

Im Folgenden wird das Modell nicht für einen Bezirk, sondern für die gesamte beobachtete Wahl angewandt, wodurch der Index i wegfällt. Somit lassen sich nun mit Hilfe der anteiligen Wahlergebnisse der ersten Wahl die absoluten Stimmzahlen der zweiten Wahl schätzen. Seien nun N die Gesamtstimmen der ersten Wahl, M die Gesamtstimmen der zweiten Wahl und x die Anteile der Parteien in der ersten Wahl, so lassen sich mit $P(Y = r|x)$ die Anteile der zweiten Wahl und mit $P(Y = r|x) * M$ die absoluten Stimmzahlen der zweiten Wahl schätzen. Fügt man nun bei der ersten Wahl einen zusätzlichen Wähler bei einer Partei $j \in 1, \dots, p$ (hier $j = 1$) hinzu, berechnet die Anteile von x mit \tilde{x}

$$\tilde{x} = \left(\frac{Nx_1+1}{N+1}, \frac{Nx_2}{N+1}, \dots, \frac{Nx_p}{N+1} \right)$$

neu und berechnet $P(Y = r|\tilde{x}) * (M + 1)$, erhält man die geschätzten absoluten Stimmergebnisse für die zweite Wahl mit einem zusätzlichen Wähler. Die Übergangswahrscheinlichkeit, dass ein Wähler der in der ersten Wahl t die Partei j gewählt hat in der zweiten Wahl die Partei r wählt, ergibt sich wie folgt:

$$P(Y = r|t = j) = P(Y = r|\tilde{x}) * (M + 1) - P(Y = r|x) * M$$

Man kann also eine Tabelle mit Übergangswahrscheinlichkeiten erstellen, in dem schrittweise jeder Partei genau ein zusätzlicher Wähler hinzugefügt und geschätzt wird, wie sich ein Wähler dieser Partei in der nächsten Wahl verhält. Als grundlegende Modellannahme gilt, dass $P(Y = r|t = j)$ im Intervall $[0,1]$ liegen muss.

Zudem muss $\sum_{r=1}^c P(Y = r|t = j) = 1$ gelten.

3.1 Ergebnisse und Probleme des Modells

Bei der Durchführung des Modells muss zunächst einmal geklärt werden, welche Parteien in das Modell eingerechnet werden. So sollten kleine Parteien, die aufgrund ihrer geringen Wähleranzahl eventuell zu einer Verzerrung führen könnten, zu einer Gruppe ‘Sonstige’ zusammengefasst werden. Die nächste Frage ist, wie mit Nichtwählern umgegangen wird. Schließlich hat eine Person, die beispielsweise 2009 die CSU gewählt hat, 2013 dennoch die Möglichkeit, nicht wählen zu gehen und andersrum. Es empfiehlt sich also, die Nichtwähler miteinzubeziehen. In der Tabelle 1 wurde nun das Modell für die Bundestagswahlen 2009 und 2013 in den Münchener Bezirken gerechnet. Alle Berechnungen dieser Arbeit wurden mit dem Programm Paket R [4] berechnet. Es wurden hierbei als Einflussgrößen die sechs größten Parteien der Wahl 2009, die kleineren Parteien gruppiert als Sonstige und die Nichtwähler begutachtet. Als Zielvariablen wurden die sechs großen Parteien von 2013 sowie die AfD, welche 2009 in München noch nicht angetreten war, Sonstige und Nichtwähler begutachtet. Die Werte in der Tabelle stellen Übergangswahrscheinlichkeiten in Prozent dar.

2013										
	CSU	SPD	FDP	Grüne	Linke	Piraten	AfD	Sonstige	NW	
2	CSU	91,4	4,7	1,1	-1,2	-1,5	-1,1	3,5	7,3	-4,2
0	SPD	18,9	72,2	1,7	-0,2	1,9	1,2	7,4	2,5	-5,5
0	FDP	34,5	2,4	41,5	7,6	-2,8	0,8	8,7	-5,9	13,1
9	Grüne	-6,0	22,5	10,3	68,6	9,0	3,7	-4,5	1,2	-4,7
	Linke	-26,5	25,5	-21,1	11,8	36,8	4,4	7,5	8,9	52,7
	Piraten	12,8	16,2	4,0	46,1	11,5	35,9	5,7	2,8	-34,9
	Sonstige	23,4	11,1	-18,6	-6,2	3,8	2,4	12,9	35,9	35,4
	NW	-1,4	4,0	-1,5	-3,8	1,7	1,4	-0,1	1,7	98,0

Tabelle 1: Geschätzte Wählerwanderung in Prozent der Bundestagswahlen 2009-2013 nach multinomialen Modell.

Ein kurzes Lesebeispiel: Ein Wähler der 2009 die SPD gewählt hat, wählt mit einer Wahrscheinlichkeit von 18,9% bei den Wahlen 2013 die CSU und mit einer Wahrscheinlichkeit von 72,2% wieder die SPD.

Schaut man sich nun aber die Tabelle genauer an, so erkennt man sofort eine entscheidende Schwäche des Modells: Es treten negative Werte auf. So wählt beispielsweise ein Wähler der Linken 2009 mit einer Wahrscheinlichkeit von -26,5% in der Wahl 2013 die CSU. Da jedoch die Werte dieser Tabelle lediglich im Intervall [0,1] sein können, unter der Restriktion, dass die Zeilensummen gleich eins sind, ist dieser Wert unlogisch. Es ist also festzustellen, dass Werte, die kleiner Null sind, unter- und Werte, die größer Eins sind, überschätzt werden. Um diese Unterschätzung bei den Linken nun auszugleichen, werden zwangsläufig andere Werte überschätzt. So hat ein Wähler, der 2009 die Linke gewählt hat, bei der Wahl 2013 eine Wahrscheinlichkeit von -26,5%, die CSU und -21,1%, die FDP zu wählen. Somit sind diese beiden Werte um mindestens -47,6% unterschätzt. Um dies nun auszugleichen, wird nun die Wahrscheinlichkeit, dass dieser Wähler etwas anderes wählt als CSU oder FDP, mit 147,6% deutlich überschätzt. Es findet also zwangsläufig mit jedem Wert außerhalb des Intervalls eine Verzerrung mehrerer Werte statt.

Mit Hilfe der Methode von Ambühl lassen sich die Werte wieder innerhalb des Intervalls bringen. Hierbei werden alle Werte, die außerhalb des Intervalls liegen, auf die jeweilige Grenze gelegt. Alle Werte kleiner Null werden somit auf Null und alle Werte größer Eins auf Eins gesetzt. Da nun die Restriktion, dass alle Zeilensummen Eins sein müssen, verletzt wird, werden die Zeilensummen nun normiert. Um nun die Randsummen der Übergangstabelle an die wahren Anteile der Wahlen anzupassen, wird ein iteratives Verfahren angewandt. Hierzu werden abwechselnd die Spalten- und Zeilensummen durch Multiplikation mit einem geeigneten Faktor an die wahren Anteile solange angepasst, bis ein vorgegebener maximaler Abstand eingehalten wird.[5]

In Tabelle 2 sind nun die Übergangswahrscheinlichkeiten aus Tabelle 1 mit Korrektur nach Ambühl dargestellt.

2013

	CSU	SPD	FDP	Grüne	Linke	Piraten	AfD	Sonstige	NW
2	84,1	5,3	0,9	0,0	0,0	0,0	3,1	6,6	0,0
0	SPD	15,6	72,4	1,2	0,0	1,8	1,1	5,9	2,0
0	FDP	32,9	2,7	33,1	8,7	0,0	0,9	8,1	0,0
9	Grüne	0,0	20,4	6,4	61,2	7,9	3,3	0,0	0,9
Linke	0,0	18,5	0,0	8,4	26,0	3,1	4,4	5,2	34,4
Piraten	8,3	12,7	2,1	35,5	8,7	27,3	3,6	1,8	0,0
Sonstige	17,9	10,3	0,0	0,0	3,5	2,2	9,6	27,1	29,5
NW	0,0	4,2	0,0	0,0	1,7	1,4	0,0	1,4	91,3

Tabelle 2: Geschätzte Wählerwanderung der Bundestagswahlen 2009-2013 in Prozent nach multinomialen Modell, mit Korrektur nach Ambühl.

Alle Werte, die in Tabelle 1 noch negativ waren, sind nun in Tabelle 2 auf Null gesetzt. Zudem kann man in den Zeilen, bei denen in Tabelle 1 noch hohe negative Werte standen, erkennen, dass sich die Werte deutlich verändert haben. So hat sich beispielsweise die Wahrscheinlichkeit, dass ein Wähler der Linken zum Nichtwähler wird, von 52,7% in Tabelle 1 auf 34,4% in Tabelle 2 verringert.

Zwar lässt sich durch dieses Verfahren das Problem mit Werten außerhalb des Intervalls bereinigen, jedoch werden negative Werte so zwangsläufig auf Null gesetzt. Im weiteren Verlauf dieser Arbeit wird versucht, das Modell so anzupassen, dass die Werte der Übergangstabelle innerhalb des Intervalls liegen.

Um festzustellen, ob es sich um einen Einzelfall handelt, dass negative Werte vorkommen, wird dasselbe Verfahren für die Landtagswahlen 2008 und 2013 mit den sechs großen Parteien und den restlichen Parteien, zusammengefasst als Sonstige, durchgeführt. Dieses mal wurden die Nichtwähler exemplarisch nicht berücksichtigt. Das Ergebnis ist in Tabelle 3 zu sehen.

2013

	CSU	SPD	FDP	Grüne	Linke	FW	Sonstige
2009	CSU	102,3	-5,5	2,2	-4,6	-1,5	4,6
	SPD	4,4	81,2	-3,4	5,4	4,3	-0,0
	FDP	44,2	5,5	49,9	9,7	-5,8	4,5
	Grüne	-28,4	45,1	10,8	63,6	3,7	2,0
	Linke	-40,4	84,1	-5,4	17,0	22,4	3,2
	FW	63,1	15,7	-16,2	-11,8	2,8	25,5
	Sonstige	51,8	37,4	-39,0	-14,9	5,9	13,8

Tabelle 3: Geschätzte Wählerwanderung der Landtagswahlen 2008-2013 in Prozent nach multinomialen Modell.

Auch in den Ergebnissen der Landtagswahlen 2008 und 2013, dargestellt in Tabelle 3, liegen Werte außerhalb des Intervalls [0,1]. So hat ein Wähler der CSU im Jahr 2008 eine Wahrscheinlichkeit von 102,3%, bei der Wahl 2013 nochmal die CSU zu wählen. Man sieht also, dass auch Werte oberhalb des vorgeschriebenen Intervalls [0,1] vorkommen. Zudem kommen ebenfalls wieder hohe negative Werte vor. So besteht eine Wahrscheinlichkeit von -39,0%, dass ein Wähler der sonstigen Parteien zur FDP wechselt. Es wird also auch hier eine Korrektur nach Ambühl benötigt, welche in der Tabelle 4 aufgezeigt wird.

2013

	CSU	SPD	FDP	Grüne	Linke	FW	Sonstige
CSU	89,9	0,0	1,5	0,0	0,0	5,4	3,1
SPD	2,9	81,5	0,0	5,0	3,3	0,0	7,4
FDP	39,5	7,7	35,0	12,5	0,0	5,3	0,0
Grüne	0,0	38,5	4,7	50,4	2,3	1,5	2,6
Linke	0,0	61,5	0,0	11,5	12,3	2,0	12,6
FW	40,9	15,8	0,0	0,0	2,1	22,2	19,1
Sonstige	26,1	29,3	0,0	0,0	3,5	9,3	31,8

Tabelle 4: Geschätzte Wählerwanderung der Landtagswahlen 2008-2013 in Prozent nach multinomialen Modell, mit Korrektur nach Ambühl.

3.2 Lösungsversuch

Um eine Lösung zu finden, wird nun versucht, die Struktur des Modells auszunutzen. Da es sich nicht um eine lineare Struktur handelt, sondern um eine logistische, ist der Einfluss von x_i auf $P(Y_i = r|t = j)$ in jedem Bezirk anders. Seien nun N_i die Gesamtstimmen der ersten Wahl im Bezirk i und x_i die Anteile der Parteien in der ersten Wahl im Bezirk $i = 1, \dots, m$. Als nächstes fügt man bei der ersten Wahl einen zusätzlichen Wähler bei einer Partei j (hier $j=1$) in einem Bezirk hinzu, berechnet die Anteile

$$\tilde{x}_i = \left(\frac{N_i x_{i1} + 1}{N_i + 1}, \frac{N_i x_{i2}}{N_i + 1}, \dots, \frac{N_i x_{ij}}{N_i + 1} \right)$$

neu und berechnet $P(Y_i = r|\tilde{x}_i) * (N_i + 1)$. Die Übergangswahrscheinlichkeit, dass ein Wähler der in der ersten Wahl t die Partei j gewählt hat in der zweiten Wahl die Partei r wählt, ergibt sich wie folgt:

$$P(Y_i = r|t = j) = P(Y_i = r|\tilde{x}_i) * (N_i + 1) - P(Y_i = r|x_i) * N_i$$

Damit ein großer Zuwachs der Wahlberechtigten zwischen den Wahlen das Ergebnis $P(Y_i = r|t = j)$ nicht zu sehr verzerrt, wurde in der Formel nicht mit der Bezirksgröße der zweiten, sondern mit der Bezirksgröße bei der ersten Wahl multipliziert. Somit lassen sich also nun die Übergangswahrscheinlichkeiten \hat{Y}_i für jeden einzelnen Bezirk erstellen.

Um nun eine für alle Bezirke gültige Übergangswahrscheinlichkeit \hat{Y} zu erhalten, wird eine Gewichtung nach Größe der Bezirke verwendet:

$$\hat{Y} = \frac{\sum_{i=1}^m \hat{Y}_i N_i}{N}$$

Mit dieser Modellanpassung wurden die Bundestagswahlen 2009 und 2013 wie in Tabelle 1 nochmal gerechnet und in Tabelle 5 dargestellt.

2013

	CSU	SPD	FDP	Grüne	Linke	Piraten	AfD	Sonstige	NW	
2009	CSU	88,9	5,9	0,5	-1,4	-1,3	-0,9	3,5	7,2	-2,4
	SPD	19,5	70,4	1,3	-0,6	1,9	1,2	7,3	2,6	-3,8
	FDP	33,0	3,1	42,0	6,8	-2,7	0,8	8,5	-5,5	14,0
	Grüne	-5,1	21,9	9,8	67,8	8,6	3,5	-4,2	1,3	-3,6
	Linke	-23,0	25,5	-21,6	11,5	36,2	4,3	7,5	8,7	50,9
	Piraten	13,2	16,1	3,3	45,1	11,2	35,1	5,7	2,9	-32,5
	Sonstige	24,5	12,2	-19,2	-6,1	3,9	2,4	12,6	34,9	34,8
	NW	0,5	5,1	-1,6	-3,8	1,6	1,5	0,2	1,7	94,9

Tabelle 5: Geschätzte Wählerwanderung der Bundestagswahlen 2009-2013 in Prozent nach multinomialen Modell, mit Modellanpassung. Es wurde eine Übergangstabelle für jeden einzelnen Bezirk geschätzt und dann, gewichtet mit der Bezirksgröße bei der ersten Wahl, eine für alle Bezirke gültige Tabelle berechnet.

Zwar sind die Schätzungen der Wahrscheinlichkeiten nun anders, aber das Resultat ist dennoch ähnlich. Es sind immer noch negative Werte vorhanden und auch die Größe der Werte hat sich nur bedingt verändert.

In der Abbildung 1 sind nun die Übergangswahrscheinlichkeiten für einen Wähler der Linken im Jahr 2009 zur CSU in der Wahl 2013 bei den Bundestagswahlen für jeden Bezirk dargestellt.

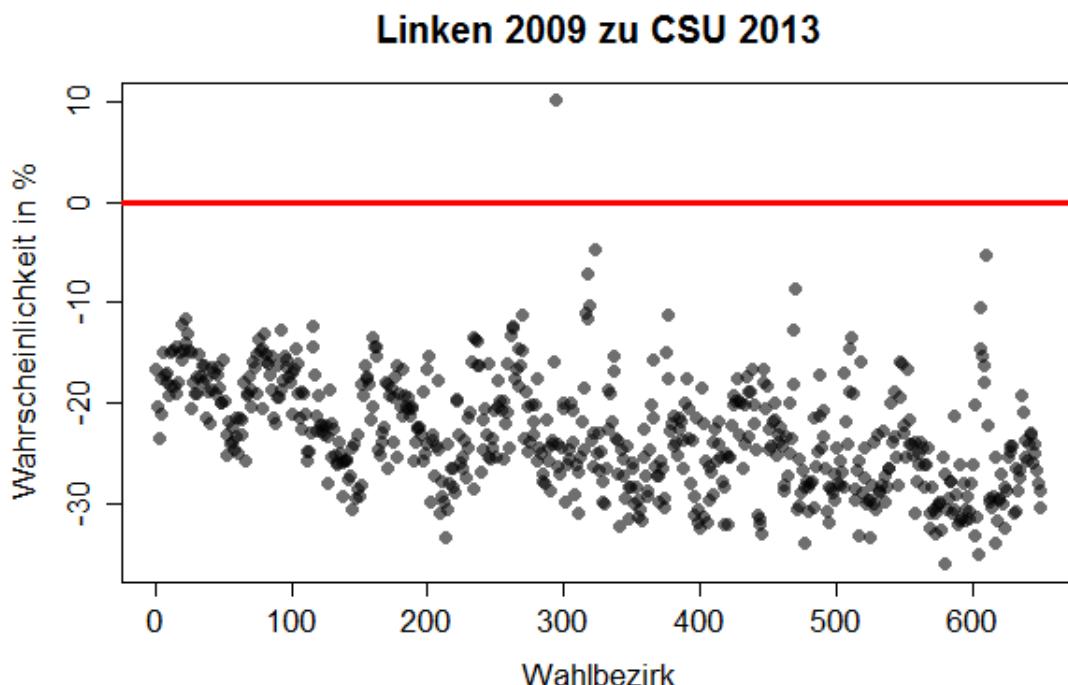


Abbildung 1: Streudiagramm der Wahrscheinlichkeiten, dass ein Wähler der Linken zur CSU wechselt.

Jeder Punkt symbolisiert hierbei einen Bezirk mit einer prozentualen Wahrscheinlichkeit, dass ein Wähler der Linken von 2009 in der Wahl 2013 die CSU wählt. Die rote Linie spiegelt hierbei die untere Grenze des Intervalls $[0,1]$ für die Werte der Übergangswahrscheinlichkeit wieder. Man kann schnell erkennen, dass lediglich ein Bezirk innerhalb des Intervall liegt und alle anderen 649 Bezirke kleiner Null sind. Zwar gibt es Unterschiede zwischen den Bezirken, jedoch ist dies keine Möglichkeit, die negativen Werte zu verhindern. Alle folgenden Modellberechnungen wurden mit dieser Modellanpassung durchgeführt.

4 Simulationen zur Modellanalyse

Um die Ursachen für die Werte außerhalb des Intervalls [0,1] herauszufinden, werden nun verschiedene Simulationen durchgeführt. Als erster Schritt wird hierbei eine theoretische Übergangsmatrix u und $w = (w_1, w_2, \dots, w_p)$ mit den relativen Wahlanteilen der Parteien in der ersten Wahl erstellt,

$$u = \begin{pmatrix} u_{11} & u_{12} & \cdots & u_{1q} \\ u_{21} & u_{22} & \cdots & u_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ u_{p1} & u_{p2} & \cdots & u_{pq} \end{pmatrix}$$

wobei p der Anzahl der Parteien bei der ersten und q der Anzahl der Parteien bei der zweiten Wahl entspricht.

Als nächstes werden n Personen aus einer Multinomialverteilung $M(n, \pi = w)$ gezogen und dies für m Bezirke wiederholt. Somit erhält man ein w^* mit simulierten absoluten Wählerstimmen pro Partei und Bezirk, welches die Wahlstimmen der ersten Wahl repräsentiert:

$$w^* = \begin{pmatrix} w_{11}^* & w_{12}^* & \cdots & w_{1p}^* \\ w_{21}^* & w_{22}^* & \cdots & w_{2p}^* \\ \vdots & \vdots & \ddots & \vdots \\ w_{m1}^* & w_{m2}^* & \cdots & w_{mp}^* \end{pmatrix}$$

Um nun das Modell erfolgreich auf die simulierten Daten rechnen zu können, fehlen noch Wählerstimmen für die zweite Wahl. Um diese zu erhalten, wird für jeden Bezirk eine Kontingenztabelle simuliert. Exemplarisch wird dies nun für den ersten Bezirk $m = 1$ und vier Parteien gezeigt, wobei $p = 4$ und $q = 4$ gilt.

	CSU13	SPD13	FDP13	Grüne13	Σ
CSU09	$y \sim M(w_{11}^*, \pi = (u_{11}, u_{12}, u_{13}, u_{14}))$				$\Sigma_{1\bullet}$
SPD09		$y \sim M(w_{12}^*, \pi = (u_{21}, u_{22}, u_{23}, u_{24}))$			$\Sigma_{2\bullet}$
FDP09			$y \sim M(w_{13}^*, \pi = (u_{31}, u_{32}, u_{33}, u_{34}))$		$\Sigma_{3\bullet}$
Grüne09				$y \sim M(w_{14}^*, \pi = (u_{41}, u_{42}, u_{43}, u_{44}))$	$\Sigma_{4\bullet}$
Σ	$\Sigma_{\bullet 1}$	$\Sigma_{\bullet 2}$	$\Sigma_{\bullet 3}$	$\Sigma_{\bullet 4}$	

Tabelle 6: Kontingenztabelle für den ersten Bezirk mit vier Parteien bei der Wahl 2009 und vier Parteien bei der Wahl 2013. Es wird eine Übergangstabelle für den Bezirk simuliert.

Es wird also zeilenweise aus einer Multinomialverteilung gezogen, wobei der Umfang n die vorher simulierte Anzahl an Wählerstimmen dieser Partei in diesem Bezirk ist und die Wahrscheinlichkeit π aus der Übergangstabelle u übernommen wird. Die Spaltensummen bilden dann die simulierten Stimmzahlen der Parteien im ersten Bezirk. Dies wird nun für die restlichen Bezirke wiederholt, wodurch man die Stimmzahlen aller Parteien in allen Bezirken erhält. Somit ist die Anzahl der Wählerstimmen bei der ersten und der zweiten Wahl in jedem der m Bezirke identisch. Nun hat man alle Daten simuliert, um das Modell rechnen zu können. Die folgenden Simulationen wurden nun für unterschiedliche Übergangstabellen u , relativen Wahlanteile w in der ersten Wahl und Anzahl Wähler n gerechnet. Die Anzahl der Bezirke war bei jeder Simulation $m = 650$.

4.1 Einfluss von geringen Wähleranteilen in der ersten Wahl

Ein Blick auf die Tabelle 1 lässt die Vermutung offen, dass gerade Parteien mit kleinen Wahlanteilen in der ersten Wahl Probleme mit negativen Werten haben. So haben die Linke, Piraten und Sonstige hohe negative Werte. Bei der CSU treten zwar ebenfalls negative Werte auf, jedoch sind diese relativ nahe an der Intervallsgrenze. Um diese Vermutung zu testen, wurden nun verschiedene Simulationen durchgeführt.

4.1.1 2x2 Tabelle

Als erstes wurde die Simulation 1 für den Fall mit $p = 2$ und $q = 2$ Parteien und den Bezirksgrößen $n = 1000$ mit 1000 Iterationen durchgeführt.

$$u = \begin{pmatrix} 0,9 & 0,1 \\ 0,01 & 0,99 \end{pmatrix} \quad w = (0,82; 0,18)$$

In Tabelle 7 sind nun die Mittelwerte und die Varianzen für die Schätzungen der Übergangswahrscheinlichkeiten dieser Simulation aufgeführt. Es wurden hierbei die 1000 simulierten geschätzten Tabellen beobachtet und dann der Mittelwert und die Varianz für die einzelnen Zellen berechnet. Die Varianzen sind, aufgrund der binomialverteilten Struktur im 2x2 Fall, zeilenweise identisch. Hat beispielsweise eine Tabelle die Schätzung 0,8 für die Zelle A09-A13, so hat sie automatisch $1 - 0,8 = 0,2$ in der Zelle A09-B13. Durch diese Struktur sind die Abweichungen zum jeweiligen Mittelwert und somit auch die Varianzen in den Zeilen gleich. Dies gilt nur für den 2x2 Fall.

	A13	B13		A13	B13		
A09	0,9001	0,0999	1	A09	0,000029	0,000029	
B09	0,0102	0,9898	1	B09	0,000588	0,000588	

Tabelle 7: Simulation 1: In der linken Tabelle sind die Mittelwerte und in der rechten Tabelle die Varianzen der 1000 Iterationen des Modells in den jeweiligen Zellen aufgeführt. $w = (0,82; 0,18)$.

Wie man erkennen kann, wird die Übergangstabelle u im Mittel sehr genau getroffen. Die Varianz ist jedoch bei der Partei B09 mit einem Wahlanteil von 0,18 um etwa das zwanzigfache höher als bei der Partei A09 mit einem Wahlanteil von 0,82. Dies spricht für eine höhere Streuung bei den Werten.

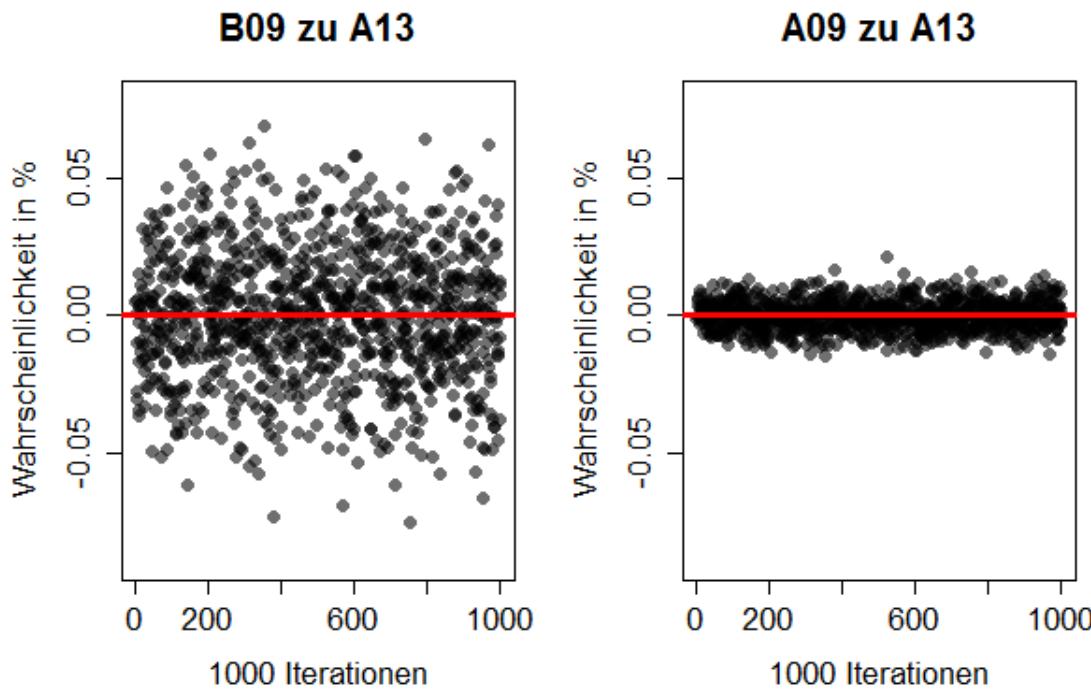


Abbildung 2: Simulation 1, mit $w = (0,82; 0,18)$:

Linke Grafik: Streudiagramm der 1000 um den Mittelwert zentrierten Werte für die geschätzte Wahrscheinlichkeit, dass ein Wähler von der Partei B09 zu A13 wechselt. In rot ist die Nulllinie eingezeichnet.

Rechte Grafik: Streudiagramm der 1000 um den Mittelwert zentrierten Werte für die geschätzte Wahrscheinlichkeit, dass ein Wähler von der Partei A09 bei A13 bleibt. In rot ist die Nulllinie eingezeichnet.

In Abbildung 2 sind nun zwei Streudiagramme für die je 1000 Werte der Simulation 1. Die linke Grafik stellt hierbei die 1000 Werte der Zelle B09 zu A13 und die rechte Grafik die 1000 Werte der Zelle A09 zu A13 dar. Die Werte wurden um den jeweiligen Mittelwert zentriert und in rot die Nulllinie eingezeichnet. Man kann erkennen, dass in beiden Zellen eine Streuung um den Mittelwert vorkommt. Hierdurch treten auch vermehrt negative Werte auf. Die Streuung in der linken ist jedoch deutlich höher als in der rechten Grafik. Dies kann an den kleineren Wahlanteilen der Partei B09 liegen.

Was geschieht jetzt, wenn man die Wahlanteile w verändert und alles andere gleich lässt? Um dies herauszufinden, wurden in der Simulation 2 die Wahlanteile $w = (0,95; 0,05)$ geändert. Alle übrigen Parameter bleiben identisch mit denen aus Simulation 1. Der Mittelwert und die Varianz der Tabelle sind in Tabelle 8 zu sehen.

	A13	B13	
A09	0,9002	0,0998	1
B09	0,0090	0,9910	1

	A13	B13	
A09	0,0000073	0,0000073	
B09	0,0025463	0,0025463	

Tabelle 8: Simulation 2, mit $w = (0,95; 0,05)$: In der linken Tabelle sind die Mittelwerte und in der rechten Tabelle die Varianzen der 1000 Iterationen des Modells in den jeweiligen Zellen aufgeführt.

Auch hier liegen die Mittelwerte nahe an der vorgegebenen Übergangstabelle. Schaut man sich nun aber die Varianzen an, so erkennt man, dass diese sich deutlich geändert haben. Bei der Partei A09 fällt die Varianz mit Anstieg der Anteile von 0,82 in Simulation 1 auf 0,95 in Simulation 2 von 0,000029 auf 0,0000073. Die Streuung nimmt hier also durch den steigenden Wahlanteil in der ersten Wahl ab. Im Gegensatz dazu steigt die Varianz der Partei B09 von 0,000588 in der Simulation 1 auf 0,0025463 in der Simulation 2. Dies stellt in etwa eine Vervierfachung dar. Die Wahlanteile der ersten Wahl haben sich hier von 0,18 auf 0,05 gesenkt.

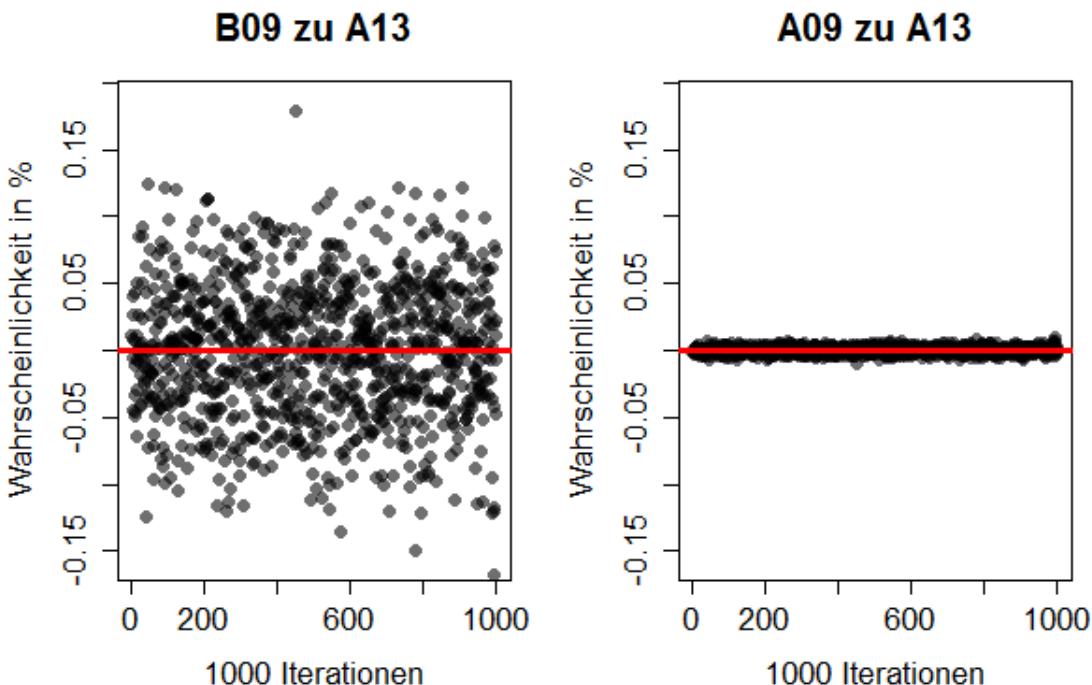


Abbildung 3: Simulation 2, mit $w = (0,95; 0,05)$:

Linke Grafik: Streudiagramm der 1000 um den Mittelwert zentrierten Werte für die geschätzte Wahrscheinlichkeit, dass ein Wähler von der Partei B09 zu A13 wechselt. In rot ist die Nulllinie eingezeichnet.

Rechte Grafik: Streudiagramm der 1000 um den Mittelwert zentrierten Werte für die geschätzte Wahrscheinlichkeit, dass ein Wähler von der Partei A09 bei A13 bleibt. In rot ist die Nulllinie eingezeichnet.

Begutachtet man nun die Streuung der Werte von Partei B09 zu A13, so erkennt man, dass die Streuweite deutlich zunimmt. Dies ist in Abbildung 3 zu sehen. Im Vergleich zur Abbildung 2, die die Streuung aus der Simulation 1 widerspiegelt, ist hier die Spannweite der Y-Achse wesentlich höher. Um dies in Zahlen zu veranschaulichen, sind in Tabelle 9 verschiedene Lagemaße zu den beiden Simulationen aufgeführt. Schaut man sich die Zelle B09-A13 an, so sieht man, dass die Spannweite der Daten bei Simulation 2 deutlich zunimmt. Bei der Zelle A09-A13 nimmt sie dagegen leicht ab.

	Simulation 1				Simulation 2			
	Minimum	u. Quartil	o. Quartil	Maximum	Minimum	u. Quartil	o. Quartil	Maximum
A09-A13	0,8851	0,8966	0,9034	0,9216	0,8903	0,8983	0,9021	0,9094
B09-A13	-0,0897	-0,0045	0,0262	0,0789	-0,1580	-0,0263	0,0445	0,1879

Tabelle 9: Simulation 1 ($w = (0,82; 0,18)$) und 2 ($w = (0,95; 0,05)$): Tabelle mit Lagemaßen der beiden Simulationen für die Werte der Zellen A09-A13 und B09-A13.

Um dies nun für noch mehr Wahlanteile w zu untersuchen, wurde die identische Simulation mit je 1000 Iterationen für $w = (1 - c, c)$ durchgeführt. In Abbildung 4 sind nun Boxplots für verschiedene Wahlanteile c aufgeführt. Die Werte der Boxplots in der linken Grafik sind die Wahrscheinlichkeiten, dass ein Wähler der Partei B09 zur Partei A09 wechselt. Die Boxplots in der rechten Grafik sind die Wahrscheinlichkeiten, dass ein Wähler A09 wieder A13 wählt. Man kann sehen, dass mit sinkenden Wahlanteil der Partei B09 die Spannweite der Boxplots größer wird. In der rechten Grafik sinkt die Spannweite der Boxplots mit steigendem Wahlanteil von A09, mit $1 - c$.

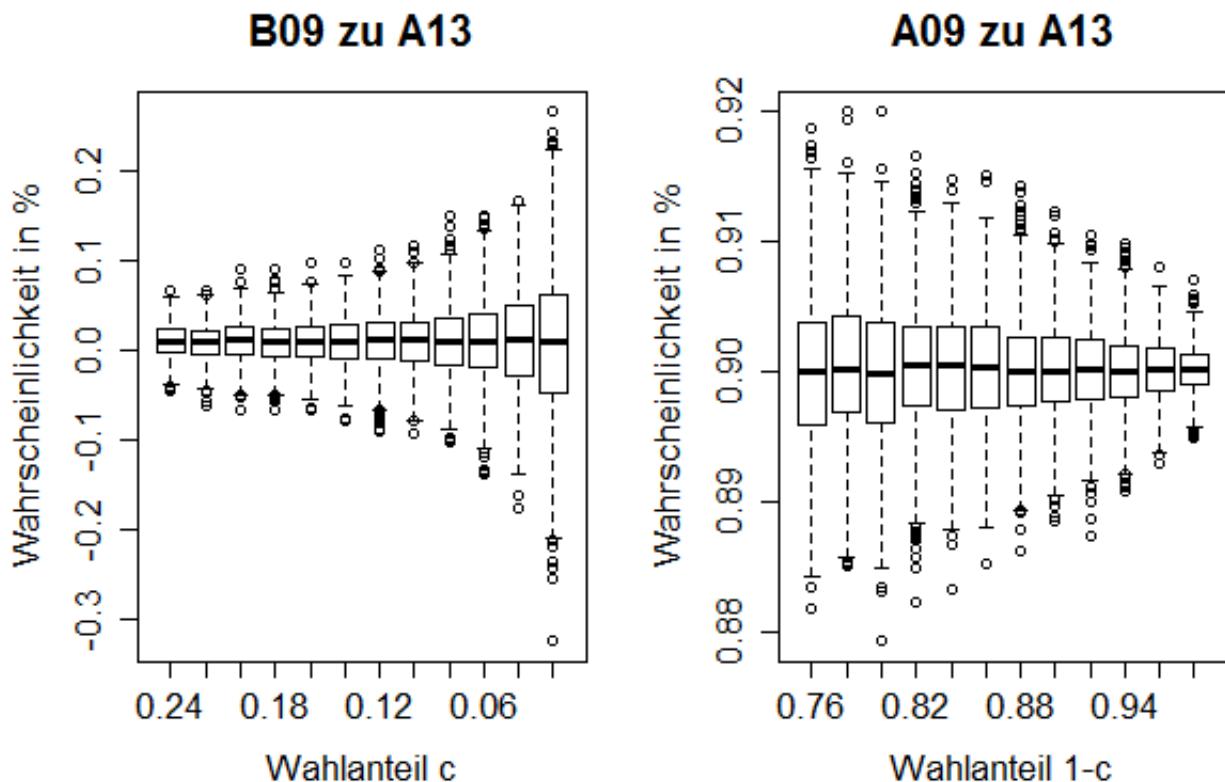


Abbildung 4: Simulation 3: Boxplots mit je 1000 Iterationen für verschiedene Wahlanteile c .
 Linke Grafik: Die Werte stellen die Übergangswahrscheinlichkeiten aus dem simulierten Modell aus der Zelle B09-A13 dar.
 Rechte Grafik: Die Werte stellen die Übergangswahrscheinlichkeiten aus dem simulierten Modell aus der Zelle A09-A13 dar.

Es kann also festgestellt werden, dass die Genauigkeit der Modellprognosen mit fallenden Wahlanteilen in der ersten Wahl abnimmt. Ebenso kann festgestellt werden, dass die Prognose im Mittel sehr genau an der tatsächlichen Übergangswahrscheinlichkeit liegt. Jedoch ist der Fall in einer 2x2 Tabelle aufgrund der binomialverteilten Struktur ein spezieller. Aus diesem Grund wird nun der Einfluss der Wahlanteile auf die Modellprognose in einem 3x3 Fall überprüft.

4.1.2 3x3 Tabellen

Als erstes wurde die Simulation 4 für $p = 3$ und $q = 3$ Parteien und den Bezirksgrößen $n = 1000$ mit 1000 Iterationen durchgeführt. Die Übergangsmatrix u und Wahlanteile w wurden wie folgt angepasst:

$$u = \begin{pmatrix} 0,9 & 0,03 & 0,07 \\ 0,2 & 0,7 & 0,1 \\ 0,005 & 0,495 & 0,5 \end{pmatrix} \quad w = (0,8; 0,18; 0,02)$$

In Tabelle 10 sind nun die Mittelwerte und die Varianzen angegeben. Auch im 3x3 Fall wird im Mittel die Übergangstabelle sehr genau vorhergesagt. Schaut man sich die Varianzen an, so kann man auch hier erkennen, dass die Varianzen bei der Partei C09 am größten sind, gefolgt von B09. Die Größe der Varianzen nimmt also auch in diesem Fall mit fallenden Wahlanteilen zu.

	A13	B13	C13		A13	B13	C13
A09	0,900	0,030	0,070	1	0,0000380	0,0000254	0,0000292
B09	0,199	0,700	0,101	1	0,0007321	0,0004871	0,0005668
C09	0,001	0,499	0,500	1	0,0073197	0,0049749	0,0054104

Tabelle 10: Simulation 4: In der linken Tabelle sind die Mittelwerte und in der rechten Tabelle die Varianzen der 1000 Iterationen des Modells in den jeweiligen Zellen aufgeführt.

In Abbildung 5 sind die um den jeweiligen Mittelwert zentrierten Ergebnisse der Zellen C09-A13 und A09-A13 in einem Streudiagramm abgebildet. Man kann sehr schnell erkennen, dass die Streuung bei dem Wahlanteil der Partei C09 mit einem Wert von 0,02 deutlich höher ist als die Streuung bei dem Wahlanteil der Partei A09 mit einem Wert von 0,8. Dennoch streuen beide Diagramme erneut um den Mittelwert, der aufgrund der Zentrierung die Nulllinie darstellt.

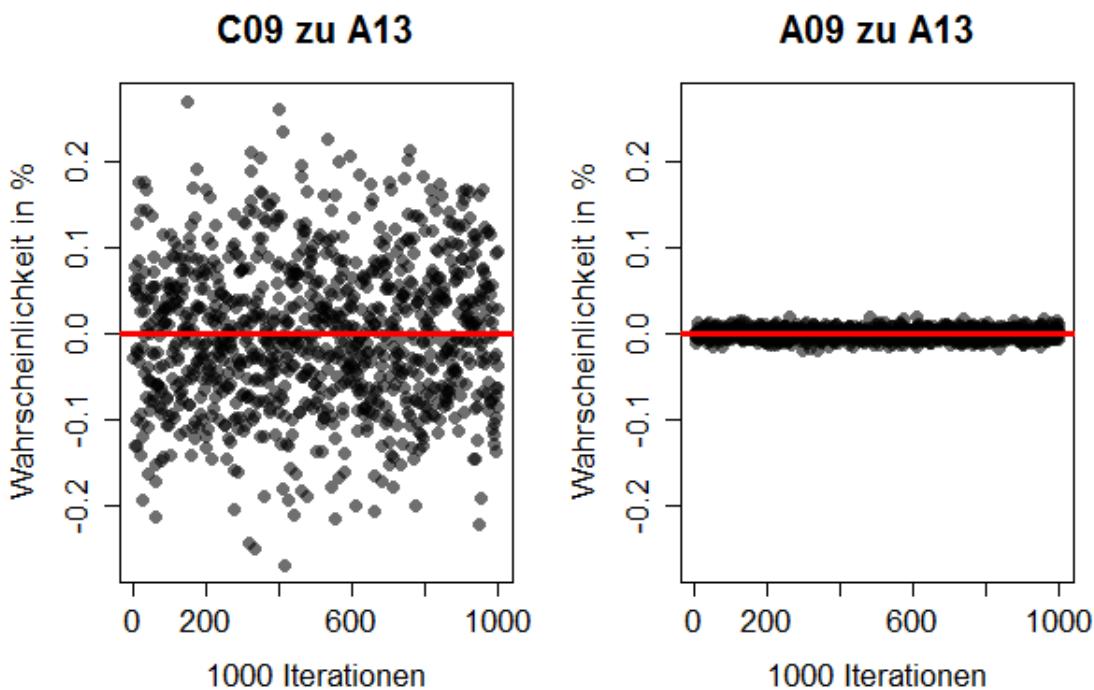


Abbildung 5: Simulation 4: Das linke Streudiagramm bildet die um den Mittelwert zentrierten Ergebnisse des Modells der Zelle C09-A13. Das rechte Streudiagramm bildet die um den Mittelwert zentrierten Ergebnisse des Modells der Zelle A09-A13.

In Simulation 5 wurde das nun für verschiedene Wahlanteile getestet.

Mit $w = (0,7 - c; 0,3; c)$ wurde die Simulation für jedes c mit 1000 Iterationen durchgeführt. Die restlichen Parameter bleiben wie in Simulation 3. Und was in der 2x2 Tabelle schon zu sehen war, bestätigt sich auch in der 3x3 Tabelle und ist in Abbildung 6 zu erkennen. Mit fallenden c , welches den Wahlanteil von C09 darstellt, steigt auch die Spannweite der Boxplots in Zelle C09-A13. Da gleichzeitig der Wahlanteil A09 steigt, fällt in der linken Grafik, welche die Zelle A09-A13 widerspiegelt, die Spannweite der Boxplots.

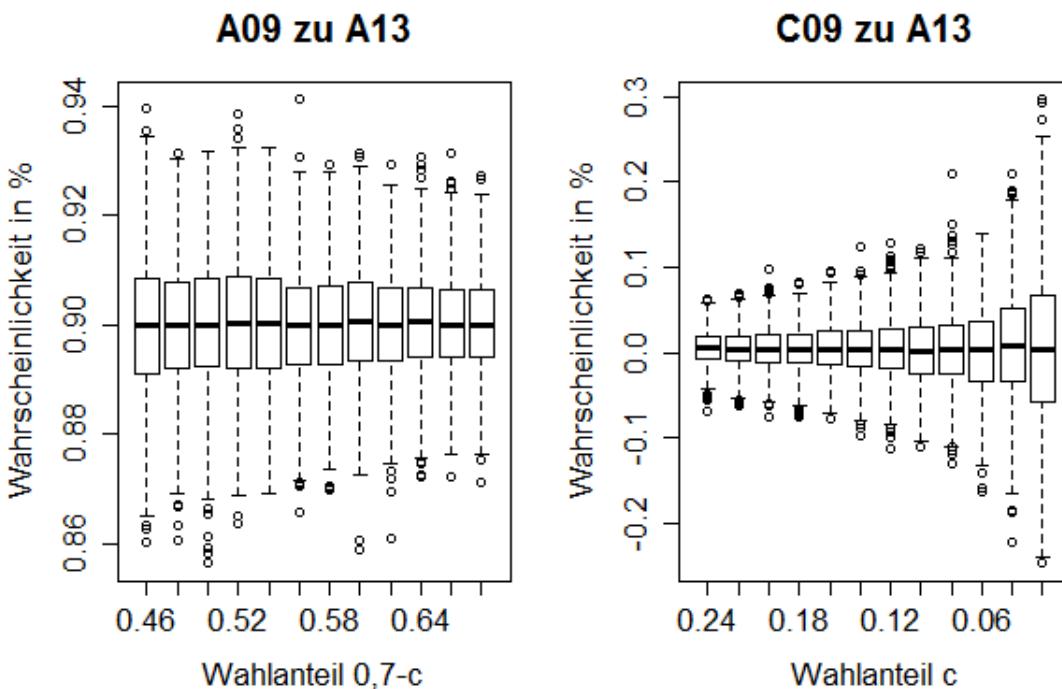


Abbildung 6: Simulation 5: Boxplots mit je 1000 Iterationen für verschiedene Wahlanteile c . Linke Grafik: Die Werte stellen die Übergangswahrscheinlichkeiten aus dem simulierten Modell aus der Zelle A09-A13 dar.

Rechte Grafik: Die Werte stellen die Übergangswahrscheinlichkeiten aus dem simulierten Modell aus der Zelle C09-A13 dar.

Sowohl in den Simulationen für 2x2 Tabelle, als auch in den Simulationen für 3x3 Tabellen war ersichtlich, dass die Varianzen einer Zeile mit fallenden Wahlanteil der Partei in der ersten Wahl steigen. Trotzdem streuen die Ergebnisse um den jeweiligen Mittelwert, wodurch bei geringen Übergangswahrscheinlichkeiten und hohen Varianzen negative Werte auftreten können.

4.2 Einfluss der Bezirksgrößen

4.2.1 2x2 Tabellen

Da es vielleicht noch andere Gründe für die hohen negativen Werte gibt, soll als nächstes der Einfluss von verschiedenen Bezirksgrößen untersucht werden. Hierzu wurde die Simulation mit 1000 Iterationen durchgeführt. Für die Simulation 6 wurde hierbei eine Bezirksgröße $n = 1000$ und für die Simulation 7 eine Bezirksgröße von $n = 100$ hergenommen. Die Wahlanteile und Übergangswahrscheinlichkeiten wurden wie folgt vorgegeben:

$$u = \begin{pmatrix} 0,7 & 0,3 \\ 0,2 & 0,8 \end{pmatrix} \quad w = (0,6; 0,4)$$

	A09	B09		A13	B13		
A09	0,6999	0,3001	1	A09	0,0001913	0,0001913 <th>A09</th>	A09
B09	0,2001	0,7999	1	B09	0,0004274	0,0004274	B09

Tabelle 11: Simulation 6, $n = 1000$: In der linken Tabelle sind die Mittelwerte und in der rechten Tabelle die Varianzen der 1000 Iterationen des Modells in den jeweiligen Zellen aufgeführt.

	A09	B09		A13	B13		
A09	0,7005	0,2995	1	A09	0,0002042	0,0002042 <th>A09</th>	A09
B09	0,1992	0,8008	1	B09	0,0004502	0,0004502	B09

Tabelle 12: Simulation 7, $n = 100$: In der linken Tabelle sind die Mittelwerte und in der rechten Tabelle die Varianzen der 1000 Iterationen des Modells in den jeweiligen Zellen aufgeführt.

In Tabelle 11 sind die Ergebnisse der Mittelwerte und Varianzen für die Simulation 6 mit $n = 1000$ und in Tabelle 12 für die Simulation 7 mit $n = 100$ zu sehen. Die Mittelwerte sind zwar bei $n = 1000$ genauer, jedoch ist die Abweichung bei $n = 100$ so gering, dass sie zu vernachlässigen ist. Auch die Varianzen ähneln sich in beiden Simulationen und sind für den Fall $n = 100$ minimal höher. Allerdings kann auf den ersten Blick kein relevanter Einfluss der Bezirksgröße auf das Ergebnis erkannt werden.

Schaut man sich nun die Lagermaße der beiden Simulationen in Tabelle 13 an, bestätigt sich der erste Eindruck. Auch hier gibt es keine Veränderung durch die Bezirksgrößen. Die Lagermaße sind für die Simulation 6 und 7 fast identisch.

	Simulation 6				Simulation 7			
	Minimum	u. Quartil	o. Quartil	Maximum	Minimum	u. Quartil	o. Quartil	Maximum
A09-A13	0,6526	0,6905	0,7091	0,7471	0,6562	0,6902	0,7099	0,7484
B09-A13	0,1294	0,1860	0,2142	0,2711	0,1279	0,1843	0,2142	0,2661

Tabelle 13: Simulation 6 ($n = 1000$) und 7 ($n = 100$): Tabelle mit Lagemaßen der beiden Simulationen für die Werte der Zellen A09-A13 und B09-A13

Um sicher zu gehen, dass die Bezirksgrößen keinen Einfluss in den 2x2 Tabellen haben, wurde die identische Simulation mit verschiedenen n durchgeführt. Doch auch die Ergebnisse in Abbildung 7 zeigen, dass sich die Spannweite der Boxplots durch die Größe n der Bezirke kaum verändert. Es kann also kein Einfluss festgestellt werden.

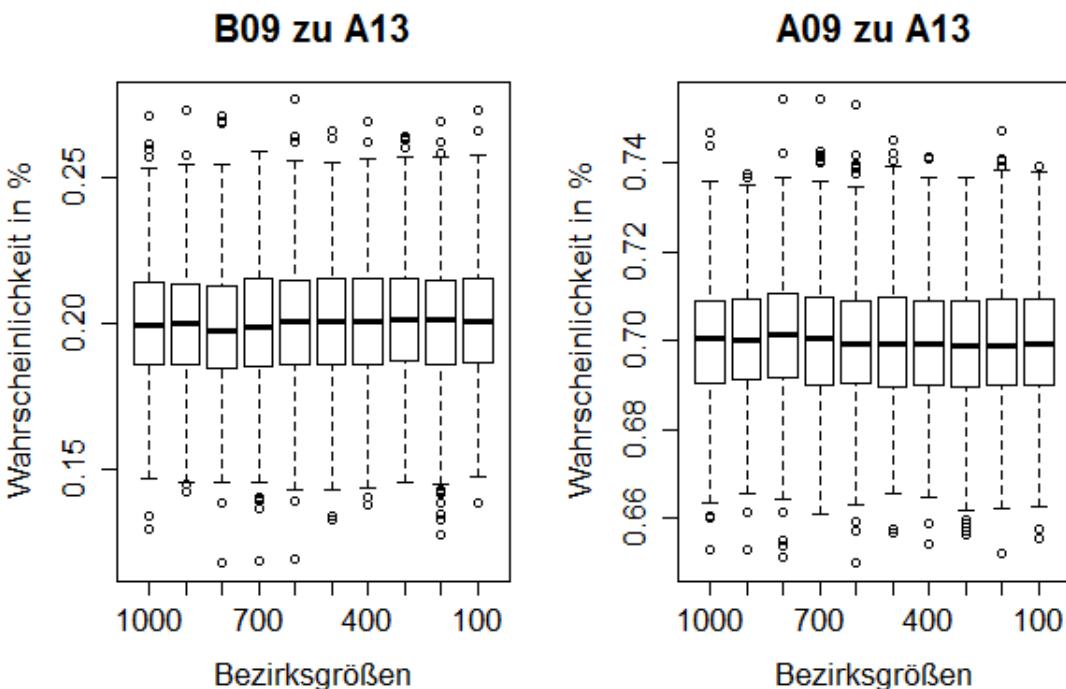


Abbildung 7: Simulation 8: Boxplots mit je 1000 Iterationen für verschiedene Bezirksgrößen n . Linke Grafik: Die Werte stellen die Übergangswahrscheinlichkeiten aus dem simulierten Modell aus der Zelle B09-A13 dar.

Rechte Grafik: Die Werte stellen die Übergangswahrscheinlichkeiten aus dem simulierten Modell aus der Zelle A09-A13 dar.

4.2.2 3x3 Tabellen

Als nächstes wurde die Simulation für eine 3x3 Tabelle mit 1000 Iterationen durchgeführt. Die Übergangswahrscheinlichkeiten und Wahlanteile waren wie folgt:

$$u = \begin{pmatrix} 0,8 & 0,1 & 0,1 \\ 0,2 & 0,7 & 0,1 \\ 0,05 & 0,35 & 0,6 \end{pmatrix} \quad w = (0,6; 0,3; 0,1)$$

Die Simulation wurde für verschiedene n im Intervall $[100, 1000]$ wiederholt und das Ergebnis ist in Abbildung 8 zu sehen.

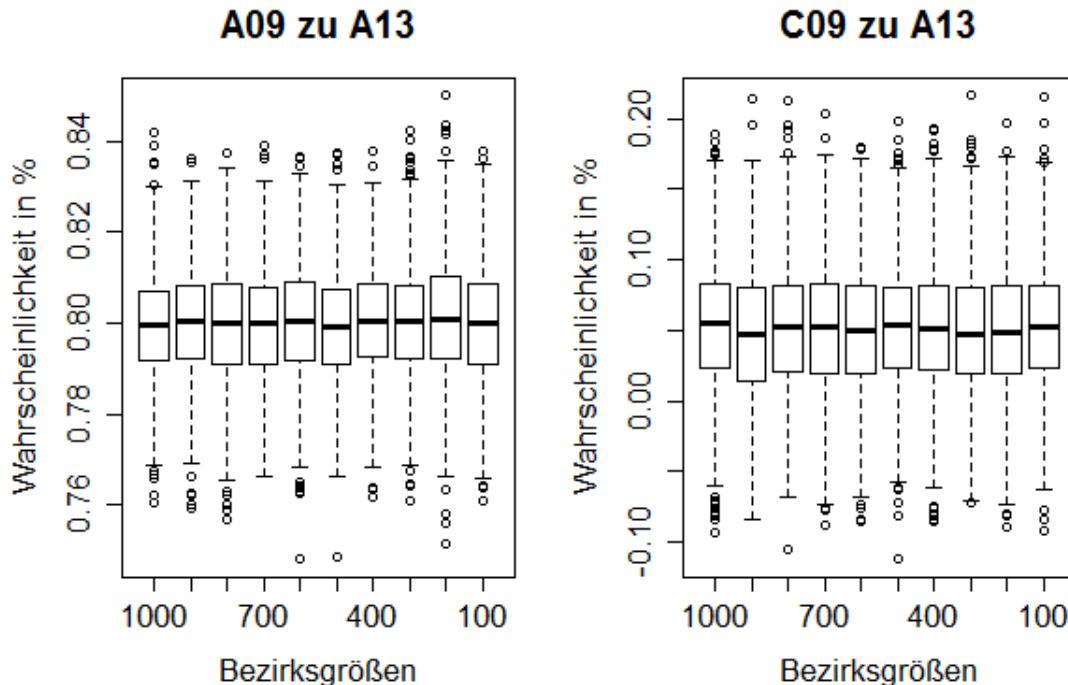


Abbildung 8: Simulation 9: Boxplots mit je 1000 Iterationen für verschiedene Bezirksgrößen n . Linke Grafik: Die Werte stellen die Übergangswahrscheinlichkeiten aus dem simulierten Modell aus der Zelle A09-A13 dar.

Rechte Grafik: Die Werte stellen die Übergangswahrscheinlichkeiten aus dem simulierten Modell aus der Zelle C09-A13 dar.

Wie bereits im 2x2 Fall ändern sich hier die Boxplots der Zellen nicht, wenn man die Bezirksgrößen verändert. Auch die Ansicht von Streudiagrammen oder der Varianzen hat keine Schlussfolgerung zugelassen, dass die Bezirksgröße einen Einfluss auf das Ergebnis des Modells hat.

4.3 Einfluss von geringen Übergangswahrscheinlichkeiten

4.3.1 2x2 Tabellen

Einen möglichen Einfluss auf die Modellergebnisse könnten auch die tatsächlichen Übergangswahrscheinlichkeiten haben. Wie bereits in vorherigen Simulationen festgestellt, schwanken die Ergebnisse stets um den tatsächlichen Wert. Wie groß diese Schwankungen sind, hängt nach derzeitigen Wissen vor allem von den Wahlanteilen der

ersten Wahl ab. Somit treten hauptsächlich bei kleinen Übergangswahrscheinlichkeiten mit geringem Wahlanteil der jeweilige Partei negative Werte auf. Im Folgenden wird nun versucht, den Einfluss von geringen Übergangswahrscheinlichkeiten herauszufinden. Haben kleine Übergangswahrscheinlichkeiten eine größere Streuung? Um diese Hypothese zu testen, wird zunächst eine Simulation mit 1000 Iterationen mit $n = 1000$ und w und u wie folgt durchgeführt:

$$u = \begin{pmatrix} 0,7 & 0,3 \\ b & 1-b \end{pmatrix} \quad w = (0,6; 0,4)$$

Es wurden also verschiedene Übergangswahrscheinlichkeiten, b und $1 - b$, für die Partei B09 gewählt. Die Ergebnisse sind in der Abbildung 9 in verschiedenen Boxplots zu sehen.

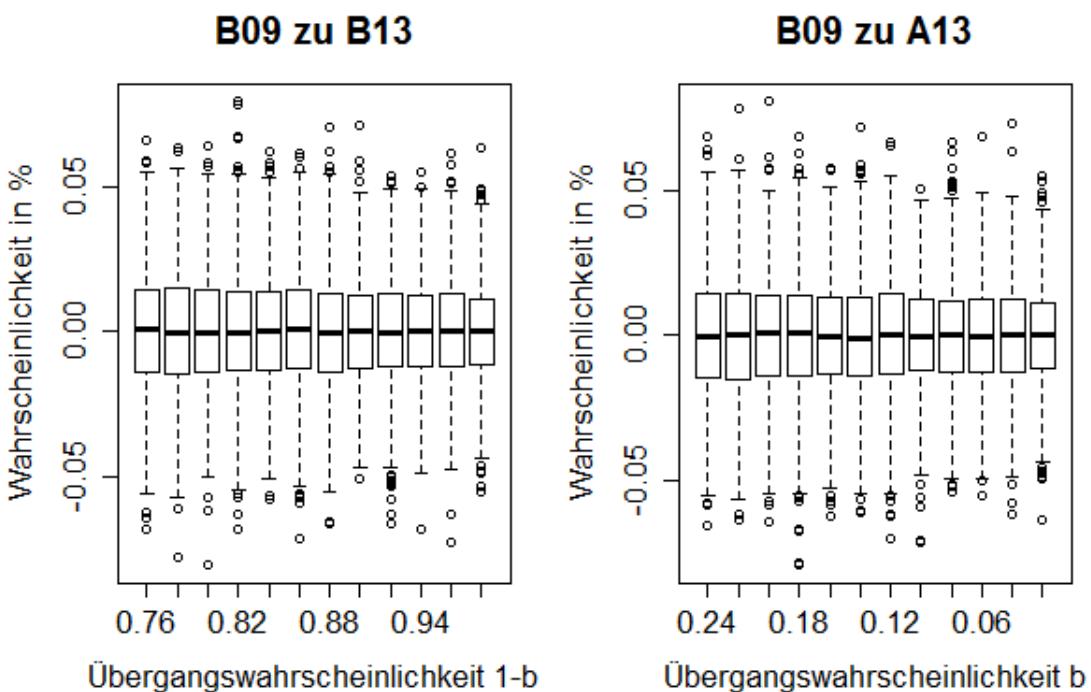


Abbildung 9: Simulation 10: Boxplots mit je 1000 Iterationen für verschiedene Übergangswahrscheinlichkeiten b .

Linke Grafik: Die Werte stellen die Übergangswahrscheinlichkeiten aus dem simulierten Modell aus der Zelle B09-B13 dar, zentriert um den jeweiligen Mittelwert.

Rechte Grafik: Die Werte stellen die Übergangswahrscheinlichkeiten aus dem simulierten Modell aus der Zelle B09-A13 dar, zentriert um den jeweiligen Mittelwert.

Begutachtet man nun beispielsweise die beiden Boxplots für $b = 0,02$ und somit automatisch $1 - b = 0,98$ genauer, erkennt man, dass es jeweils nur Spiegelungen voneinander sind. Dies trifft auch auf die anderen Boxplots bei jedem beliebigen b zu. Der Grund hierfür ist die Struktur in der 2x2 Tabelle, in der eine binomialverteilte Struktur vorliegt. Jeder Wähler, der in der ersten Wahl B09 gewählt hat und in der zweiten Wahl nicht A13

wählt, wählt automatisch B13. Da die Summe der beiden Wahrscheinlichkeiten immer eins sein muss und die jeweiligen Boxplots um den eigenen Mittelwert zentriert wurden, kommt es zu dieser Spiegelung. Schaut man sich nun aber den Verlauf der Boxplots mit kleiner werdenden b an, erkennt man, dass die Spannweite der Boxplots abnimmt. Da es jedoch auch möglich ist, dass die Verringerung der Spannweite durch die zeitgleiche Steigung der Wahrscheinlichkeit von $1 - b$ entsteht, muss die Simulation als nächstes in einer 3x3 Tabelle getestet werden.

4.3.2 3x3 Tabellen

Um dies nun auch in einer 3x3 Tabelle zu testen, wurde eine Simulation mit 1000 Iterationen, $n = 1000$ und folgenden Übergangswahrscheinlichkeiten u und Wahlanteile w durchgeführt:

$$u = \begin{pmatrix} 0,8 & 0,1 & 0,1 \\ 0,2 & 0,7 & 0,1 \\ b & 0,3 & 0,6-b \end{pmatrix} \quad w = (0,3; 0,4; 0,3)$$

In Abbildung 10 sind Boxplots für verschiedene b gegeben. Auch hier sieht man eine leichte Verkleinerung der Spannweite in der linken Grafik mit fallenden b . Der Einfluss ist jedoch extrem gering und kann als Problem für die hohen negativen Werte vernachlässigt werden. Es bleibt also dabei, dass die hohen negativen Werte hauptsächlich durch geringe Wahlanteile bei der ersten Wahl entstehen.

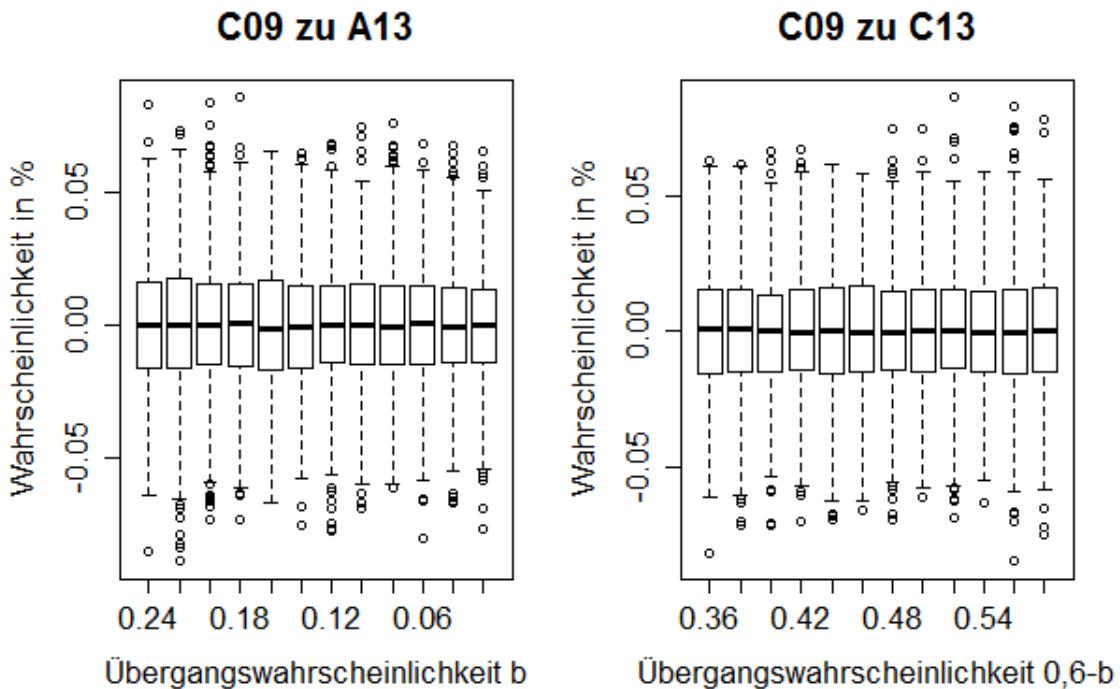


Abbildung 10: Simulation 11: Boxplots mit je 1000 Iterationen für verschiedene Übergangswahrscheinlichkeiten b .

Linke Grafik: Die Werte stellen die Übergangswahrscheinlichkeiten aus dem simulierten Modell aus der Zelle B09-B13 dar, zentriert um den jeweiligen Mittelwert.

Rechte Grafik: Die Werte stellen die Übergangswahrscheinlichkeiten aus dem simulierten Modell aus der Zelle B09-A13 dar, zentriert um den jeweiligen Mittelwert.

4.4 Lösungsversuch mit Bootstrap

Da die Werte um den tatsächlichen Wert der Übergangstabelle streuen, wird als nächstes versucht, mit Hilfe von Bootstrap eine Verbesserung der Modellschätzung zu erreichen. Hierbei handelt es sich um ein Resampling-Verfahren, um eine Stichprobe aus dem Datensatz zu ziehen. Hierzu werden aus den $m = 650$ Bezirken 650 neue Bezirke mit Zurücklegen gezogen. Es werden dann die jeweiligen Wahlanteile von der ersten Wahl und die Stimmzahlen bei der zweiten Wahl im jeweiligen Bezirk gezogen und als neuer Bezirk betrachtet. Ein Bezirk kann nun also in der Stichprobe überhaupt nicht, einmal oder mehrfach vorkommen. Mit diesem neuen Datensatz berechnet man anschließend das Modell neu. [6]

Um dieses Verfahren zu testen, wurde eine erneute Simulation in einer 3x3 Tabelle durchgeführt. Hierfür wurde ein Datensatz mit Wahlanteilen und Stimmzahlen für $m = 650$ Bezirke wie bisher simuliert. Als nächstes wurden aus diesen 650 Bezirken durch Bootstrap 650 neue Bezirke gezogen. Mit dieser Stichprobe wurde dann das Modell berechnet. Die Ziehung und die Modellberechnung wurden dann 100 mal wiederholt und aus diesen Ergebnissen ein Mittelwert gebildet. Das Ergebnis gibt dann die neue Schätzung für diesen einen simulierten Datensatz wieder. Das gesamte Prozedere wurde dann mit 600 Iteratio-

nen durchgeführt. Das heißt, es wurden 600 Datensätze simuliert und das neue Modell mit Bootstrap berechnet. Die Übergangstabelle u und die Wahlanteile w stellen sich hierbei wie folgt dar:

$$u = \begin{pmatrix} 0,8 & 0,1 & 0,1 \\ 0,2 & 0,7 & 0,1 \\ 0,1 & 0,3 & 0,6 \end{pmatrix} \quad w = (0,3; 0,4; 0,3)$$

In Abbildung 11 sind die Ergebnisse dieser Simulation zu sehen. Die geschätzten Werte wurden hierbei um die tatsächliche Übergangswahrscheinlichkeit zentriert. Die linke Grafik stellt hierbei die Werte der geschätzten Tabelle aus der Zelle C09-A13 um den Wert 0,1 zentriert dar. In der rechten Grafik sind die Wahrscheinlichkeiten der Zelle A09-A13 zentriert um den Wert 0,8 zu sehen. Trotz des Bootstraps sind die Abweichungen der geschätzten Werte von den wahren Werten deutlich sichtbar. Um dies zu verdeutlichen, wurde die identische Simulation ohne Bootstrap gerechnet, was in Abbildung 12 sichtbar ist. Im Vergleich der beiden Abbildungen erkennt man, dass durch das Bootstrap-Verfahren keine Verbesserung zustande kommt. Die Schätzungen streuen weiter um den wahren Wert, ähnlich wie ohne Bootstrap.

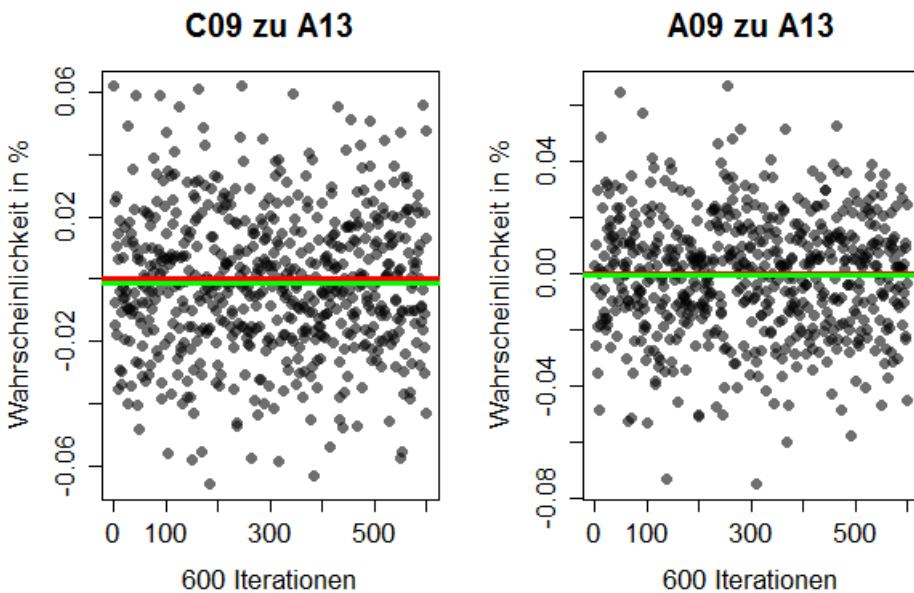


Abbildung 11: Simulation 12: Streudiagramme mit je 600 Iterationen. Die Werte wurden nach dem Bootstrap-Verfahren geschätzt und sind um den jeweiligen wahren Wert in der Übergangstabelle u zentriert. In rot ist die Nulllinie und in grün der Mittelwert eingezeichnet. Linke Grafik: Die Werte stellen die Übergangswahrscheinlichkeiten aus dem simulierten Modell aus der Zelle C09-A13 dar. Rechte Grafik: Die Werte stellen die Übergangswahrscheinlichkeiten aus dem simulierten Modell aus der Zelle A09-A13 dar.

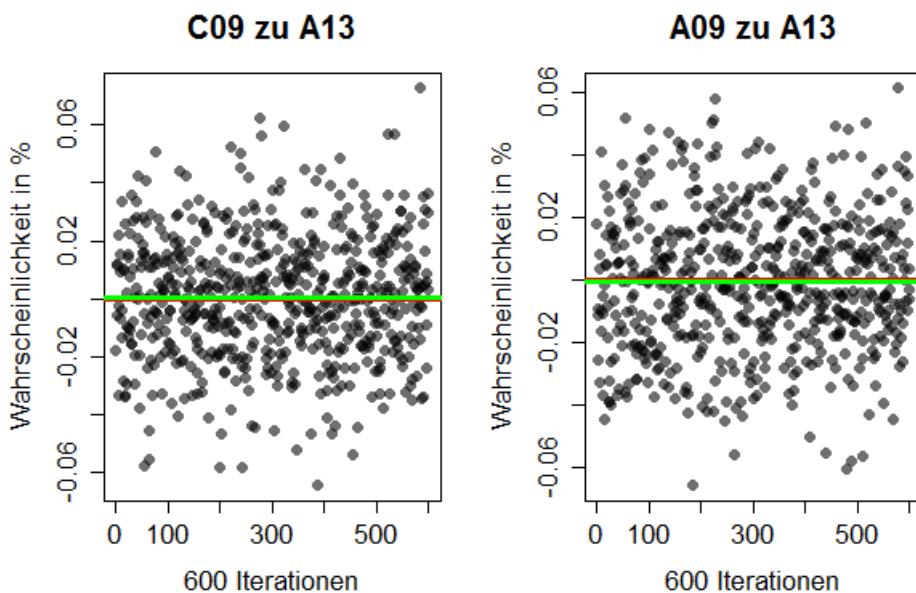


Abbildung 12: Simulation 12: Streudiagramme mit je 600 Iterationen. Die Werte sind um den jeweiligen wahren Wert in der Übergangstabelle u zentriert (ohne Bootstrap). In rot ist die Nulllinie und in grün der Mittelwert eingezeichnet.

Linke Grafik: Die Werte stellen die Übergangswahrscheinlichkeiten aus dem simulierten Modell aus der Zelle C09-A13 dar.

Rechte Grafik: Die Werte stellen die Übergangswahrscheinlichkeiten aus dem simulierten Modell aus der Zelle A09-A13 dar.

4.5 Bootstrap bei den Bundestagswahlen

Zum Abschluss soll das Bootstrap-Verfahren noch bei den offiziellen Wahldaten der Bundestagswahlen getestet werden. Hierzu wurden aus den 650 Bezirken durch Resampling 650 neue Bezirke mit Zurücklegen gezogen und das Modell berechnet. Dieser Schritt wurde 1000 mal wiederholt und die Ergebnisse wurden gemittelt. Das Ergebnis ist in Tabelle 14 zu sehen. Im Vergleich zur Tabelle 5 treten kaum Veränderungen auf. Das Bootstrap-Verfahren liefert also bei diesem Datensatz im Mittel das gleiche Ergebnis wie ohne Bootstrap. Somit bleiben auch die hohen negativen Werte erhalten.

In Abbildung 13 sind zwei Streudiagramme zu Visualisierung dargestellt. In der linken Grafik sind die Wahrscheinlichkeiten für die Zelle Linke09-CSU13, in der rechten Grafik die Wahrscheinlichkeiten der Zelle CSU09-CSU13. Die Beobachtungen streuen wieder um den Mittelwert, welcher in grün eingezeichnet ist. Dennoch treten in der linken Grafik fast durchgängig negative Werte auf. Es kann also keine Verbesserung dieser Daten durch das Verfahren festgestellt werden.

		2013								
		CSU	SPD	FDP	Grüne	Linke	Piraten	AfD	Sonstige	NW
2 0 0 9	CSU	88,9	5,9	0,5	-1,3	-1,3	-0,9	3,5	7,1	-2,4
	SPD	19,4	70,4	1,3	-0,6	2,0	1,2	7,4	2,7	-3,7
	FDP	32,9	3,1	42,0	6,7	-2,7	0,8	8,5	-5,4	14,0
	Grüne	-5,1	22,0	9,7	67,9	8,6	3,5	-4,2	1,3	-3,8
	Linke	-22,8	25,6	-21,4	11,5	36,2	4,3	7,4	8,5	50,8
	Piraten	13,0	15,4	2,7	45,2	11,3	34,9	5,6	3,0	-31,2
	Sonstige	24,4	12,3	-19,3	-6,2	3,7	2,5	12,6	35,0	35,0
	NW	0,5	5,1	-1,6	-3,8	1,6	1,5	0,2	1,7	94,8

Tabelle 14: Geschätzte Wählerwanderung in Prozent der Bundestagswahlen 2009-2013 nach multinomialen Modell mit Bootstrap.

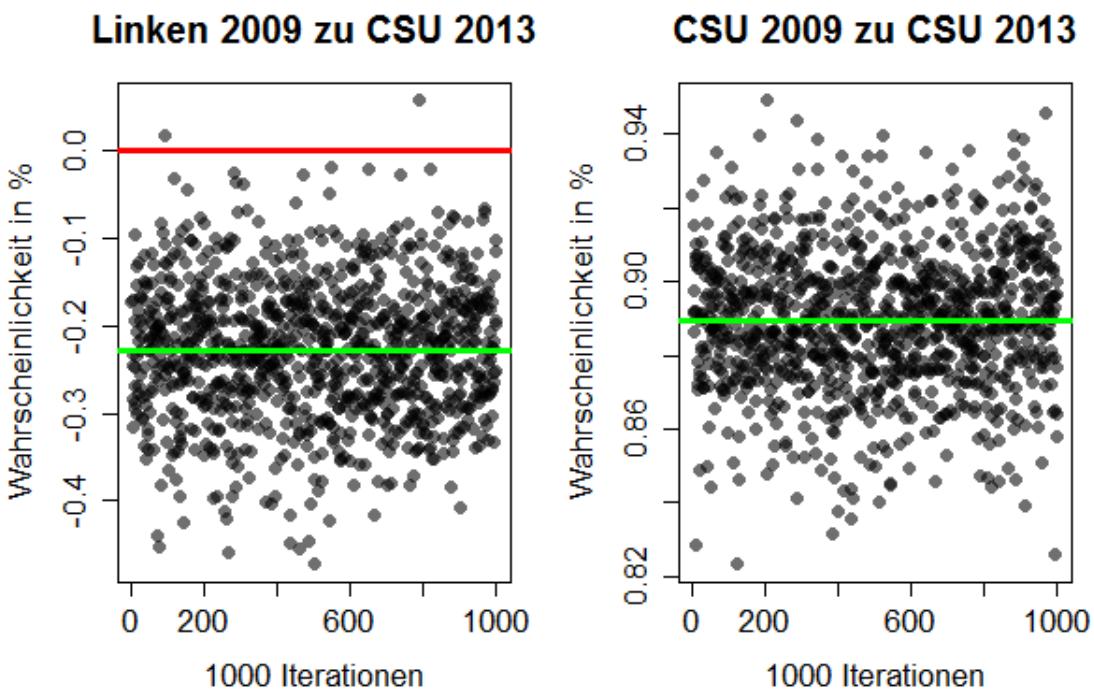


Abbildung 13: Bundestagswahlen: Streudiagramme mit je 1000 Iterationen. Die Werte wurden mit dem multinomialen Modell nach dem Bootstrap-Verfahren geschätzt. In grün ist der Mittelwert und in rot die Nulllinie eingezeichnet.

Linke Grafik: Die Werte stellen die Übergangswahrscheinlichkeiten aus der Zelle Linken09-CSU13 dar.

Rechte Grafik: Die Werte stellen die Übergangswahrscheinlichkeiten aus der Zelle CSU09-CSU13 dar.

5 Fazit

Wie in den Simulationen festgestellt wurde, kann das Modell die tatsächlichen Wählerwanderung im Mittel gut schätzen. Gerade bei Parteien, die in der ersten Wahl einen hohen Anteil haben, ist die Streuung sehr gering und die Schätzungen sind recht genau. Probleme entstehen bei Parteien mit geringeren Wahlanteilen. Beispielsweise haben bei den Bundestagswahlen 2009 und 2013 vor allem die Linke und die Piraten, die geringe Wahlanteile bei der Wahl 2009 hatten, Probleme mit hohen negativen Werten. Auch die CSU hat negative Werte, jedoch ist hier der Wahlanteil 2009 deutlich höher, weshalb davon ausgegangen werden kann, dass diese Werte nur leicht verzerrt geschätzt werden und die wahren Werte lediglich knapp über 0 liegen. Beide vorgestellten Lösungsversuche konnten die negativen Werte nicht verhindern. Dennoch kann mit der Methode von Ambühl eine Übergangstabelle erstellt werden, deren Werte innerhalb des Intervalls liegen.

6 Literatur

- [1] <http://www.muenchen.de/rathaus/Stadtverwaltung/Kreisverwaltungsreferat/Wahlen-und-Abstimmungen/Wahlergebnisse-Bundestagswahl.html> (letzter Abruf am 14. Juni 2014 um 13:23 Uhr).
- [2] <http://www.muenchen.de/rathaus/Stadtverwaltung/Kreisverwaltungsreferat/Wahlen-und-Abstimmungen/Wahlergebnisse.html> (letzter Abruf am 14. Juni 2014 um 13:23 Uhr).
- [3] Fahrmeir, Ludwig / Kneib, Thomas / Lang, Stefan: Regression. Modelle, Methoden und Anwendungen, München Oldenburg & Innsbruck, 2009.
- [4] R Development Core Team (2014). R: eine freie Programmiersprache für statistisches Rechnen und statistische Grafiken. <http://www.R-project.org>.
- [5] Dr. Mathias Ambühl, Methoden zur Rekonstruktion von Wählerströmen aus Aggregatdaten, Neuchatel 2003.
- [6] <http://www.statistik.lmu.de/institut/lehrstuhl/semwiso/schaetzenteisten2-ss10/skript/ST2-ss10-kap05.pdf> (letzter Abruf am 26. Juni 2014 um 16:20 Uhr).
- [7] Alan Genz, Frank Bretz, Tetsuhisa Miwa, Xuefei Mi, Friedrich Leisch, Fabian Scheipl, Bjoern Bornkamp, Martin Maechler, Torsten Hothorn, mvtnorm: Multivariate Normal and t Distributions. R package version 0.9-99992. <http://cran.r-project.org/web/packages/mvtnorm/index.html> (letzter Abruf am 02. Juli 2014 um 09:37 Uhr).
- [8] Thomas W. Yee(2014), VGAM: Vector Generalized Linear and Additive Models. R package version 0.9-4. <http://cran.r-project.org/web/packages/VGAM/index.html> (letzter Abruf am 26. Juni 2014 um 16:20 Uhr).
- [9] David B. Dahl(2014), xtable: Export tables to LaTeX or HTML. R package version 1.7-3. <http://cran.r-project.org/web/packages/xtable/index.html> (letzter Abruf am 08. Juli 2014 um 13:10 Uhr).

7 Abbildungsverzeichnis

Abb. 1	Streudiagramm der Wahrscheinlichkeiten, dass ein Wähler der Linken zur CSU wechselt.	8
Abb. 2	Simulation 1, mit $w = (0,82; 0,18)$: Linke Grafik: Streudiagramm der 1000 um den Mittelwert zentrierten Werte für die geschätzte Wahrscheinlichkeit, dass ein Wähler von der Partei B09 zu A13 wechselt. In rot ist die Nulllinie eingezeichnet. Rechte Grafik: Streudiagramm der 1000 um den Mittelwert zentrierten Werte für die geschätzte Wahrscheinlichkeit, dass ein Wähler von der Partei A09 bei A13 bleibt. In rot ist die Nulllinie eingezeichnet.	11
Abb. 3	Simulation 2, mit $w = (0,95; 0,05)$: Linke Grafik: Streudiagramm der 1000 um den Mittelwert zentrierten Werte für die geschätzte Wahrscheinlichkeit, dass ein Wähler von der Partei B09 zu A13 wechselt. In rot ist die Nulllinie eingezeichnet. Rechte Grafik: Streudiagramm der 1000 um den Mittelwert zentrierten Werte für die geschätzte Wahrscheinlichkeit, dass ein Wähler von der Partei A09 bei A13 bleibt. In rot ist die Nulllinie eingezeichnet.	12
Abb. 4	Simulation 3: Boxplots mit je 1000 Iterationen für verschiedene Wahlanteile c. Linke Grafik: Die Werte stellen die Übergangswahrscheinlichkeiten aus dem simulierten Modell aus der Zelle B09-A13 dar. Rechte Grafik: Die Werte stellen die Übergangswahrscheinlichkeiten aus dem simulierten Modell aus der Zelle A09-A13 dar.	14
Abb. 5	Simulation 4: Das linke Streudiagramm bildet die um den Mittelwert zentrierten Ergebnisse des Modells der Zelle C09-A13. Das rechte Streudiagramm bildet die um den Mittelwert zentrierten Ergebnisse des Modells der Zelle A09-A13.	15
Abb. 6	Simulation 5: Boxplots mit je 1000 Iterationen für verschiedene Wahlanteile c. Linke Grafik: Die Werte stellen die Übergangswahrscheinlichkeiten aus dem simulierten Modell aus der Zelle A09-A13 dar. Rechte Grafik: Die Werte stellen die Übergangswahrscheinlichkeiten aus dem simulierten Modell aus der Zelle C09-A13 dar.	16

Abb. 7	Simulation 8: Boxplots mit je 1000 Iterationen für verschiedene Bezirksgrößen n. Linke Grafik: Die Werte stellen die Übergangswahrscheinlichkeiten aus dem simulierten Modell aus der Zelle B09-A13 dar. Rechte Grafik: Die Werte stellen die Übergangswahrscheinlichkeiten aus dem simulierten Modell aus der Zelle A09-A13 dar.	18
Abb. 8	Simulation 9: Boxplots mit je 1000 Iterationen für verschiedene Bezirksgrößen n. Linke Grafik: Die Werte stellen die Übergangswahrscheinlichkeiten aus dem simulierten Modell aus der Zelle A09-A13 dar. Rechte Grafik: Die Werte stellen die Übergangswahrscheinlichkeiten aus dem simulierten Modell aus der Zelle C09-A13 dar.	19
Abb. 9	Simulation 10: Boxplots mit je 1000 Iterationen für verschiedene Übergangswahrscheinlichkeiten b. Linke Grafik: Die Werte stellen die Übergangswahrscheinlichkeiten aus dem simulierten Modell aus der Zelle B09-B13 dar, zentriert um den jeweiligen Mittelwert. Rechte Grafik: Die Werte stellen die Übergangswahrscheinlichkeiten aus dem simulierten Modell aus der Zelle B09-A13 dar, zentriert um den jeweiligen Mittelwert.	20
Abb. 10	Simulation 11: Boxplots mit je 1000 Iterationen für verschiedene Übergangswahrscheinlichkeiten b. Linke Grafik: Die Werte stellen die Übergangswahrscheinlichkeiten aus dem simulierten Modell aus der Zelle B09-B13 dar, zentriert um den jeweiligen Mittelwert. Rechte Grafik: Die Werte stellen die Übergangswahrscheinlichkeiten aus dem simulierten Modell aus der Zelle B09-A13 dar, zentriert um den jeweiligen Mittelwert.	22
Abb. 11	Simulation 12: Streudiagramme mit je 600 Iterationen. Die Werte wurden nach dem Bootstrap-Verfahren geschätzt und sind um den jeweiligen wahren Wert in der Übergangstabelle u zentriert. In rot ist die Nulllinie und in grün der Mittelwert eingezeichnet. Linke Grafik: Die Werte stellen die Übergangswahrscheinlichkeiten aus dem simulierten Modell aus der Zelle C09-A13 dar. Rechte Grafik: Die Werte stellen die Übergangswahrscheinlichkeiten aus dem simulierten Modell aus der Zelle A09-A13 dar.	23

- Abb. 12 Simulation 12: Streudiagramme mit je 600 Iterationen. Die Werte sind um den jeweiligen wahren Wert in der Übergangstabellen zentriert (ohne Bootstrap). In rot ist die Nulllinie und in grün der Mittelwert eingezeichnet. Linke Grafik: Die Werte stellen die Übergangswahrscheinlichkeiten aus dem simulierten Modell aus der Zelle C09-A13 dar. Rechte Grafik: Die Werte stellen die Übergangswahrscheinlichkeiten aus dem simulierten Modell aus der Zelle A09-A13 dar. 24
- Abb. 13 Bundestagswahlen: Streudiagramme mit je 1000 Iterationen. Die Werte wurden mit dem multinomialen Modell nach dem Bootstrap-Verfahren geschätzt. In grün ist der Mittelwert und in rot die Nulllinie eingezeichnet. Linke Grafik: Die Werte stellen die Übergangswahrscheinlichkeiten aus der Zelle Linken09-CSU13 dar. Rechte Grafik: Die Werte stellen die Übergangswahrscheinlichkeiten aus der Zelle CSU09-CSU13 dar. 25

8 Tabellenverzeichnis

Tab. 1	Wählerwanderung Bundestagswahl 2009-2013 nach multinomialem Modell	3
Tab. 2	Wählerwanderung Bundestagswahl 2009-2013 nach multinomialem Modell mit Korrektur nach Ambühl	4
Tab. 3	Wählerwanderung Landtagswahl 2009-2013 nach multinomialem Modell	5
Tab. 4	Wählerwanderung Landtagswahl 2009-2013 nach multinomialem Modell mit Korrektur nach Ambühl	6
Tab. 5	Wählerwanderung Bundestagswahl 2009-2013 nach multinomialem Modell mit Modellanpassung	7
Tab. 6	Kontingenztabelle für den ersten Bezirk mit vier Parteien bei der Wahl 2009 und vier Parteien bei der Wahl 2013. Es wird eine Übergangstabelle für den Bezirk simuliert.	9
Tab. 7	Simulation 1: In der linken Tabelle sind die Mittelwerte und in der rechten Tabelle die Varianzen der 1000 Iterationen des Modells in den jeweiligen Zellen aufgeführt. $w = (0,82; 0,18)$	10
Tab. 8	Simulation 2, mit $w = (0,95; 0,05)$: In der linken Tabelle sind die Mittelwerte und in der rechten Tabelle die Varianzen der 1000 Iterationen des Modells in den jeweiligen Zellen aufgeführt.	12
Tab. 9	Simulation 1 ($w = (0,82; 0,18)$) und 2 ($w = (0,95; 0,05)$): Tabelle mit Lagemaßen der beiden Simulationen für die Werte der Zellen A09-A13 und B09-A13.	13
Tab. 10	Simulation 4: In der linken Tabelle sind die Mittelwerte und in der rechten Tabelle die Varianzen der 1000 Iterationen des Modells in den jeweiligen Zellen aufgeführt.	15
Tab. 11	Simulation 6, $n = 1000$: In der linken Tabelle sind die Mittelwerte und in der rechten Tabelle die Varianzen der 1000 Iterationen des Modells in den jeweiligen Zellen aufgeführt.	17
Tab. 12	Simulation 7, $n = 100$: In der linken Tabelle sind die Mittelwerte und in der rechten Tabelle die Varianzen der 1000 Iterationen des Modells in den jeweiligen Zellen aufgeführt.	17
Tab. 13	Simulation 6 ($n = 1000$) und 7 ($n = 100$): Tabelle mit Lagemaßen der beiden Simulationen für die Werte der Zellen A09-A13 und B09-A13	18
Tab. 14	Wählerwanderung Bundestagswahl 2009-2013 nach multinomialem Modell mit Bootstrap	25

9 Inhalt der CD-ROM

- Im Ordner ‘Grafiken‘ befinden sich alle in der Bachelorarbeit vorkommenden Grafiken im png-Format.
- Im Ordner ‘R_Code‘ befinden sich folgende Dateien
 - Die Datei ‘Daten_BW.RData‘ enthält den bearbeiteten Datensatz aus dem statistischen Praktikum zur Bundestagswahl.
 - Die Datei ‘Daten_LW.RData‘ enthält den bearbeiteten Datensatz aus dem statistischen Praktikum zur Landtagswahl.
 - Die Datei ‘Funktionen.R‘ enthält alle für die Arbeit selbstgeschriebenen Funktionen für R.
 - Die Datei ‘Simulationen.R‘ enthält alle Simulationen die für die Bachelorarbeit durchgeführt wurden.
 - Die Datei ‘BW_Multinomial.R‘ enthält den R-Code für die Modellschätzungen bei den Bundestagswahlen.
 - Die Datei ‘LW_Multinomial.R‘ enthält den R-Code für die Modellschätzungen bei den Landtagswahlen.
- Die Bachelorarbeit als PDF Datei.

Erklärung

Hiermit versichere ich, dass ich meine Bachelorarbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

München, den 09.07.2014:
(Unterschrift)