



Centraal Bureau voor de Statistiek

Divisie Methodologie en Kwaliteit Sector

Methodologie

Postbus 4481

6401 CZ Heerlen

**Ein Vergleich verschiedener Schätzmethoden von
Abstimmungsübergänge mit einer Bewerbung in der
Niederländische Nationalwahlen**

Carin van der Ploeg

Samenvatting: Er zijn vele uitdagingen in het schatten van het stemgedrag van kiezers. Eén hiervan is het schatten van verschuivingen in het kiesgedrag, engagaamde kiezersstromen. Deze kiezersstromen kunnen worden gepresenteerd in een transitie matrix. Modellierungs- und Umfragemethode zijn twee verschillende algemene methoden die deze transitie matrix kunnen schatten. Onderzoek laat zien dat beide methoden voordelen en nadelen hebben. In dit paper wurden deze twee methoden elkaar vergeleken op basis van de voorspelde tweede verkiezingsuitslag getroffen. Das Wort verkregen Tür de uitslag van het eerste Moment te vermenigvuldigen traf de Transitie matrix. De focus van het onderzoek vindt plaats binnen de context van de Nederlandse tweede kamerverkiezingen van 2003 en 2006, waar de uitslagen kunnen worden geobserveerd op gemeenteniveau. De resultaten laten zien dat op gemeenteniveau er realistische schattingen kunnen worden gemaakt van het stemgedrag. Tevens blijkt dat Surveytechnieken en modelleringstechnieken elkaar kunnen aanvullen, wat de validiteit van de resultaten verder versterkt. Dit biedt een ging uitgangspunt voor verdere Studien.

Trefwoorden: **Stemgedrag,** **Lineaire** **Regression,** **Kwadratisch**
Programmierer, Latente Markov ketens, Iteratief Proportioneel Fitten, Ontrouwe-trouwe
kiezersmodel, Nederlands KiezersOnderzoek (NKO)

Zusammenfassung: Die Einschätzung des Wahlverhaltens ist mit vielen Herausforderungen verbunden. Eine davon ist die Schätzung von Verschiebungen der Abstimmungspräferenzen, so genannten Abstimmungsübergängen. Diese Abstimmungsübergänge können in einer Übergangsmatrix dargestellt werden. Die Abschätzung von Änderungen im Abstimmungsverhalten kann auf zwei Arten erfolgen, durch Modellierung und mit einer Umfrage. Untersuchungen zeigen, dass beide ihre Vorteile, aber auch ihre Nachteile haben. Wir vergleichen diese beiden Ansätze unter Verwendung der Wahlergebnisse im ersten Moment multipliziert mit dieser Übergangsmatrix, um die Wahlergebnisse im zweiten Moment zu schätzen. Der Schwerpunkt dieser Forschung liegt auf den niederländischen Nationalwahlen 2003 und 2006, bei denen die Ergebnisse auf Gemeindeebene beobachtet werden können. Dieses Papier zeigt, dass realistische Schätzungen der Wahlergebnisse auf Gemeindeebene vorgenommen werden können.

*Schlüsselwörter: Wählen Übergänge, Linear Regression, Quadratisch
Programmierung, Latent Markov, Iterative Proportionalanpassung, MoverStayer-Modell,
Dutch National Survey (NKO)*

Danksagung

Ich möchte meinen Vorgesetzten, Dr. René Bekker, Dr. Jarl Kampen und Dr. Frank Pol für ihre Anleitung, Kommentare und Ratschläge während der Praktikumszeit. Besonders die Zusammenarbeit mit Dr. Jarl Kampen und die Diskussionen und seine Beiträge waren interessant und nützlich. Der Beitrag meines zweiten Betreuers Prof. Dr. Bert Kersten war auch sehr wertvoll. Ich möchte mich bei dr. Peter Kruiskamp für seine Liebe, sein Feedback und seine Geduld und auch meine anderen Kollegen vom CBS für ihre Unterstützung und Kommentare. Darüber hinaus bin ich Joost Bosman für seine verschiedenen nützlichen Beiträge sehr dankbar.

Inhalt

Inhalt.....	4
1. Einführung	6
1.1 Einführung in die Organisation	8
1.2 Fragestellung.....	9
2. Etablierte Methodik zur Beschreibung des Verhaltens von Mover-Stayer	11
2.1 Einführung.....	11
2.2 Benchmark-Methoden	13
2.3 Umfragemethode	13
2.4 Modellbasierte Methodik	16
2.4.1 Lineare Regression.....	16
2.4.2 Quadratische Programmierung	17
2.4.3 Latente Klassenanalyse	18
2.4.4 Iterative Proportionalanpassung	21
2.4.5 Kombinationen	21
2.4.6 Kurze Zusammenfassung	23
3. Anwendung im niederländischen Hoheitsgebiet	24
3.1 Einführung.....	24
3.2 Verfügbare Daten	27
3.3 Passmaße	28
4. Ergebnisse.....	31
4.1 Statistik der Anpassungsgüte	31
4.2 Benchmark-Methoden	33
4.3 NKO	36
4.4 Quadratische Programmierung	36
4.5 Latente Klassenanalyse	37
4.6 Iterative Proportionalanpassung	37
4.7 Kombinationen	38
4.8 Diskussion über die Wahlen 2003-2006	39
4.9 Sitzverteilung 2003-2006	41
5. Diskussion und weitere Forschung	45
5.1 Ökologische Folgerung	45
5.1.1 El-Schätzer	46
5.2 Diskussion.....	47
6. Literaturverzeichnis.....	50
7. Anhang A	55
7.1 Analyse der Parteien	55

7.2	Übergangsmatrizen 1998-2002, 2003-2006	56
7.3	Vergleich aller Modelle pro Wahljahr	68
7.4	Erweiterungen des Mover-Stayer-Modells	72
7.5	Verwendete Programme und Code	73

1. Einleitung

Politiker, Politikwissenschaftler und Medien sind sehr daran interessiert, die Übergänge in der Parteipräferenz der Wähler zwischen den Wahlen abzubilden. Dies zeigt sich in der großen Menge an Literatur zu *Mover* und *Steher* das ist in der Literatur zu finden und geht zumindest auf Leo Goodmans Artikel von 1961 zurück, *Statistische Methoden für das Mover-Stayer-Modell*. Viele niederländische Meinungsumfragen, die von Maurice D'Hondt, Interview NSS bis TNS NIPO, variieren, veröffentlichen regelmäßig Vorhersagen über politische Übergänge und die Unterstützung politischer Parteien. Ziel war es immer, das Abstimmungsverhalten so genau wie möglich abzuschätzen. Dies ist eine große Herausforderung, da aufgrund der Vertraulichkeit der Stimmabgabe nur Daten auf aggregierter Ebene (national, regional oder kommunal) verfügbar sind, während wir das Verhalten von Einzelpersonen wissen möchten. Viele Forscher versuchen, den Mangel an Daten zum individuellen Abstimmungsverhalten durch die Organisation von Umfragen zu überwinden. Aus einer Reihe von Gründen, die nachstehend erläutert werden, sind Umfragen jedoch nicht unbedingt eine Garantie für zuverlässige Ergebnisse.

Hintergrund für diese spezielle Studie ist eine Arbeit von Keller und Ten Cate (1977). Ihr Artikel mit dem Titel "De verschuiving van kiezersvoorkeur" warf die interessante Frage auf, wohin die Wähler gehen. Sie konstruierten eine einzige Matrix, um zu versuchen, die Abstimmungsübergänge von 1972 bis 1977 in den Niederlanden zu beschreiben. Wir haben diese Arbeit als Ausgangspunkt für die Entwicklung und den Vergleich mehrerer neuer und bestehender Modelle genommen. Das ursprüngliche Modell von Keller und Ten Cate wurde ebenfalls in diese Arbeit aufgenommen. Diese mathematischen Modelle zielen darauf ab, die Übergänge von Parteipräferenzen auf der Grundlage aggregierter Informationen auf Gemeindeebene zu beschreiben.

Um sich einen schnellen Überblick über das Thema dieser Dissertation zu verschaffen, wird ein kurzes hypothetisches Beispiel vorgestellt, das von Keller & ten Cate (1977) gegeben wurde. In diesem Beispiel entspricht wie in den Analysen in diesem Papier die Gesamtebene den Wahlergebnissen in den Gemeinden. Stellen Sie sich vor, es gibt zwei Parteien (Partei A und Partei B) und zwei Städte (Gemeinde 1 und Gemeinde 2). Wie aus Tabelle 1 hervorgeht, ist Partei A die größte Partei in beiden Gemeinden und hat 2003 in Gemeinde 1 1000 Stimmen und in Gemeinde 2 300 Stimmen erhalten. Die Stimmen für Partei B können auf die gleiche Weise abgeleitet werden. In Gemeinde 2 erhielt Partei B 2003 200 Stimmen und gewann 2006 70 Stimmen, was einer Gesamtzahl von 270 Stimmen entspricht. Bei den Wahlen 2006 verliert Partei A in beiden Gemeinden Stimmen, während Partei B Unterstützung erhält.

Tabelle 1. Zwei Parteien, zwei Städte und zwei Wahlen				
	2003		2006	
	Partei A.	Party B	Partei A.	Party B
Gemeinde 1	1000	800	780	1020
Gemeinde 2	300	200	230	270

Um die Übergangsrate von Partei A zu Partei B zu berechnen, können wir die

Anzahl der abgegebenen Stimmen in einer einfachen Gleichung. Lassen Sie P_{AA} sei die Übergangsrate von

Partei A zu Partei A, die sogenannten Steher, und P_{BA} sei der Übergang von Partei B zu Partei A, den sogenannten Movern. Dann,

$$1000 \times P_{AA} + 800 \times P_{BA} = 780$$

$$300 \times P_{AA} + 200 \times P_{BA} = 230$$

Wir können eine einfache Berechnung durchführen und die Übergangsraten finden, die zu diesen Gleichungen gehören.

Man kann leicht überprüfen, ob sich die Übergangsraten zu 1 summieren. Diese Werte zeigen, dass 70% der

Unterstützung für Partei A verblieben sind, aber dass 30% ihrer Anhänger 2006 zu Partei B gewechselt sind. Die

Unterstützer für Partei B waren noch loyaler da blieben sogar 90% bei dieser party.

$$P_{AA} = 0,70 \quad P_{AB} = 0,30$$

$$P_{BA} = 0,10 \quad P_{BB} = 0,90$$

Der letzte Schritt besteht darin, die Übergangsraten in einer Matrix darzustellen, die eine sehr einfache

Übergangsmatrix ergibt. Diese Matrix wird auf alle Gemeinden und mehr Parteien ausgedehnt und ist das Zentrum der Analyse.

	2006	
2003 A.		B.
EIN	0,7	0,3
B.	0,1	0,9

Diese Werte können dann wieder in die Gleichung eingetragen werden, um die ersten Abstimmungsergebnisse zu erhalten:

$$1000 \times 0,7 + 800 \times 0,1 = 780$$

$$1000 \times 0,3 + 800 \times 0,9 = 1020$$

$$300 \times 0,7 + 200 \times 0,1 = 230$$

$$300 \times 0,3 + 200 \times 0,9 = 270$$

Dieses einfache und gut ausgewählte Beispiel liefert eine einzigartige Übergangsmatrix. Das Hinzufügen von mehr Parteien und mehr Kommunen erhöht jedoch schnell die Komplexität des Problems und erfordert die Anwendung ausgefeilterer Modellierungsansätze. Die Analyse zeigt, dass es möglich ist, eine eindeutige Übergangsmatrix in einer 1x1-, 2x2- und 3x3-Situation zu identifizieren. Wenn das Problem auf 3 Parteien und 3 Städte erhöht wird, ist es immer noch möglich, eine eindeutige Übergangsmatrix zu schätzen. Es ist nicht mehr möglich, eine eindeutige Lösung zu finden, wenn die Komplexität zunimmt. Unter strengen Voraussetzungen sind gewöhnliche kleinste Quadrate tatsächlich unvoreingenommen. Wenn jedoch Abhängigkeiten auftreten, ist es leicht zu zeigen, dass es unendlich viele Lösungen und keine eindeutige Übergangsmatrix gibt.

Als Alternative zu modellbasierten Ansätzen, die auf einer aggregierten Datenanalyse basieren, wird häufig Umfrageforschung eingesetzt, um detaillierte Informationen von einem repräsentativen Teil der Bevölkerung zu sammeln. Anhand der Antworten einer Stichprobe versuchen die Forscher, Rückschlüsse auf den Rest der Bevölkerung zu ziehen. Diese Methode liefert sehr nützliche Daten und wird in vielen Anwendungen verwendet. Dieses Verfahren hat jedoch auch seine Nachteile. Es wird viel Forschung betrieben, um diese Probleme zu überwinden, aber einige bleiben noch bestehen. Daher war es interessant, mögliche Kombinationen sowohl mit aggregierten als auch mit Umfrageinformationen zu untersuchen. In den vorliegenden Analysen von Abstimmungsübergängen werden die modellbasierten Ansätze mit dem umfragebasierten Ansatz verglichen, um mehr Einblick in Abstimmungsübergänge zu erhalten.

1.1 Einführung in die Organisation

Statistik Niederlande hat die Aufgabe, Statistiken zu sammeln, zu bearbeiten und zu veröffentlichen, die in der Praxis sowie für politische und Forschungszwecke relevant sind. Neben der Verantwortung für die nationalen (offiziellen) Statistiken ist Statistics Netherlands auch für die Erstellung europäischer (Gemeinschafts-) Statistiken verantwortlich. Rechtsgrundlage für die niederländische Statistik ist das "Gesetz über das Zentralamt für Statistik" vom November 2003 (Staatsblad, 2003, S. 516).¹ Es gibt zwei Einrichtungen: eine in Voorburg in der Nähe des niederländischen Regierungszentrums und eine in Heerlen im Süden der Niederlande.

¹ "Het Centraal Bureau voor de Statistiek (CBS) heeft tot taak het verzamelen, bewerken en publiceren van statistieken zehn behoeve van praktijk, beleid en wetenschap. Naast de verantwoordelijkheid voor de nationale (officiële) statistieken ist het CBS ook belast met de productie van Europese (communautaire) statistieken. De wettelijke grondslag voor het CBS ist de "Wet op het Centraal Bureau voor de Statistiek" vom 20. November 2003 "(Staatsblad, 2003, S. 516)

Statistics Netherlands forscht aktiv, um ihre Methoden im Allgemeinen zu verbessern, und macht Forschung und Entwicklung zu einer wichtigen Aufgabe von Statistics Netherlands. Der Sektor Methodik und Qualität (DMK) befindet sich an beiden Standorten. Dieser Sektor trägt zur Wissensentwicklung der Statistik Niederlande bei und zielt darauf ab, ihren statistischen Prozess kontinuierlich zu verbessern.

1.2 Forschungsfrage

Ziel dieser Studie ist es, die Methodik zur Schätzung von Abstimmungsübergängen zu verbessern und zu vergleichen. Es wird ein umfassender Vergleich zwischen der modellbasierten Methodik und der umfragebasierten Methodik durchgeführt. Die Wahlergebnisse von zwei aufeinanderfolgenden Wahlen werden verwendet, um die Abstimmungsübergänge zwischen zwei aufeinander folgenden Wahlen zu schätzen. Der erste Zeitpunkt ist das Wahljahr 2003 und der zweite Zeitpunkt ist das Wahljahr 2006. Die Qualität der geschätzten Übergangsmatrizen wird bewertet, indem die Abstimmungsergebnisse zum zweiten Zeitpunkt anhand der in der ersten Wahl und der Übergangsmatrix enthaltenen Informationen geschätzt werden ausgewertet. Diese Schätzungen des zweiten Augenblicks können mit den tatsächlichen Wahlergebnissen verglichen werden, und ein Maß für die Anpassung der Schätzungen (z. B. das Wahrscheinlichkeitsverhältnis) kann berechnet werden.

Diese Methode zur Schätzung der Wahlergebnisse mittels einer Übergangsmatrix ist in einem breiteren Feld von Kohortenstudien mit kategorialen Variablen anwendbar. Zwei Vorteile der vorliegenden Studie, die spezifisch für Wahlstudien sind, werden hier hervorgehoben. Erstens gibt es einen Mehrwert für das spezifische Gebiet der Wahlstudien, wenn eine Antwort darauf gegeben wird, ob die verwendete Methodik zur Schätzung von Abstimmungsübergängen auf Gemeindeebene geeignet ist. Zweitens gibt es eine mögliche Verbesserung der Genauigkeit bei der Schätzung des Abstimmungsverhaltens. Im Vergleich zum NKO kann eine Verbesserung mithilfe modellbasierter Methoden möglich sein.

Die obigen Überlegungen führen zu der Hauptforschungsfrage:

*Inwieweit können Abstimmungsübergänge in den Niederlanden auf Gemeindeebene sein?
genau beschrieben, wo die notwendigen Übergangsmatrizen fehlen?*

Um diese Forschungsfrage zu beantworten, wird sie in folgende Unterfragen unterteilt:

Welcher modellbasierte Ansatz eignet sich am besten zur Schätzung von Abstimmungsübergängen auf Gemeindeebene?

Wie können sich Erhebungsmethodik und modellbasierte Methodik ergänzen und verbessern?

Die folgenden allgemeinen Hypothesen werden gezogen:

Die Schätzung des Abstimmungsverhaltens anhand geschätzter Übergangsmatrizen ist eine zuverlässige Alternative für die Methodik zur Ableitung von Übergängen in der Umfrageanalyse.

Für jede Gemeinde gibt es unterschiedliche Übergangsmatrizen, die teilweise durch die demografische Variabilität erklärt werden können

Der Rest dieses Papiers ist wie folgt strukturiert: In Kapitel 2 wird der Stand der Technik der etablierten Methodik und die Erläuterung aller verwendeten Modelle vorgestellt, gefolgt von einer Einführung in den Kontext dieser Forschung. In Kapitel 3 werden die Dateneigenschaften und Einschränkungen vorgestellt. In Kapitel 4 werden die Ergebnisse und die Analyse der Ergebnisse vorgestellt, und schließlich enthält Kapitel 5 die Schlussfolgerungen und Themen für die weitere Forschung.

2. Etablierte Methodik zur Beschreibung des Verhaltens von Mover-Stayer

2.1 Einleitung

Wahlstudien bilden eine eigene sozialwissenschaftliche Disziplin. Ein Großteil der Forschung im Rahmen von Wahlstudien konzentrierte sich auf Zweiparteiensysteme und weniger auf Mehrparteiensysteme. In den letzten Jahren wurde mehr Forschung zu letzterem Thema betrieben, aber dies ist im Vergleich zur Forschung zu Zweiparteiensystemen (Quinn) immer noch gering *et al.*, 1999). In diesem Kapitel wird ein kurzer Überblick über wichtige Beiträge sowohl zu Zweiparteien- als auch zu Mehrparteiensystemen gegeben. Der Schwerpunkt der Übersicht liegt auf der im Rahmen dieser Forschung verwendeten Methodik, es werden jedoch auch andere verwandte Forschungen kurz erwähnt, um ein vollständiges Bild der im Bereich der Wahlstudien durchgeführten Forschung zu erhalten.

Um das Abstimmungsverhalten zu beschreiben, müssen kategoriale Daten untersucht werden. Eine kategoriale Variable ist einfach eine Variable, für die die Messskala aus einer Reihe von Kategorien besteht. Sie können auf einer nominalen (nicht geordneten) oder ordinalen (geordneten) Skala liegen (Agresti, 1996, S. 2-3; Kampen & Swyngedouw, 2000: S. 87). Insbesondere in den Sozialwissenschaften sind Variablen jedoch häufig nicht direkt beobachtbar, und latente Variablen werden zur Erklärung der Realität verwendet. Wenn beispielsweise nur Wahlergebnisse für zwei aufeinanderfolgende Wahlen bekannt sind, kann die Übergangsmatrix als latente Variable modelliert werden. Um dies klarer zu sehen, muss zuerst eine Notation eingeführt werden. Die Formeln werden in bedingten Wahrscheinlichkeiten ausgedrückt. In einer Markov-Kette werden Übergangswahrscheinlichkeiten ausgedrückt als

$$P(X_n = j | X_{n-1} = i) = p_{ij}$$

Wo die Parteien aus der ersten Wahl mit Index vertreten sind $i = (1, \dots, m_{ich})$

und die Parteien aus der zweiten Wahl mit Index $j = (1, \dots, m_{ii})$. In diesem Artikel haben wir

Wenden Sie die folgende Grundnotation für die in der Forschung verwendeten Wahrscheinlichkeiten an:

$p_i = \Pr(i_{ch})$, die Wahrscheinlichkeit der stimmberechtigten Partei i_{ch} bei der ersten Wahl bei nationalen Niveau,

$p_j = \Pr(j)$, die beobachtete Wahrscheinlichkeit der stimmberechtigten Partei j beim zweiten Wahl auf nationaler Ebene,

$p_{ij} = \Pr(i, j)$, die unbeobachteten Gelenkwahrscheinlichkeiten der beobachteten Ränder,

$$p_{j|i} = \Pr(j|i) = \frac{\Pr(i,j)}{\Pr(ich)}, \text{ das } \textit{unbeobachtet} \text{ Anteil der Wähler}$$

Übertragung von der Partei *ich* feiern *j*,

$$p_{i|g} = \Pr(i|g) = \frac{\Pr(ich\ G)}{\Pr(G)}, \text{ der beobachtete Anteil der Wähler, die gewählt haben}$$

für Party *ich* in der Gemeinde *g* (Wahlpräferenz zum ersten Mal zum Beispiel 2003),

$$p_{j|g} = \Pr(j|g) = \frac{\Pr(j, g)}{\Pr(G)}, \text{ der beobachtete Anteil der Wähler, die gewählt haben}$$

für Party *j* in der Gemeinde *g* (Wahlpräferenz im zweiten Augenblick, zum Beispiel 2006).

Nach Einführung dieser Grundnotation kann die allgemeinere Situation in der folgenden Grundformel ausgedrückt werden:

$$\Pr(j|i) \notin \sum_{ich}^m \Pr(i|g) \times \Pr(j|ich), \quad (2.1)$$

Die Werte von $\Pr(j|g)$ und $\Pr(i|g)$ bekannt sind und der Wert von $\Pr(ji)$ (das

Elemente der Übergangsmatrix) müssen geschätzt werden. Wie zu sehen ist, kann diese Formel auch im obigen Beispiel verwendet werden. Analog zu Keller & ten Cate (1977) wird eine einzige nationale Übergangsmatrix verwendet, um alle Übergänge zu berechnen. Dies bedeutet, dass die Elemente von $\Pr(j|ich)$ sind unabhängig von der Gemeinde *G*. Die bedingten Wahrscheinlichkeiten

oder Übergangsgewichte $\Pr(ji)$ Geben Sie die Abstimmungsübergänge zwischen zwei Wahlen an.

$\Pr(i|g)$ und $\Pr(j|g)$ und $\Pr(ji)$ kann nicht direkt aus der Wahl abgeleitet werden

Ergebnisse. Im Allgemeinen lautet die folgende Formel für die Elemente der Übergangsmatrix:

$$\Pr(ji) \neq \Pr(i, j) / \Pr(ich), \quad (2.2)$$

Die gemeinsamen Wahrscheinlichkeiten $\Pr(i, j)$ kann im Allgemeinen bei der Verwendung nicht beobachtet werden

aggregierte Daten. Sie müssen anhand empirischer Daten oder wie bei der Umfrageanalyse mithilfe von Rückrufdaten zur Schätzung der zugrunde liegenden (latenten) Übergangsmatrix geschätzt werden.

Ein mathematischer Ansatz zur Schätzung der zugrunde liegenden Übergangsmatrix ist die lineare Regression, eine intuitive und einfache Technik, die jedoch in der Praxis zu unrealistischen Ergebnissen führen kann. Die einfache Technik der linearen Regression kann als quadratisches Programmierproblem ausgedrückt werden, indem mehr Einschränkungen hinzugefügt werden, um realistischere Ergebnisse zu erzielen. Latente Markov-Theorie und latente Klassenmodelle ermöglichen es, latente Variablen mit den in beobachteten Informationen zu kombinieren

Variablen bei der Schätzung von Abstimmungsübergängen. Eng verwandt mit dem Ansatz der latenten Klasse ist die loglineare Modellierung. Diese mathematischen Techniken und die Erhebungsmethode bilden die Grundlage für diese Studie.

2.2 Benchmark-Methoden

Anstatt Übergänge zu schätzen, können wir einfach (konzeptionell vertretbare) Werte in die Übergangsmatrix eintragen und die Anpassung eines solchen Ad-hoc- oder Benchmark-Modells bewerten. Zwei Methoden in dieser Studie werden lediglich als Benchmark verwendet. Die erste Methode ist das Unabhängigkeitsmodell oder das sogenannte perfekte Mobilitätsmodell. Dies bedeutet, dass die Abstimmung des Mitglieds im zweiten Augenblick völlig unabhängig von seiner Abstimmung im ersten Augenblick ist. Dies kann durch diese Formel dargestellt werden:

$$\Pr(j|i) = \Pr(j). \quad (2.3)$$

Formel (2.3) besagt, dass die Wahrscheinlichkeit der Wahl für die Partei j gegebene Wahl der Partei i bei der vorherigen Wahl ist gleich der Wahrscheinlichkeit, dass über die Partei abgestimmt wird j .

Die Information, dass der Wähler zuerst für die Partei gestimmt hat i wird nicht berücksichtigt.

Die zweite Methode wird als 100% Steher-Modell bezeichnet. Dies bedeutet, dass die Wähler ihrer Partei vollkommen treu sind. Ein Wähler, der für die Partei stimmt j im ersten Moment wird im zweiten Moment für genau dieselbe Partei gestimmt, was zu der folgenden Formel führt:

$$\Pr(j|i) = \begin{cases} 1 & \text{if } j = i \\ 0 & \text{wenn } j \neq i \end{cases} \quad (2.4)$$

Die Formel (2.4) besagt, dass die Wähler einfach ihre ursprüngliche Wahl wiederholen werden i . Aufgrund des Fehlens einer „goldenen“ Messmethode ist es interessant, diese Extreme zum Vergleich zu verwenden und die Leistung anderer Modelle zu bewerten.

2.3 Erhebungsmethode

Es gibt verschiedene Forschungsbüros, die alle erklären wollen, was der niederländische Wähler bei den kommenden Wahlen tun wird. Die niederländischen Parlamentswahlstudien sind die größte durchgeführte Umfrage und auf die Beschreibung und das Verständnis des niederländischen Wahlverhaltens spezialisiert. Die niederländische Wahlumfrage (*Nederlands Kiesers Onderzoek*)

wird bei jeder nationalen Wahl durchgeführt. Die letzte Umfrage wurde von Statistics Netherlands durchgeführt.

Ähnliche Erhebungen werden in vielen Demokratien durchgeführt, um einen Einblick in den wichtigsten demokratischen Prozess zu erhalten, bei dem es um die Abstimmung geht. In der niederländischen Wahlumfrage (NKO) werden die Befragten gefragt, was ihre jüngste Wahlpräferenz sowie ihre Präferenz bei der vorherigen Wahl ist. Mit diesen Daten wird die

reale Übergänge zwischen Parteien hin und her können in einer Übergangsmatrix ausgedrückt werden. Für diese Studie ist das Buch relevant, das über die niederländischen Wahlen von 2006 "Een verdeeld Electoraat De Tweede Kamerverkiezingen van 2006" unter Verwendung der NKO-Forschung von Aarts, Van der Kolk & Rosema (2007) geschrieben wurde. Die Übergangsmatrix finden Sie in Aarts *et al.* ((2007: S. 224).

Der Mehrwert von Umfragen besteht darin, dass es sich um eine Beobachtungsmethode handelt, die eine allgemeine Fähigkeit zur Beschreibung der Merkmale einer großen Bevölkerung bietet. Da die Messung der Variablen auf individueller Ebene erfolgt, können Umfragen auch die Wahlpräferenzen zwischen zwei Wahlen auf Mikroebene analysieren. Übergänge können daher in beide Richtungen analysiert werden. Es ist zum Beispiel zu sehen, wer bei der ersten Wahl in der PvdA und der zweiten Wahl in der CDA gewählt hat. Wir können aber auch sehen, wer bei der ersten Wahl für die CDA und bei der zweiten Wahl für die PvdA gestimmt hat. Auf diese Weise können wir sehen, wie viele Wähler zwischen zwei Parteien gewechselt haben, und mehr als nur die gesamten Nettoübergänge zwischen zwei Parteien visualisieren, was bedeutet, dass wir nur die Partei sehen können, die bei den letzten Wahlen die größte Unterstützung erhalten hat. Dann sehen wir nur den Nettoübergang von PvdA und CDA. Umfragen sind in gewisser Weise flexibel, da sie eine Vielzahl von Informationen sammeln können. Es ist daher möglich, zwischen soziodemografischen Eigenschaften, Einstellungen, Werten, Überzeugungen und (früherem) Wahlverhalten zu korrelieren und mögliche Erklärungsfaktoren für ein bestimmtes Verhalten zu berücksichtigen.

Das Problem bei dieser Art von Daten ist, dass Rückrufdaten oft sehr unzuverlässig sind (Weir, 1975). Es gibt mehrere Effekte, die ihre Konten unzuverlässig machen (Voogt, 2004; Keller & ten Cate, 1977; Upton, 1977):

Die Befragten erinnern sich nicht an ihre vorherige Abstimmung.

Die Befragten wollen nicht launisch sein, daher besteht die Tendenz, die gleiche Wahl der Partei zwischen der vorherigen und der aktuellen Wahl zu beantworten.

Die Befragten antworten falsch und nennen die Partei, für die sie bei den Kommunal-, Provinz- oder Europawahlen gestimmt haben.

Die Befragten geben an, dass sie gewählt haben, aber bei den vorherigen Wahlen nicht gewählt haben, um eine gesellige, akzeptierte Antwort zu geben.

Die Befragten möchten zur Gewinnerseite gehören und falsch antworten, indem sie eine Partei angeben, die bei den Wahlen gewonnen hat.

Darüber hinaus gibt es mehrere Effekte, die die Umfrageergebnisse erheblich verzerren. Es ist jedoch nicht einfach, genau zu bestimmen, wie groß diese Vorspannung ist. Voogt (2004) hat die niederländische nationale Abstimmungserhebung untersucht. gemäß

Für seine Forschung sind die wichtigsten Effekte nicht reagierende Verzerrung, Antwortverzerrung und Stimuluseffekt (Voogt, 2004). Non-Responsive Bias ist der wichtigste Effekt und für eine Unterrepräsentation bestimmter Gruppen in der Bevölkerung verantwortlich. Umfangreiche Literatur zu diesem Problem zeigt, dass Nicht-Befragte im Allgemeinen höher ausgebildet, jünger und häufiger ledig sind und in den städtischen Gebieten überrepräsentiert sind (siehe eine Liste der veröffentlichten Literatur Voogt, 2004: S. 35). Die Antwortverzerrung wird auch als "Antwort- und Gedächtniseffekt" bezeichnet und spiegelt die oben genannten Gründe für unzuverlässige Ergebnisse wider (Bethelehem & Kersten, 1986). Die letzte Verzerrung (Stimulus-Effekt) tritt insbesondere in der Panel-Forschung auf, bei der die Befragten mindestens zweimal befragt werden (Greenwald *et al.*, 1987). Personen, die normalerweise nicht vorhaben, ihre Stimme abzugeben, werden wählen, weil sie an dem Teil der Wahlforschung teilgenommen haben, der vor den Wahlen stattgefunden hat. Für eine weitere Übersicht über die Stärken und Schwächen von Umfragen siehe Voogt, (2004). Aufgrund des Fehlens einer „goldenen Messmethode“ ist es nicht einfach, genau zu bestimmen, welche Auswirkungen diese Ursachen für Verzerrungen haben. Daher müssen wir uns auf Methoden stützen, die diese Fehler nur teilweise korrigieren können, so dass die Bestimmung der Übergangsmatrix auf der Grundlage von Erhebungen nicht einfach ist, was zu Schwierigkeiten führen kann.

Die teilnehmenden Befragten werden gefragt, für welche Partei sie bei den aktuellen nationalen Wahlen und bei den vorherigen Wahlen gestimmt haben. Die Kreuztabelle dieser beiden Erinnerungsvariablen geteilt durch die Zeilensummen ergibt den angeforderten Übergang

Matrix, die sogenannten bedingten Wahrscheinlichkeiten $Pr(j|i)$. Genau wie in Aarts *et al.* ((2007)

Die Stichprobe wird mit einem Gewichtungsfaktor multipliziert, um die wahrheitsgetreueste Darstellung der niederländischen Bevölkerung zu erhalten.² Um zu testen, ob die NKO-Übergangsmatrix die Realität darstellt, werden die geschätzten Werte auf der rechten Seite der Formel (2.1) angewendet, wo anschließend die Wahlergebnisse des zweiten Augenblicks pro Gemeinde aus den Ergebnissen des ersten Augenblicks vorhergesagt werden.

Eine interessante Ergänzung zur Überwindung von Verzerrungen in der Umfrageforschung wird von Barbosa und Goldstein (2000) vorgestellt. Ihr Beitrag ist nützlich in der Längsschnittforschung. Sie präsentieren eine Methode, die die Goldstein-Methode, ein mehrstufiges Zeitreihenmodell, um diskrete Variablen für normalverteilte Antworten erweitert. Dies bedeutet, dass die Antworten der Befragten einer Normalverteilung folgen. Diese Methode ist besser in der Lage, den tatsächlichen Anteil von Umzugsunternehmen und Stehern abzuschätzen und strategische Abstimmungen zu berücksichtigen (Barbosa & Goldstein, 2000).

² Die Werte dieser Übergangsmatrix unterscheiden sich geringfügig von den Werten im Buch von Aarts *et al.* (2007), da die Matrix in diesem Artikel die korrekt gerundeten Werte verwendet und sich die Zeilensummen zu 1 addieren. Dies ist bei der im Buch von Aarts *et al.*

2.4 Modellbasierte Methodik

2.4.1 Lineare Regression

Keller und Ten Cate stellten einen interessanten Ansatz zur Schätzung von Abstimmungsübergängen vor. Sie verwendeten eine lineare Regression, um die Abstimmungsübergänge innerhalb der Niederlande zu schätzen. Ihre Methode kann als Stichprobe angesehen werden, bei der Stichprobengröße und Populationsgröße übereinstimmen. Da es nicht möglich ist, den Wähler zu fragen, was er oder sie bei den vorherigen Wahlen gewählt hat, sind die Ergebnisse nur auf aggregierter Ebene verfügbar, dh auf Gemeindeebene. 1977 gab es in den Niederlanden 842 Gemeinden. Die Eigenschaft, dass die Wahlergebnisse pro Gemeinde auf mehrere Parteien verteilt sind, wird folgendermaßen verwendet: Wenn eine Partei verliert und eine andere Partei in einer Gemeinde gewinnt, gehen sie davon aus, dass zwischen diesen beiden Parteien ein Überlauf auftritt. Die Übergangswahrscheinlichkeiten gelten für alle Gemeinden als konstant. Sie vergleichen ihre Ergebnisse mit der damals größten Umfrage, der Intromart-Umfrage, und den tatsächlichen Abstimmungsergebnissen und kommen zu dem Schluss, dass ihre Ergebnisse im Vergleich relativ gut sind. Sie stellen jedoch fest, dass einige Wahrscheinlichkeiten extrem hoch sind und dass es im Vergleich zur Intromart-Umfrage viele Wahrscheinlichkeiten von Null gibt, bei denen sie nicht erwartet werden (Keller & ten Cate, 1977).

Die Grundmethode von Keller und Ten Cate ist das lineare Wahrscheinlichkeitsmodell

$$\Pr(j) = \sum_{ich} \Pr(ich) \times \Pr(j|i) + \varepsilon(j) \quad (2.5)$$

Wo in ihrem Fall $\Pr(ich)$ ist die Anzahl der Stimmen auf Partei *ich* 1972 wurde $\Pr(j)$ ist der Anzahl der Stimmen auf Partei *j* im Jahr 1977, $\Pr(j|i)$ bezeichnen die Übergangswahrscheinlichkeiten und $\varepsilon(j)$ ist ein zufälliger Interferenzterm oder -fehler. Man kann die Übergangsmatrix abschätzen

Verwenden von gewöhnlichen kleinsten Quadraten, dh durch Minimieren der quadratischen Diskrepanzen zwischen den tatsächlichen und geschätzten Werten, was einer Minimierung von entspricht

$$\sum_j \sum_G (p_{jg} - c_j - \sum_{ich} \hat{p}_{ij})^2 \quad \text{Opsterhoff \& Vaart, 2003, S. 40}.$$

Keller und Tencate stellen jedoch fest, dass sich die Bevölkerung zwischen zwei Wahlen ändert. Die Wähler, die zu Hause geblieben sind, sind eine zusätzliche Kategorie, und die neuen / zurückgetretenen Wähler werden als zusätzliche Herkunftspartei hinzugefügt. Dies fügt Fehler in die Berechnung ein. Sie erzwingen Gleichheit in der Gesamtzahl der Stimmen pro Gemeinde, aber die geschätzten Wahrscheinlichkeiten $\Pr(i, j)$ sind manchmal außerhalb der Grenzen, was bedeutet, dass sie nicht immer zwischen 0 und 1 notwendig sind. Das Problem bei dieser Methode ist, dass negative Wahrscheinlichkeiten und Wahrscheinlichkeiten, die größer als eins sind, erhalten werden können (Keller & ten Cate, 1977, S. 4). Mit den Einschränkungen, dass die Schätzungen zwischen 0 und 0 liegen müssen

1, und die Einschränkung, dass die Übergangswahrscheinlichkeiten das Problem auf 1 summieren müssen, wird zu einem quadratischen Programmierproblem.

2.4.2 Quadratische Programmierung

Wie bereits erwähnt, sind die Schätzungen von $\Pr(i, j)$ ist manchmal out-of-bound. Mehrere Autoren haben das Mittel der eingeschränkten Regression angepasst, um dieses Problem zu beheben (Telser, 1963). Quadratische Programmierung ist (auch) eine der angebotenen Lösungen (siehe zB Judge & Takayama, 1966; Irwin & Meeler, 1969; McCarthy & Ryan, 1977 und Keller & Ten Cate, 1977). Die Kritik an der Verwendung dieser Methode stammt von Cleave et al. (1995), der angibt, dass diese Methode die Formulierung eines geeigneten Modells vermeidet. Cleave et al., 1995, S. 5). Die Autoren bieten jedoch keine alternative Lösung oder eine eingehendere Kritik an und bieten daher wenig Diskussionsgrundlage.

Die quadratische Programmierung erzwingt Einschränkungen für die Zielfunktion.

$$\text{Mindest} \left\| \Pr(i, j) - (\hat{\beta}_{ich} - \Pr(i, j)) \right\|^2$$

Unter:

$$\beta \geq 0$$

Die Einschränkungen sind:

1. $\sum_{i=1}^{10} \Pr(i, j) = \Pr(ich)$ und $\sum_{j=1}^{10} \Pr(i, j) = \Pr(j)$.
2. $\sum_{j=1}^m \Pr(j/i) = 1$.

Wo:

ich ist Parteipräferenz im ersten Moment (zum Beispiel 2003),

j ist Parteipräferenz im zweiten Moment (zum Beispiel 2006),

m ist die Gesamtzahl der politischen Parteien,

$\hat{\beta}$ erzeugt die erforderlichen Übergangskoeffizienten unter den Einschränkungen 1 und 2.

Die erste Einschränkung besagt, dass die Ränder von $\Pr(i, j)$ muss sich zu den beobachteten Margen addieren (dies bedeutet das Wahlergebnis). Die zweite Einschränkung besagt, dass die Wahrscheinlichkeiten für die Partei zu stimmen j gegeben ich muss sich zu 1 addieren. Denken Sie daran, dass dies bei der linearen Regression nicht der Fall war.

Der Hauptunterschied zur normalen linearen Regression besteht darin, dass die Matrix mit den Ergebnissen von 2003 und die Matrix mit den Ergebnissen von 2006 zu einem Vektor mit zusammengesetzt werden

Länge 100. β enthält die Koeffizienten, über die die Einschränkungen des Ziels

Funktion werden erzwungen. Die einzige Einschränkung für β ist, dass es gleich oder größer als 0 sein sollte. Dies bedeutet, dass alle kleinen quadratischen Programmierprobleme beschrieben wurden

pro Partei in einem großen Vektor. Zusätzliche Einschränkungen könnten daher durchgesetzt werden. β ist die Matrix mit dem $\Pr(i, j)$. Die Übergangswahrscheinlichkeiten werden in einen Vektor mit der Länge 100 eingegeben. Durch die Neuformulierung von OLS entsteht ein quadratisches Programmierproblem mit dem Ziel, den mittleren quadratischen Fehler zu minimieren (Tijms & Ridder, 2003). Sowohl die statistischen Softwareprogramme S-plus als auch R können verwendet werden, um tatsächliche Berechnungen der linearen quadratischen Programmierung zu erzeugen.

2.4.3 Latente Klassenanalyse

Das traditionelle Markov-Modell ist nicht geeignet, um langfristige soziale Daten wie das Wahlverhalten vorherzusagen (Van de Pol & Langheime, 2004, S. 2). Die Analyse latenter Klassen eignet sich jedoch sehr gut zur Analyse von Längsschnittdaten (Langheime & Pol, 1990; Van de Pol & Langheime, 2004; Vermunt et al., 1999) und aufgrund ihrer Fähigkeit, nicht beobachtbare Variablen zu verwenden, zur Schätzung von Abstimmungsübergängen. Das zugrunde liegende Modell für latente Markov-Ketten ist das Mover-Stayer-Modell, das im Laufe der Zeit erweitert wurde und in diesem Abschnitt vorgestellt wird. Das erste Vorkommen des Mover-Stayer-Modells war 1955. Der Artikel von Goodman (1961) präsentiert einen Ansatz, der als Mover-Stayer-Modell bekannt geworden ist. Es wurde in einem Jobwechselkontext (Blumen) eingeführt (et al., 1955). Es ist ein Sonderfall der Mischung Markov-Modell (Poulsen, 1982). Die grundlegende Struktur dieses Modells besteht darin, dass es Bewegende und Stehende gibt. In der Bevölkerung gibt es Menschen, die bei aufeinanderfolgenden Wahlen für dieselbe Partei stimmen, und Menschen, die mit Wahrscheinlichkeit zu einer anderen Partei wechseln $p(j)$, unabhängig von der vorherigen Wahl. Es bietet eine elegante

Erklärung für den offensichtlichen Mangel an Korrelation zwischen Parteiwechsel. Goodman verwendete den Begriff der Halbindependenz (Unabhängigkeit zwischen den Wahlen), um die Wahlen von 1959 bis 1974 zu schätzen (Goodman, 1968). Seine Ergebnisse zeigten, dass das Mover-Stayer-Modell für die Kurzzeitanalyse tatsächlich eine interessante Möglichkeit zur Visualisierung von Abstimmungsübergängen darstellt. Langfristig ist es nicht geeignet, da die Wähler im Gegensatz zur kurzfristigen über einen langen Zeitraum ihre Wahlpräferenz ändern. Daher ist der geschätzte Anteil der Stehenden zu groß (Upton, 1977, S. 2). Weitere Ausführungen zu Erweiterungen des Mover-Stayer-Modells finden Sie in Abschnitt 7.4 des Anhangs

EIN.

Um eine weitere Verbesserung des Mover-Stayer-Modells zu erzielen, können wir Messfehler in das Modell einbeziehen. Dies kann mit latenten oder indirekt gemessenen Markov-Modellen erfolgen. Latente Markov-Modelle eignen sich gut für die individuelle Beschreibung von Änderungen in kategorialen Daten. Das Schlüsselement dieses Modells ist, dass es beobachtete Veränderungen teilweise als Messfehler latenter Veränderungen (indirekt gemessene Veränderungen) betrachtet. In diesem Modell gibt es keine zeitkonstante latente Variable, da angenommen wird, dass keine unbeobachtete latente Heterogenität vorliegt (Van de Pol & Langheime, 2004, S. 4). Methoden wie parametrische, nichtparametrische und semiparametrische Modelle eignen sich nur für kontinuierliche Beobachtungen. Da es sich bei diesem Problem um diskrete Daten handelt und keine genauen Übergangszeiten verfügbar sind, werden sie häufig als Paneldaten bezeichnet (Cook *et al.*, 2002).

Van de Pol und Langheime (1990) beschreiben ein gemischtes latentes Markov-Klassenmodell, das ursprünglich von Wiggins (1973) und Poulsen (1982) entwickelt wurde. Das Modell kann Fluktuationsraten zwischen zwei oder mehr Panelwellen beschreiben, was bei Abstimmungsdaten der Fall ist. In diesem Modell gibt es mehr latente Klassen als beobachtete Kategorien und ist nur identifizierbar, wenn genügend Umsatz beobachtet wird (Van de Pol & Langheime, 1990, S. 29).

Das in dieser Studie verwendete Modell basiert auf der Arbeit von Clogg und Goodman (1984). Es heißt Latent Class Analysis (LCA). Diese Methode sucht nach den vernünftigsten Werten der bedingten Wahrscheinlichkeiten $\Pr(j|i)$, angesichts der Wahlergebnisse in beiden Momenten. Die Parameter können unter Verwendung der maximalen Wahrscheinlichkeit geschätzt werden. Diese Funktion kann für die vollständige Tabelle $\Pr(i, j, g)$ (formuliert werden). i, j, g , die Anzahl der Stimmen auf Partei *ich* und *j* in der Gemeinde *G* für mindestens einen von beiden Momenten. Die Maximum-Likelihood-Schätzung ist der Parameterwert, für den die beobachteten Daten die größte Wahrscheinlichkeit haben. Dieser Parameter wird durch die Wahrscheinlichkeitsfunktion mit ihrem Maximalwert angegeben (Agresti, 1996, S. 8-9). Aufgrund der analytischen Unmöglichkeit einer direkten Berechnung der maximalen Wahrscheinlichkeit (ML) kann dieses iterative Verfahren verwendet werden. Hier ist es nicht notwendig, die hessischen Matrizen zu berechnen.

Die Wahrscheinlichkeitsfunktion (L) ist gegeben durch:

$$L = \prod_{i,j,g} \Pr(i, j, g) = \prod_G \Pr(G) \prod_{ich} \Pr(i|G) \prod_j \Pr(j|ich, G) \quad (2.7)$$

Wo f entspricht Frequenzen. f_{ijg} sind die Wahlergebnisse für jede Gemeinde. $f(i|g)$ sind die Wahlergebnisse für die Partei *ich* im ersten Augenblick gegeben Gemeinde *G*. $f(j|i, g)$ sind die Wahlergebnisse für die Partei *j* im zweiten Augenblick gegebene Wahlpräferenz im ersten Augenblick in der Gemeinde *G*.

Dies kann mithilfe des Expectation Maximization-Algorithmus (EM) geschätzt werden, für den PANMARK entwickelt wurde. Aufgrund der Implementierung von Heterogenität im Modell ist die Ökobilanz realistischer als die Mover-Stayer-Varianten (Van de Pol & Langheime, 1990, S. 9; Vermunt) *et al.*, 2006, S. 5).

Das 1977 erschienene Papier von Dempster *et al.* führt den verallgemeinerten EM-Algorithmus ein. Es handelt sich um eine iterative Berechnung von Maximum-Likelihood-Schätzungen für den Fall, dass die Daten unvollständig sind (Dempster) *et al.*, 1977, S. 1). Es basiert auf einer relativ einfachen Idee, um unvollständige Datenprobleme zu lösen (Dempster *et al.*, 1977):

Ersetzen Sie fehlende Werte durch geschätzte Werte

Schätzen Sie die Parameter

Schätzen Sie fehlende Werte erneut unter der Voraussetzung, dass die neue Parameterschätzung korrekt ist

Schätzen Sie die Parameter erneut

Wiederholen Sie die Parameterschätzung bis zur Konvergenz

Im E-Schritt wird die bedingte Erwartung einer Log-Likelihood-Funktion für vollständige Daten der fehlenden Daten basierend auf den beobachteten Daten und der aktuell geschätzten Erwartung der fehlenden Parameter angegeben. Im M-Schritt werden die Parameter aktualisiert, damit die Erwartung maximiert wird. Das Maximieren der Untergrenze in jedem Schritt ist oft eine einfachere Methode als das direkte Maximieren der Log-Likelihood-Funktion. Diese Untergrenze nimmt aufgrund des M-Schritts im Algorithmus immer zu, wodurch die Konvergenz zu einer Lösung sichergestellt wird. Das Problem ist jedoch, dass dieser Algorithmus manchmal zu lokalen Optima konvergiert und keine Garantie dafür gegeben wird, dass er zum globalen Optimum konvergiert. (Dempster *et al.*, 1977).

Aufgrund der Tatsache, dass die vollständige Tabelle $\Pr(i, j, g)$, nicht beachtet wird, muss es berechnet werden. Durch Erstellen einer temporären Übergangsmatrix auf der Grundlage der Wahlergebnisse im ersten Moment wird der E-Schritt des Algorithmus ausgeführt (Little &

Rubin, 1987). Der M-Schritt erfolgt durch Teilen $\sum_g \Pr(i, j, g)$ von $\Pr(i, j)$ (siehe Formel

(2.2)) und Erstellen einer neuen Schätzung der Übergangsmatrix. Die Optimierung der ML erfolgt durch Anpassen der (temporären) vollständigen Tabelle $\Pr(i, j, g)$ abwechselnd zu den Wahlen des ersten und des zweiten Augenblicks. Dies ist der Beginn des M-Schritts, der vor der Schätzung einer neuen Übergangsmatrix erfolgt.

Die Ökobilanz, die den EM-Algorithmus verwendet, löst das Problem unrealistischer Schätzungen und unterscheidet sich von den gewöhnlichen Schätzungen der kleinsten Quadrate, bei denen es sich um eine Methode handelt, die den Fehler minimiert, während diese Methode die Schätzungen maximiert. Latente Klassenanalyse und Iterative Proportional Fitting (IPF) sind beide Methoden, um eine maximale Wahrscheinlichkeit zu erreichen

schätzen. Der zugrunde liegende Algorithmus unterscheidet sich jedoch. Die Ökobilanz basiert auf dem EM-Algorithmus und die IPF basiert auf dem IPF-Algorithmus.

2.4.4 Iterative Proportionalanpassung

Iterative Proportional Fitting (IPF) ist ein mathematisches Verfahren, das ursprünglich entwickelt wurde, um Informationen aus zwei oder mehr Datensätzen zu kombinieren (Bishop *et al.*, 1975). In diesem Fall verwenden wir Informationen aus zwei Augenblicken, die Wahlen von 2003 und 2006. In einem ersten Schritt wird das Wahrscheinlichkeitsmodell beispielsweise in Form der Formel (2.1) ausgedrückt. Dieses Modell enthält eine unbekannte Wahrscheinlichkeitstabelle, genauer gesagt die Übergänge $\Pr(j | i, t)$.

Der IPF-Schätzungsansatz besteht zunächst darin, Werte für diese Übergänge einzugeben und diese Schätzungen dann schrittweise zu verbessern, wobei die Schätzungen kontinuierlich aktualisiert werden. In unserem Fall werden die ersten Wahlergebnisse zufällig ausgewählt, um die anfänglichen Werte zu erhalten. Die Eingabe ist die Schätzung, die im Fall des Benchmark-Modells der vollständigen Unabhängigkeit angegeben wurde:

$$f_{ijg}^{(0)} = n \times \frac{f_{g++} f_{+i+}}{n} \frac{f_{++j}}{n} \quad (2.9)$$

Der nächste Schritt ist die Optimierung der anfänglichen Schätzung:

$$f_{ijg}^{t+1} = \frac{f_{i+} f_{+j}}{f_{g++}^{t+1}} \frac{f_{g++}^{t+1} f_{+i+}}{f_{g++}^{t+1}} \frac{f_{++j}}{f_{++j}^{t+2}}$$

Teilen durch n kann am Ende der Berechnung durchgeführt werden, um die Wahrscheinlichkeiten von zu erhalten f .

Bei dem die f 's sind offensichtlich die beobachteten Randeinträge und die f sind die geschätzte Randeinträge. Das iterative Verfahren wird beendet, sobald die Grenzsummen eng mit den beobachteten Grenzwerten übereinstimmen und kleiner als ein vorgegebenes (kleines) Kriterium sind. Unter bestimmten Bedingungen der Regelmäßigkeit konvergieren die IPF-Updates zu einer ML-Lösung.

2.4.5 Kombinationen

Ein weiterer interessanter Gedanke besteht darin, den Nutzen der Umfrageforschung mit realen Wahlkosten, wirtschaftlichen Variablen und demografischen Variablen zu kombinieren. Es wurden nur wenige Untersuchungen durchgeführt, um diese Informationsquellen wirklich in ein Modell zu integrieren. Meist werden Modellergebnisse und Umfrageforschung verglichen, wobei Umfragedaten als eine Art Benchmark verwendet werden. Es bleibt schwierig, das wahre individuelle Verhalten abzuschätzen, und es sind immer noch Schlussfolgerungen erforderlich.

King schlägt vor, aggregierte Daten (auf der niedrigstmöglichen Ebene) in die amerikanische Version der niederländischen NKO-Forschung einzufügen (King, 1996). Egmond *et al.*, (1998) zeigen jedoch in ihren Untersuchungen zur politischen Partizipation, in denen sie den Zeitraum von 1971 bis 1994 betrachten, dass die meisten Kontextvariablen die Unterschiede bei den Wahlbeteiligungsniveaus nicht wesentlich erklären. Darüber hinaus kommen sie zu dem Schluss, dass die meisten Variationen und Erklärungskräfte aus dem Kontext der Wahlen selbst resultieren (Egmond *et al.*, 1998). Dies kann auch für das Wahlverhalten zutreffen und möglicherweise das Potenzial von Kontextvariablen einschränken.

Eine interessante alternative Sichtweise zur Verwendung von aggregierten Informationen und Umfrageinformationen wurde von Thomsen (2004) implementiert. In diesem Artikel beschreibt er eine Mehrparteien-Situation aus Dänemark unter Verwendung eines bedingten Logit-Modells sowie von Umfrageinformationen. Thomsen beschreibt das allgemeine Abstimmungsverhalten auf nationaler Ebene nur mit einem Gebrauchsmuster (Thomsen, 2004). Diese Arbeit setzt seine Arbeit von 1987 fort, in der er seinen nichtlinearen Schätzer vorstellt. Er behandelt aufeinanderfolgende Wahlen als symmetrisch und gibt an, dass sie das Ergebnis eines gemeinsamen latenten Faktors sind, den er als Parteiidentifikation bezeichnet (Thomsen, 1987). In seiner Arbeit von 2004 fügt er Partei-Sympathie und Abstimmungsergebnisse in die Mischung ein und verleiht seinem Modell mehr Erklärungskraft und Genauigkeit.

In dieser Studie werden zwei Möglichkeiten verwendet, um die Ergebnisse der NKO-Umfrage mit dem quadratischen Programmiermodell zu kombinieren. Das QP-Modell wird verwendet, weil es relativ einfach ist, mehr Einschränkungen in die Zielfunktion aufzunehmen. Diese Einschränkungen werden mit einem sehr einfachen Konfidenzintervall hinzugefügt. Wir setzen die NKO-Werte in das QP-Modell ein und lassen dem Modell eine gewisse Freiheit, die Werte zu schätzen, jedoch nur innerhalb eines Konfidenzintervalls (CI):

$$CI: \bullet - 2 \times S_{x_i} \leq x_i \leq \bullet + 2 \times S_{x_i} \quad (2.10)$$

Die Standardabweichung der Stichprobe (S_{x_i}) ist dann:

$$S_{x_i} = \frac{s}{\sqrt{n}} \quad (2.11)$$

Wo:

$$s = \sqrt{\frac{1}{N - 1} \times \sum_{ich} (x_{ich} - \bar{x})^2}$$

$$n = 459 \text{ Gemeinden}$$

(Oosterhoff & van der Vaart, 2003).

Um interessante Kombinationen zwischen dem NKO und modellbasierten Ansätzen zu erhalten (in diesem Fall wurde quadratische Programmierung verwendet), wurden zwei Kombinationsmodelle erstellt:

Ersetzen Sie Werte auf der Diagonale innerhalb des CI durch NKO-Werte

Ersetzen Sie Nullen innerhalb des CI durch NKO-Werte

Der Grund für das erste Modell war, unrealistische Werte auf der Diagonale zu ersetzen. Zum Beispiel sind Werte von eins nicht sinnvoll und können unter Verwendung der NKO-Werte der Diagonale korrigiert werden. Dies beeinflusst auch die anderen Werte in der Matrix, da alle Zeilen zu 1 addiert werden müssen. Das andere Motiv für die Kombination der beiden Ansätze besteht darin, zu versuchen, mögliche unrealistische Nullen in der Matrix zu ersetzen (zweites Modell). Aufgrund der Tatsache, dass modellbasierte Ansätze nur die Nettoübergänge abschätzen können, präsentierte der NKO einige Startwerte für diese Nullen basierend auf dem NKO. Auch dies hat aufgrund der Einschränkungen auch Einfluss auf die anderen Werte. Letztendlich führen eingeschränkte Modelle immer zu weniger optimalen Lösungen als nicht eingeschränkte Modelle.

2.4.6 Kurze Zusammenfassung

Tabelle 2 zeigt, dass die meisten Analysen auf der Analogie von Keller & ten Cate (1977) basieren. Die neuen Elemente in dieser Studie sind, dass neue Modelle miteinander verglichen werden. Wenn Schätzungen aus dem zweiten Moment verwendet werden, die anhand der Wahlergebnisse aus dem ersten Moment berechnet werden, können diese Werte mit den tatsächlichen Wahlergebnissen verglichen und die Anpassung dieser Schätzungen bestimmt werden. Die Modelle können mit dieser Methodik bewertet werden. Dies ist eine weitere Ergänzung der bestehenden Forschung zu diesem Thema. Außerdem wurde versucht, das Modell mithilfe von Kombinationen aus Umfrage- und Modellergebnissen einzuschränken, um die Gültigkeit der Modelle zu erhöhen.

Tabelle 2. Vergleichstabelle

	Gewichtete Umfrage	Lineare Regression	Quadratische Programmierung	Maximale Wahrscheinlichkeit *
NKO	X.			
Keller & Ten Cate		X.	X.	X.
Grenzergebnisse gleich				X.
Combo 1	X.		X.	
Combo 2	X.		X.	

*  maximale Wahrscheinlichkeit gilt sowohl für Ökobilanzen als auch für

3. Anwendung im niederländischen Hoheitsgebiet

3.1 Einleitung

Die Niederlande haben ein Mehrparteiensystem mit proportionaler Vertretung, was bedeutet, dass der Zugang für kleinere Parteien relativ einfach ist. Dies führt zu einer höheren Volatilität der Anzahl der Parteien im niederländischen System. Die Niederlande sind ein sehr interessanter Fall, da in den letzten Jahrzehnten immer mehr Wähler einer politischen Partei nicht treu geblieben sind, was es schwieriger macht, vorherzusagen, was die Wähler bei den kommenden Wahlen tun werden. Quin *et al.* ((1999) stellen sogar fest, dass die Niederlande ein kritischer Fall sind, da das niederländische Wahlverhalten eingehend untersucht wurde. Die Niederlande haben ein System der proportionalen Vertretung ohne das Problem der Distriktfaktoren. Darüber hinaus haben niederländische Wähler in den letzten Jahren unterschiedliche Unterstützung geleistet (Quinn) *et al.*, 1999, S. 3).

Diese Studie kann einen interessanten Beitrag zur Erklärung des Aufstiegs und Niedergangs politischer Parteien und des Wahlverhaltens der Wähler leisten. Zum Beispiel war der Aufstieg von Pim Fortuyn und der jüngste Sieg des rechten Geert Wilders für viele eine Überraschung.

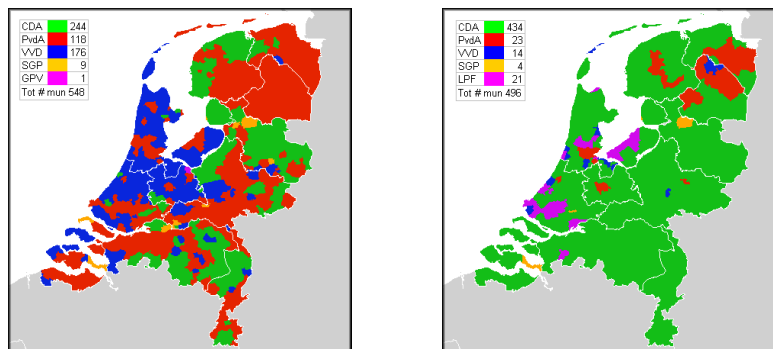


Abbildung 1a. Niederländische Wahlergebnisse 1998 Abbildung 1b. Niederländische Wahlergebnisse 2002

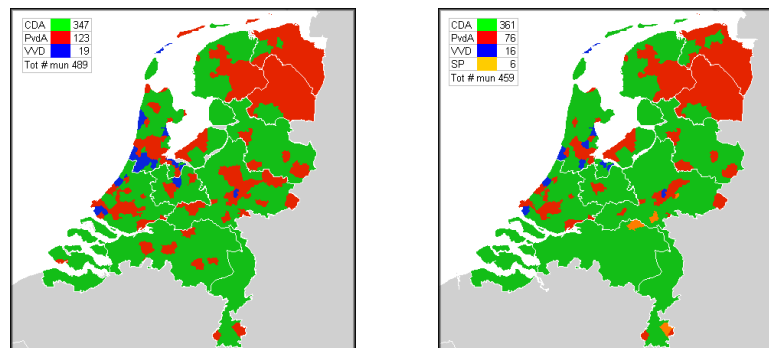


Abbildung 1c. Niederländische Wahlergebnisse 2003 Abbildung 1d. Niederländische Wahlergebnisse 2006

(Kiesraad.nl, 2008)

Die Abbildungen 1a bis 1d zeigen, dass die Wahlergebnisse auf Makroebene während der letzten Wahlen noch relativ stabil sind. Studien in verschiedenen Bereichen haben gezeigt, dass die Menschen ihrer politischen Partei immer noch relativ treu sind (siehe z. B. Shachar & Shamir, 1996). Die Christdemokraten (CDA) blieben in den meisten Gemeinden ununterbrochen die größte Partei. Die Legende zeigt, dass die Zahl der Gemeinden von 548 im Jahr 1998 auf 459 im Jahr 2006 gesunken ist. Obwohl die Übergänge zwischen den Parteien im letzten Jahrzehnt volatiler geworden sind, sind sie immer noch relativ stabil. Mehr als 50% der Wähler entscheiden noch einige Monate vor den Wahlen, was sie wählen sollen, aber diese Zahl war von 71% im Jahr 1977 auf etwas mehr als 50% im Jahr 2003 gesunken (NKO, 1977-2003). Die Niederlage der sogenannten lila Parteien (Regierungsparteien PvdA, D'66 und VVD) im Jahr 2002 ist auf der Karte gut sichtbar. Auch die Niederlage und Erholung der Soziodemokraten (PvdA) bei den letzten beiden Wahlen ist offensichtlich vorhanden. Trotz der Nettoveränderungen, die in den Fällen von Lijst Pim Fortuyn in den Jahren 2002 und 2003 und der Sozialistischen Partei in 2006 groß waren, bleiben viele Wähler in ihrem Wahlverhalten stabil (siehe auch Abbildung 3).

Tabelle 3. Wahlergebnisse 2003 und 2006 in Prozent

Party	Wahlergebnisse (%)				Veränderung		
	1998	2002	2003	2006	Veränderung 1998-2002	Veränderung 2002-2003	Veränderung 2003-2006
Christdemokraten (CDA)	18,37%	27,93%	28,6%	26,51%	+ 9,56%	+ 0,69%	- 2,13%
Labour Party (PvdA)	28,98%	15,11%	27,26%	21,19%	- 13,87%	+ 12,15%	- 6,07%
Sozialistische Partei (SP)	3,53%	5,90%	6,32%	16,58%	+ 2,37%	+ 0,42%	+ 10,26%
Liberalen Partei (VVD)	24,69%	15,44%	17,91%	14,67%	- 9,25%	+ 2,47%	- 3,25%
Gruppe für die Freiheit (PVV)	--	--	--	5,98%	--	--	--
Grüne Linke (GL)	7,27%	6,95%	5,14%	4,60%	- 0,32%	- 1,82%	- 0,54%
Christliche Union (CU)	--	2,54%	2,12%	3,97%	- 0,76% *	- 0,42%	+ 1,85%
Demokraten 66 (D66)	8,99%	5,10%	4,07%	1,96%	- 3,89%	- 1,03%	- 2,11%
Party für Tiere (PvdD)	--	--	0,49%	1,83%	--	--	+ 1,34%
Christlich-reformierte Partei (SGP)	1,78%	1,72%	1,56%	1,56%	- 0,06%	- 0,16%	0,00%
Liste Pim Fortuyn (LPF)	--	17,00%	5,70%	0,21%	--	- 11,30%	- 4,93%
Reformierte politische Union (GPV)	1,26%	--	--	--	--	--	--
Reformpolitisch Föderation (RPF)	2,03%	--	--	--	--	--	--
Lebenswerte Niederlande (LN)	--	1,61%	0,40%	--	--	- 1,21%	--
Andere	3,10%	0,71%	1,31%	1,00%	- 2,39%	+ 2,02%	- 0,31%

* GPV und RPF werden addiert, um diesen Vergleich durchzuführen

(Quelle: CBS.nl:Statline)

Tabelle 3 zeigt deutlich die Verschiebungen zwischen den Parteien. Wie bereits erwähnt, können Sie die größten Verschiebungen bei der Unterstützung der PvdA im Jahr 2002, der LPF im Jahr 2003 und der SP im Jahr 2006 feststellen. Die traditionelle christliche Partei CU weist ein stetiges Wachstum auf, während andere Parteien wie der GL und der D'66 a stetigen Rückgang. Trotz einiger vorübergehender und interessanter Änderungen bleiben die etablierten Parteien im niederländischen Parlament gut vertreten.

3.2 Verfügbare Daten

Die für diese Art der Analyse benötigten Daten sind die Wahlergebnisse auf der niedrigsten verfügbaren Ebene, in diesem Fall auf der Ebene der Gemeinden. Für alle Gemeinden sind die Häufigkeiten aller Parteienentscheidungen bekannt. Ein erschwerender Faktor ist, dass die Gemeinden im Laufe der Zeit neu verteilt wurden. Zwischen 1998 und 2006 ist die Anzahl der Gemeinden von 548 auf 459 gesunken, was eine gewisse Datenverarbeitung erforderlich macht. Die Lösung für diese Komplikation bestand darin, die Ergebnisse aller Bestandteile umverteilter Gemeinden zusammenzufassen, damit die Ergebnisse vergleichbar sind. Die Ergebnisse der Übersee-Abstimmung wurden mit einem separaten Code in „Den Haag Postal Voting“ zusammengefasst und werden als eine Gemeinde behandelt.

Es gibt eine weitere Komplikation beim Vergleichen der Ergebnisse. Einige Parteien existieren nur in einer oder mehreren aufeinander folgenden Wahlperioden. Zum Beispiel war die Liste Pim Fortuyn (LPF) nur in den Jahren 2002 und 2003 beliebt. Anhand der NKO-Ergebnisse kann man leicht erkennen, wohin die Stimmen der LPF gegangen sind: an die Partij voor de Vrijheid (PVV). Die meisten PVV-Stimmen stammen von den ehemaligen Anhängern der LPF, daher hat die PVV die LPF 2006 gewissermaßen ersetzt.

Nicht alle Parteien sind groß genug, um berücksichtigt zu werden. Sie konkurrieren nur in einer begrenzten Anzahl von Wahlkreisen oder erfüllen nicht die Wahlschwelle, die in das Parlament gewählt werden soll. Bei so wenigen Stimmen ist es nicht möglich, die Abstimmungsübergänge dieser Parteien separat zu visualisieren. Wir haben diese Stimmen im „Rest“ zusammengefasst.

Kategorie. Wir betrachten 10 Parteien, dh $m_i = m_j = 10$. Der Grund für diese Auswahl von

Die Parteien stützten sich auf die Vergleichbarkeit und die Anzahl der Sitze im Parlament. In Anhang A Abschnitt 7.1 sind die Parteien der Wahlen von 1998, 2002, 2003 und 2006 zusammengefasst.

Eine weitere Anomalie in den Daten war, dass die Wahlbeteiligung auf den niederländischen Inseln in einigen Wahljahren über 100 lag. Dies kann durch die Tatsache erklärt werden, dass viele Menschen diese Inseln in den Ferien besuchen und ihren sogenannten Abstimmungsausweis verwenden, um in einer anderen Gemeinde als der, in der sie offiziell registriert sind, abzustimmen. Auf diese Weise ist es möglich, dass mehr Menschen ihre Stimme abgeben als es Einwohner in dieser Gemeinde gibt. Interessanterweise sind diese Werte über 100% nur in den Jahren 1998 und 2002 aufgetreten, als im Mai Wahlen abgehalten wurden.

Eine interessante Frage ist, ob Daten auf einer stärker disaggregierten Ebene, wie z. B. die Wahlergebnisse auf Gemeindeebene, bessere Ergebnisse liefern als Wahlergebnisse nur auf nationaler Ebene. Die Daten auf Gemeindeebene haben Vorteile (Park, 2004: S. 14):

Das Aggregationsproblem ist auf dieser unteren Ebene weniger stark vorhanden.

Mehr Beobachtungseinheiten (anstelle nur der Wahlergebnisse des Landes oder der Provinz) führen zu größeren Abweichungen und damit zu einer besseren Genauigkeit der Schätzungen.

Er zeigt, dass die Aggregationsverzerrung tatsächlich abnimmt. Er gibt jedoch an, dass die durch die Nichtlinearität der Ergebnisse verursachte Verzerrung sogar zunehmen kann, insbesondere wenn ein lineares Modell wie das lineare Regressionsmodell angenommen wird (Park, 2004).

In jeder Wahlperiode treten neue Wähler in das Register ein und andere verlassen das Register aufgrund von Tod, Auswanderung usw. Dies wird sorgfältig registriert und hat Einfluss auf die Ergebnisse. Aufgrund des hohen Arbeitsaufwands wurden diese Bedenken in dieser Studie nicht berücksichtigt. Bei jeder Wahl ist eine kleine Anzahl von Stimmen aus irgendeinem Grund nicht gültig. Diese Stimmen wurden nicht berücksichtigt, da sie nicht in die Berechnung der Wahlschwelle einbezogen werden. (Siehe Abschnitt 4.9).

Die elektronische Verfügbarkeit niederländischer Wahlergebnisse auf Gemeindeebene ist noch recht begrenzt. Nur die Ergebnisse der letzten vier Wahlen 1998, 2002, 2003 und 2006 sind elektronisch verfügbar. Wir mussten unsere Pläne aufgeben, die Forschung von Keller & Ten Cate von 1972/1977 mit den neueren Wahldaten zu vergleichen, da die Autoren die Daten leider nicht mehr hatten. Das niederländische Parlament hat jedoch alle Wahlergebnisse auf Papier. Über den Dienst „Bild und Ton“ (*beeld en geluid*), Es ist auch möglich, Wahlergebnisse zu erhalten, jedoch wiederum nicht in einem geeigneten Format. Zum Glück hat der Abstimmungsrat (*kiesraad*) erstellt derzeit ein elektronisches Archiv aller Wahlen auf Gemeindeebene. Dies soll irgendwann im nächsten Jahr abgeschlossen sein, was für diese Studie zu spät ist. Da es zu zeitintensiv war, alle Daten selbst einzugeben, wurde beschlossen, dies nicht weiter zu verfolgen.

3.3 Maßnahmen anpassen

Um alle Modelle korrekt miteinander zu vergleichen, sind einige Methoden erforderlich, die eingeführt werden. Um die Diskrepanzen zwischen Vorhersagen und tatsächlichen Wahlergebnissen zusammenzufassen, verwenden wir das bekannte Pearson-Anpassungsgüte-Chi-Quadrat (GFX)². Die Formel lautet:

$$GFX \triangleq \sum_j \sum_g \frac{(\Pr(j|G) - \Pr(j|G))^2}{\Pr(j|G)} \quad (4.1)$$

Diese Formel berechnet die relativen quadratischen Unterschiede zwischen beobachteten und erwarteten Werten, summiert über Parteien, j , und Gemeinden, g (Long, 1997). Niedrigere Werte zeigen eine bessere Anpassung des Modells an. Dieser Wert wird jedoch nicht als bewertet

solche, aber im Vergleich zu dem Wert, der unter der Annahme erhalten wird, dass nachfolgende Wahlergebnisse in jeder Gemeinde unabhängig sind.

Die Log Likelihood Ratio-Statistik wurde auch als Kontrollstatistik berechnet, um die gefundene Anpassungsgütestatistik zu verifizieren. Die Statistik wird berechnet als:

$$LRX \hat{=} 2 \sum_j \sum_g \text{Log} \left(\frac{\text{Pr}(j/g)}{\text{Pr}(j)} \right) \quad (4.2)$$

Diese Statistik ist analog zur Anpassungsgütestatistik. Sein Verhalten unterscheidet sich nur asymptotisch vom GFX_2 . Wenn die Werte des GFX_2 und der LRX_2

unterscheiden sich stark, man sollte sehr vorsichtig mit der Gültigkeit der Ergebnisse sein. (Bischof, Fienberg und Holland, 1975)

Um den Unterschied zwischen dem Benchmark-Modell der völligen Unabhängigkeit und den anderen Modellen besser visualisieren zu können, wird auch der Unterschied zwischen zwei Chi-Quadraten berechnet. Dies gibt einen Hinweis auf die Verbesserung der Passform der anderen Modelle und wird als DLRX bezeichnet.

Der Unterschied in Freiheitsgraden (df) im Vergleich zum Benchmark-Modell wird als Ddf dargestellt. Für das Unabhängigkeitsmodell müssen $9 + 9 = 18$ Parameter geschätzt werden, für die NKO-Matrix beträgt die Anzahl der Parameter $10 \times 9 = 90$. Daher beträgt Ddf zwischen den beiden 72. Mit derselben Analogie können auch die übrigen Parameter der anderen Modelle berechnet werden.

McFaddens Pseudo R_2 wird verwendet, um die Passform aller Modelle anzuzeigen. Sie wird berechnet, indem das Chi-Quadrat des Modells durch das Chi-Quadrat des Unabhängigkeitsmodells geteilt wird.

Die Formel für das McFadden-Pseudo- R_2 ist (McFadden, 1973, S. 121):

$$Ps-R_2 \hat{=} \frac{((I_{\max} - I_0))}{((I_m - I_0))} = 1 - \frac{I_m}{I_0} \quad (4.3)$$

Wo I_{\max} ist die perfekte Passform, I_m ist das Chi-Quadrat des Vollmodells, zum Beispiel LCA

und I_0 ist das Chi-Quadrat des Benchmark-Modells der Unabhängigkeit.

Größere Werte des R_2 zeigen eine bessere Anpassung des beteiligten Modells an. Wenn $R_2 = 1$ dann ergibt sich auf der Grundlage der verwendeten Übergangsmatrix die Wahl aus der zweiten

Der Moment auf Gemeindeebene kann anhand der Wahlergebnisse des ersten Augenblicks perfekt vorhergesagt werden³.

³ Veall & Zimmermann (1996) haben eine Umfrage zu den verschiedenen Formen von Pseudo-R₂ verfasst² bei der Darstellung der Modellanpassung. Ihre Hauptschlussfolgerung ist, dass es kein offensichtliches Kriterium gibt, um zu entscheiden, welches spezifische Pseudo-R₂ ist das Beste. Das McFadden Pseudo-R₂ ist das am häufigsten verwendete Pseudo-R₂. Cameron & Windmeijer (1993) haben eine Verallgemeinerung vorgenommen, um eine größere Vielfalt von Situationen abzudecken. Laut Vaell & Zimmermann ist das Pseudo-R₂ ist konzeptionell nah an der R₂ Wert, der aus OLS berechnet werden kann. Dies ist eine Empfehlung, da das OLS-Pseudo-R₂ traditionell als Maß für die Anpassungsgüte für lineare Modelle wie die Regression verwendet wird. Weil es keinen Konsens darüber gibt, welches Pseudo-R₂ ist in einer bestimmten Situation am besten, muss man vorsichtig sein, wenn man Ergebnisse verwendet, vergleicht und interpretiert. Wir stellen fest, dass ein anderes Pseudo-R₂

kann andere Ergebnisse liefern und führt zu anderen Interpretationen der Werte. Zum Beispiel das McFadden Pseudo-R₂ kann bei gleicher Berechnung einen Wert von 0,25 erzeugen, während der McKelvey-Zavoina R₂ erzeugt einen Wert von 0,5.

4. Ergebnisse

In diesem Kapitel werden die Ergebnisse der Analysen vorgestellt. Zunächst gibt es einen Abschnitt über die Ergebnisse der Anpassungsgüte. In den nächsten Abschnitten werden alle Modelle separat behandelt. Der folgende Abschnitt enthält die Diskussion über das 100% Stayer-Modell, das NKO-Modell und das LCA-Modell für die Wahlen 2003-2006. Für alle Wahlen werden unterschiedliche Übergangsmatrizen geschätzt, und für die Wahlen 2003-2006 werden das NKO-Modell und das LCA-Modell vorgestellt und überprüft. Das Kapitel endet mit einem Abschnitt über Sitzverteilungen und die diesbezüglichen Modellleistungen. Die anderen Übergangsmatrizen und Analysen von 1998-2002 und 2002-2003 sind in Anhang A, Abschnitt 7.2 und 7.3 dargestellt.

4.1 Statistik der Anpassungsgüte

Tabelle 4. Statistik der Anpassungsgüte für alle Modelle 1998-2002

Modell	GFX ₂	LRX ₂	DLRX ₂	Ddf	P.	McFadden Pseudo-R ₂
Unabhängigkeit	18427358	8057151		--		--
100% Steher	288695202	12499000	- 4441849	- 9	<.0001	- 14.6667
NKO	478995.8	465193.5	7591957.5	81	<.0001	0,974006
LPM	300381.6	284876.9	7772274.1	81	<.0001	0,988832
Ökobilanz	205794.6	201438.4	7855712.6	81	<.0001	0..983699
IPF	1771672	1652404	6404747	81	<.0001	0,903856
Combo	321557.4	318446.1	7738704.9	81	<.0001	0,98255
Combo 2	321557.4	318446.1	7738704.9	81	<.0001	0,98255

Quelle: (Eigene Berechnungen des Autors)

Tabelle 5. Statistik der Anpassungsgüte für alle Modelle 2002-2003

Modell	GFX ₂	LRX ₂	DLRX ₂	Ddf	P.	McFadden Pseudo-R ₂
Unabhängigkeit	3776480	3580063		--		--
100% Steher	2418201	2012710	1567353	- 9	<.0001	0,3596679
NKO	393554.2	405704.9	3174358.1	81	<.0001	0,895788
Ökobilanz	157278.7	154362.7	3425700.3	81	<.0001	0,958353
LPM	98510.9	98410.4	3481652.6	81	<.0001	0,973915
IPF	1999259	1840727	1739336	81	<.0001	0,4706024
Combo	223587.7	221840.4	3358222.6	81	<.0001	0,9407947
Combo 2	223587.7	221840.4	3358222.6	81	<.0001	0,9407947

Quelle: (Eigene Berechnungen des Autors)

Tabelle 6. Statistik der Anpassungsgüte für alle Modelle 2003-2006

Modell	GFX ₂	LRX ₂	DLRX ₂	Ddf	P.	McFadden Pseudo-R ₂
Unabhängigkeit	4548314	3290676		--		
100% Steher	2643817	1956575	1334101	- 9	<.0001	0,4187
NKO	362460.2	342284.8	2948391.2	81	<.0001	0,9203
Ökobilanz	146497.4	139716.3	3150959.7	81	<.0001	0,9678
LPM	178298.9	172427.8	3118248.2	81	<.0001	0,9608
IPF	1856685	1715659	1575017	81	<.0001	0,5918
Combo	238607.4	223775.1	3066900.9	81	<.0001	0,9475
Combo 2	181484.6	175740.6	3114935.4	81	<.0001	0,9601

Quelle: (Eigene Berechnungen des Autors)

In den Tabellen 4, 5 und 6 sind die Statistiken zur Anpassungsgüte für die Wahlen 1998- aufgeführt.

2002, 2002-2003 und 2003-2006. Die Tabellen zeigen die Ergebnisse der Benchmark-Modelle für Unabhängigkeit und 100% Steher. Weiterhin werden die Ergebnisse des NKO-Modells gezeigt. Die modellbasierten Ansätze von LCA, LPM = QP und IPF werden ebenfalls vorgestellt. Beide Kombinationsmodelle werden als Combo (Ersetzen von Werten in der Diagonale) und Combo 2 (Ersetzen von Nullwerten) dargestellt.

Alle Wahlen können mit allen Modellen recht gut geschätzt werden. Die Wahlen von 1998-2002 haben die höchsten Werte in Bezug auf die Modellanpassung. Nur das 100% Steher-Modell schneidet relativ schlecht ab. Die Kombinationsmodelle 1998-2002 und 2002-2003 übertreffen sich nicht und führen zu identischen Ergebnissen. Dies kann durch die Tatsache erklärt werden, dass die Werte der NKO-Matrix in den Jahren 1998-2002 und 2002-2003 näher bei Null liegen als in den Jahren 2003-2006, und für diese Zeiträume enthält das Konfidenzintervall den Wert Null.

Der große Unterschied in den Ergebnissen zwischen Ökobilanz und IPF kann durch die Tatsache erklärt werden, dass der IPF-Algorithmus die Berechnung wesentlich stärker einschränkt. Der modellbasierte Ansatz der Ökobilanz übertrifft den QP-Ansatz nur geringfügig. Dies liegt daran, dass die Ökobilanz Protokollschatzungen verwendet, um die Wahrscheinlichkeit zu maximieren, während QP quadratische Schätzungen verwendet, um den Fehler zu minimieren. Log-Schätzungen sind besser in der Lage, Werte nahe 0 zu verarbeiten, was zu etwas besseren Ergebnissen führt. Die modellbasierten Ansätze zeigen signifikante Verbesserungen des NKO (mit Ausnahme von IPF) und natürlich auch der Unabhängigkeits- und 100%-Stayer-Ansätze.

Das Modell mit der besten Leistung ist das der Ökobilanz mit einem Pseudo- R^2 von .98, .96 und .97. Aufgrund der Anpassungsgütestatistik kann dieses Modell die meisten Übergänge abschätzen und übertrifft die NKO-Ergebnisse. Dies ist zu erwarten, da die Ökobilanz im Gegensatz zu einer Umfrage darauf ausgelegt ist, die Ergebnisse zu optimieren.

4.2 Benchmark-Methoden

Das Unabhängigkeitsmodell und das 100% Steher-Modell geben einen Hinweis auf die Leistung der anderen Modelle. Beide erzeugen eine Übergangsmatrix, die nach der Grundformel (2.1) berechnet wird. Es ist jedoch nicht sehr nützlich, diese Matrizen hier darzustellen. Das Unabhängigkeitsmodell erzeugt eine Matrix mit verschiedenen Spalten, aber identischen Zeilen, die der vollständigen Unabhängigkeit zwischen den Wahlen entspricht. Dies erfolgt auf der Grundlage einer anfänglichen Grenzwahrscheinlichkeit für die erste Wahl. Der 100% Steher ist eine sehr einfache Matrix, die die Identitätsmatrix mit nur einer auf der Diagonale ist.

Da es keine goldene Regel gibt, wie der Erfolg gemessen und die Ergebnisse interpretiert werden können, werden diese Modelle als Benchmark verwendet. Die Werte des Pseudo- R^2 sind auf der niederländischen Karte eingezeichnet⁴. Auf diese Weise werden Abweichungen und andere interessante Fakten sichtbar. Da das Unabhängigkeitsmodell zur Berechnung des Pseudo-R verwendet wird², Dieses Modell kann auf diese Weise nicht gezeichnet werden.

⁴ Denken Sie daran, dass in Formel 3 für 1998-2002 angenommen wird, dass SGP = LPF und für 2003-2006 LPF = PVV. Dies geschieht nur mit diesem Modell.



Abbildung 2a. 100% Steher 98-02

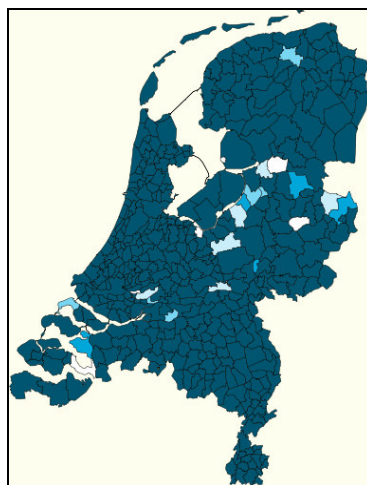


Abbildung 2b. 100% Steher 02-03

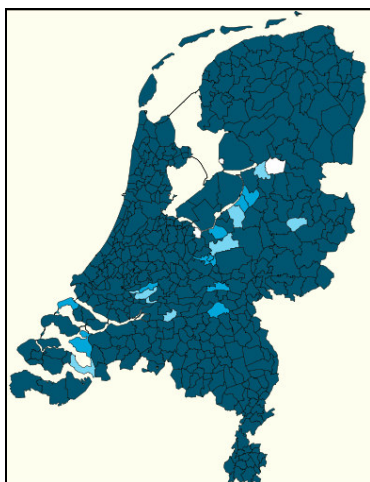


Abbildung 2c 100% Steher Modell 03-06

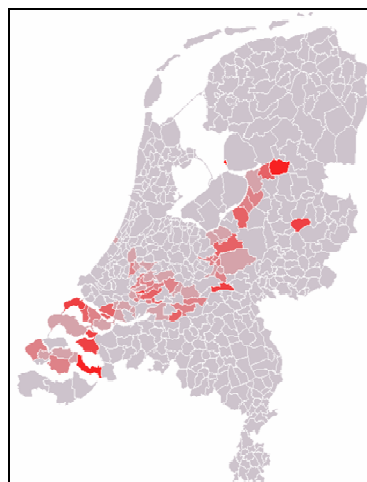


Abbildung 2d. Niederländischer Bibelgürtel

(Kiesraad.nl, 2008)

Legende 1998-2002:	$R^2 < 0.9598$	$R^2 < 0.9702$	$R^2 < 0.9778$	$R^2 < 0.9839$	$R^2 < 1$
Legende 2002-2003:	$R^2 < 0.8376$	$R^2 < 0.8632$	$R^2 < 0.8873$	$R^2 < 0.9116$	$R^2 < 1$
Legende 2003-2006:	$R^2 < 0.8668$	$R^2 < 0.9245$	$R^2 < 0.9464$	$R^2 < 0.9622$	$R^2 < 1$

Die Legenden zeigen die Pseudo-R-Statistiken pro Gemeinde. Je heller die Farbe, desto höher das entsprechende Pseudo-R und die Anpassung. Sie basieren auf dem 20., 40., 60. und 80. Perzentil der NKO-Pseudo-R-Werte. Diese Werte werden verwendet, um die Leistung aller Modelle in Bezug auf die Farben anzuzeigen. Die erste Abbildung zeigt die Übereinstimmung des 100% -Stayer-Modells bei den Wahlen 1998-2002, 2002-2003 und 2003-2006. Wie in Abbildung 2a zu sehen ist, ist die Passform des 100% -Stayer-Modells bei einer der volatilsten Wahlen in der niederländischen Geschichte außergewöhnlich schlecht.

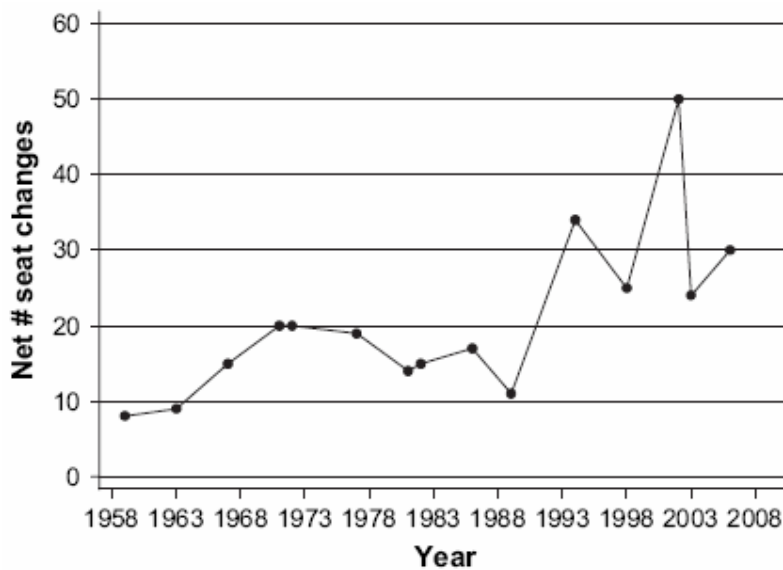


Abbildung 3. Volatilität der Sitze bei allen niederländischen Parlamentswahlen von 1959 bis 2006 (Aarts & van der Kolk, 2007, S. 837)

Aus Abbildung 3 geht hervor, dass die Nettozahl der Sitzveränderungen bei den Wahlen 2002 am höchsten war. Offensichtlich ist das 100% Steher-Modell dort nicht anwendbar. Bei den Wahlen 2002-2003 und 2003-2006 kann das 100% -Stayer-Modell das Wahlverhalten anhand der Wahlergebnisse von 2002 bzw. 2003 besser vorhersagen. 2003 finden wir sehr gute Vorhersagen bei Staphorst, Urk und Bunschoten (McF. Pseudo R. \geq .95). Der sogenannte Bibelgürtel ist sehr gut sichtbar und zeigt, dass die Bürger dieser Gemeinden auch sehr loyale Wähler sind (McF. Pseudo R. \geq .81). Der niederländische Bibelgürtel ist grob auf den Zeeuwse-Inseln, im gesamten Flussgebiet der Provinzen Südholland, Utrecht, Gelderland und teilweise in Nordbrabant (Werkendam und Wijk en Aalburg) sowie im nördlichsten Teil von Overijssel zu finden. Besonders Orte wie Staphorst, Genemuiden, Nieuw Lekkerland, Elspeet, Opheusden, Kesteren, Barneveld, Ederveen, Oudorp, Tholen, Arnhem, Meliskerke, Aagtekerke, Yserke und Krabbendijke befinden sich in der Mitte dieses Gebiets. Einige andere Orte liegen außerhalb dieses Gebiets, aber in diesen Gemeinden gibt es auch eine hohe Konzentration reformierter Menschen. Diese Gemeinden sind Urk, Rijssen, Scheveningen und Katwijk (Wikipedia: Bijbelgordel, 2008).

4,3 NKO

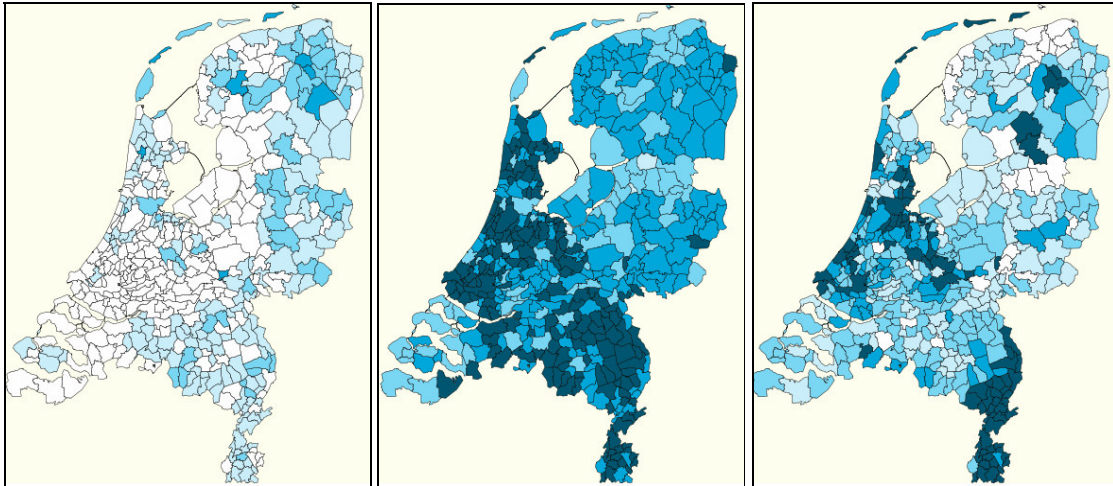


Abbildung 4a. NKO Modell 98-02

Abbildung 4b. NKO Modell 02-03

Abbildung 4c. NKO Modell 03-06

Legende: $R^2 < 0.8650$ $R^2 < 0.9163$ $R^2 < 0.9522$ $R^2 < 0.9728$ $R^2 < 1$

Die Legende, mit der die Modellleistung während der drei aufeinander folgenden Wahlen verglichen wurde, basiert ebenfalls auf den 20th, 40th, 60th und 80th Perzentil der NKO-Pseudo-R-Ergebnisse. Diesmal wurden sie über die drei Wahlen gemittelt, um eine vergleichbare Situation zu schaffen. Die Karte der Niederlande ist viel leichter, wenn die NKO-Übergangsmatrix verwendet wird, verglichen mit dem 100% -Stayer-Modell. Besonders in Zeeland, Nordbrabant und Overijssel sind die Ergebnisse viel besser. In Randstad, großen Teilen der nördlichen und südlichen Niederlande, sind die Ergebnisse jedoch immer noch schlecht, was es interessant macht, die modellbasierten Ansätze zu betrachten, um festzustellen, ob sie eine bessere Leistung erbringen können.

4.4 Quadratische Programmierung

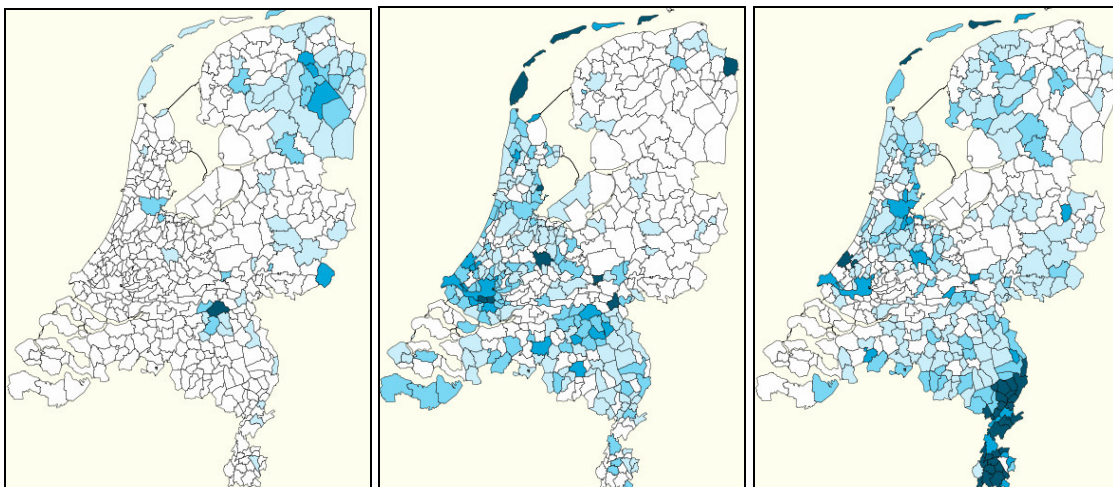


Abbildung 5a. LPM Modell 98-02

Abbildung 5b. LPM-Modell 02-03

Abbildung 5c. LPM-Modell 03-06

Legende: $R^2 < 0.8650$ $R^2 < 0.9163$ $R^2 < 0.9522$ $R^2 < 0.9728$ $R^2 < 1$

Das LPM-Modell zeigt bei allen Wahlen viel hellere Bereiche mit derselben Legende wie das NKO-Modell. Dennoch bleibt die Randstad ein allgemeines Problem. Es gibt keine wirklichen Beispiele, in denen dieses Modell merklich schlechter abschneidet als andere Modelle oder bestimmte Bereiche, in denen dieses Modell nicht in der Lage ist, korrekte Werte zu schätzen.

4.5 Analyse latenter Klassen

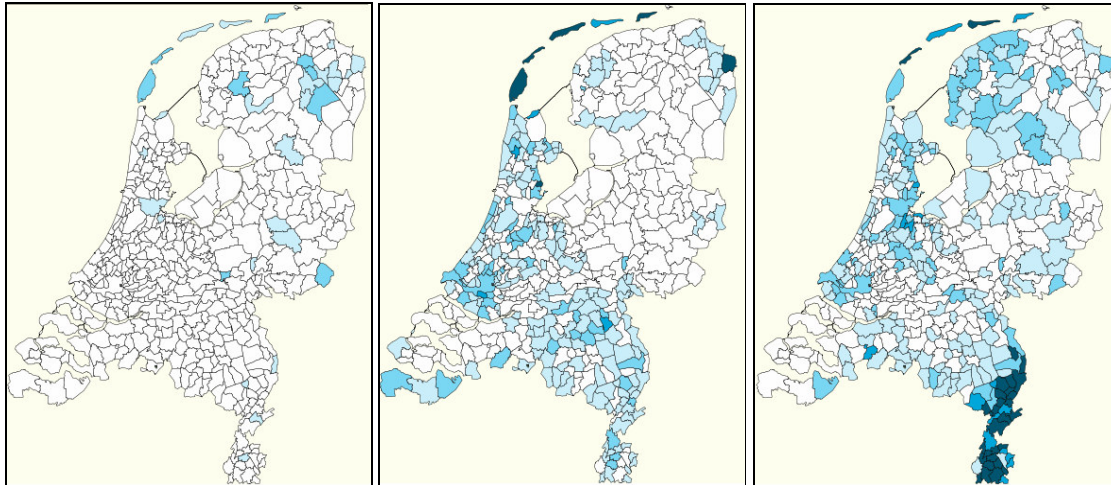


Abbildung 6a. Ökobilanzmodell 98-02

Abbildung 6b. Ökobilanzmodell 02-03

Abbildung 6c. Ökobilanzmodell 03-06

Legende: $R^2 < 0.8650$ $R^2 < 0.9163$ $R^2 < 0.9522$ $R^2 < 0.9728$ $R^2 < 1$

Mit dem LCA-Modell hat die Karte noch hellere Bereiche als mit dem LPM-Modell. Auch dieses Modell kommt mit den Inseln nicht sehr gut zurecht, wie dies auch bei allen anderen Modellen beobachtet wurde.

4.6 Iterative Proportionalanpassung

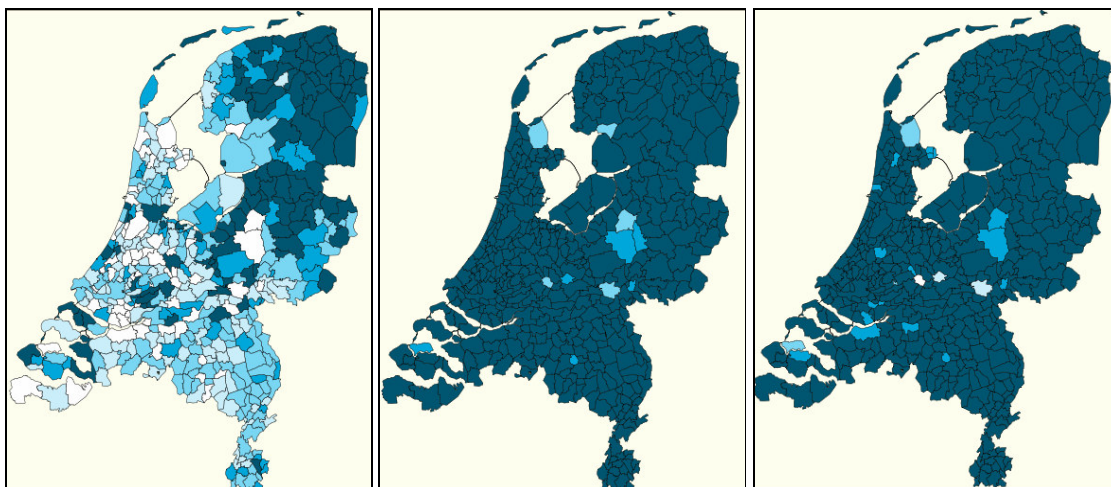


Abbildung 7a. IPF-Modell 98-02

Abbildung 7b. IPF-Modell 02-03

Abbildung 7c. IPF-Modell 03-06

Legende: $R^2 < 0.8650$ $R^2 < 0.9163$ $R^2 < 0.9522$ $R^2 < 0.9728$ $R^2 < 1$

Es ist relativ schwierig, interessante Fakten auf diesen Karten für das IPF-Modell zu sehen, da es allgemein mittelmäßig passt. Es ist ziemlich offensichtlich, dass einige Gemeinden bei allen Wahlen farbiges Licht haben. Diese Gemeinden sind Overbetuwe, Lemsterland, Vianen, Wieringermeer, Epe, Apeldoorn und Northern-Beveland. Die Leistung von IPF ist relativ schlecht, da das unter IPF (Formel (2.1)) geschätzte Modell viel eingeschränkter ist als die anderen bisher diskutierten Modelle.

4.7 Kombinationen

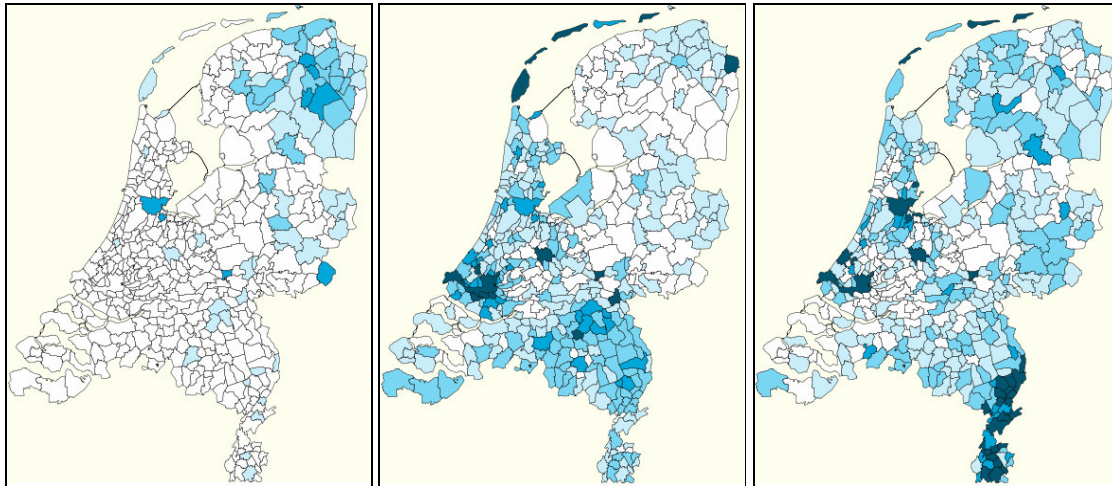


Abbildung 8a. Kombimodell 98-02

Abbildung 8b. Kombimodell 02-03

Abbildung 8c. Kombimodell 03-06

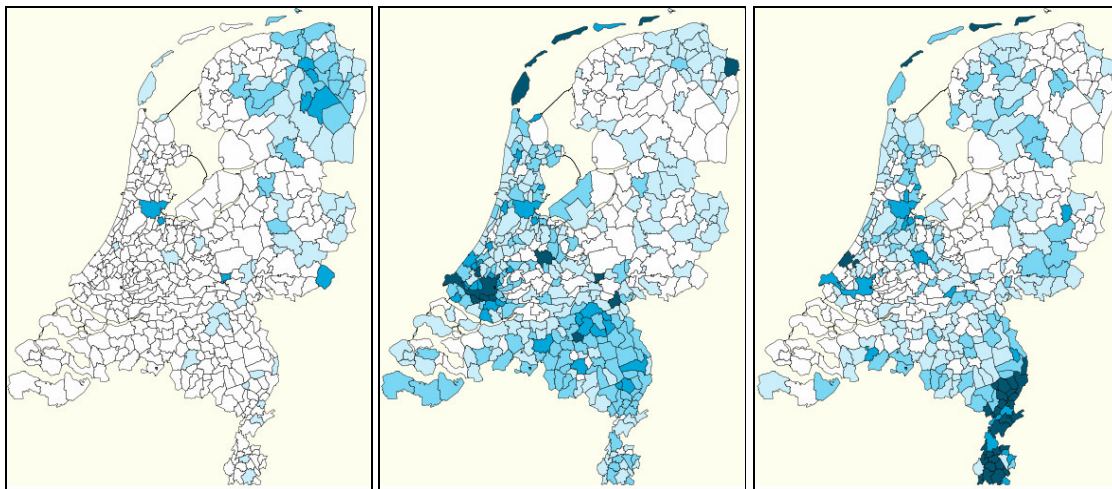


Abbildung 9a. Combo2 Modell 98-02

Abbildung 9b. Combo2 Modell 02-03

Abbildung 9c. Combo2 Modell 03-06

Legende: $R^2 < 0.8650$ $R^2 < 0.9163$ $R^2 < 0.9522$ $R^2 < 0.9728$ $R^2 < 1$

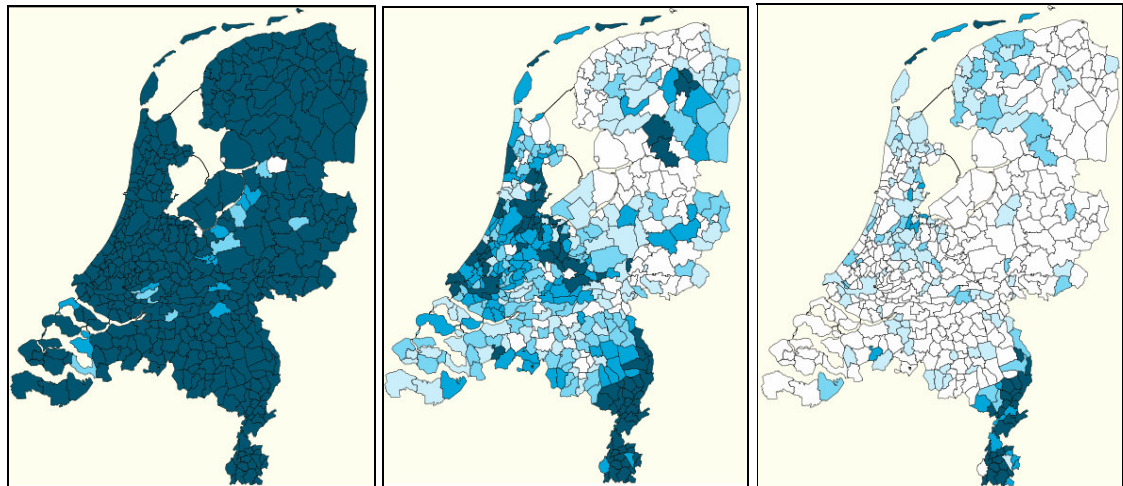
Die Bilder des Kombinationsmodells unterscheiden sich erst 2006. Die Muster in diesen Abbildungen zeigen die allgemeine Linie des LPM-Modells mit etwas schlechteren Passungen.

4.8 Diskussion über die Wahlen 2003-2006

Für alle Wahlen wird eine weitere Analyse durchgeführt, bei der alle Modelle für eine Wahl miteinander verglichen werden. In diesem Abschnitt gehen wir auf die Wahlen von 2003 bis 2006 ein und nur auf das 100% Steher-Benchmark-Modell, das NKO-Modell und den leistungsstärksten modellbasierten Ansatz der Ökobilanz. Die

anderen Wahlen und die Beschreibung aller Modelle sind in Anhang A Abschnitt 7.3 beschrieben.

Der Grund, warum die einfache lineare Regressionsanalyse hier nicht vorgestellt wurde, liegt darin, dass dieses Modell Ergebnisse außerhalb des Bereichs zwischen 0 und 1 liefert. Interessanterweise beziehen sich die meisten Werte, die über 1 lagen, auf die Inseln. Besonders hervorzuheben ist Schiermonnikoog, der zweimal Werte über 1 hatte. Tubbergen ist eine weitere Gemeinde, die zweimal auftaucht. Alle modellbasierten Ansätze haben Probleme, die Abstimmungsübergänge für die Inseln korrekt vorherzusagen. Die Erklärung für diese Beobachtung liegt in der Tatsache, dass die Inseln beliebte Urlaubsorte sind. Die Bevölkerung der Inseln ändert sich erheblich in Abhängigkeit von den Touristen, die dort Urlaub machen. Daher haben alle Modelle Schwierigkeiten, mit dieser Komplikation umzugehen.



Eine ähnliche Argumentation gilt für die Postwähler. Auch Limburg wird in allen Modellen nicht richtig eingeschätzt. Weitere Analysen zeigen, dass die Unterstützung für das PVV in Limburg systematisch unterschätzt wird. Wenn man bedenkt, dass der Gründer des PVV aus diesem Teil der Niederlande stammt, kann man leicht verstehen, warum Limburg vom nationalen Muster abweicht.

Übergangsmatrizen

Alle Modelle haben eine Übergangsmatrix, die für jedes Modell unterschiedlich ist. Wir gehen näher auf die Matrizen der Wahlen von 2003 bis 2006 ein. Die anderen Übergangsmatrizen von

1998-2002, 2002-2003 und 2003-2006, einschließlich Erläuterungen, sind in Anhang A Abschnitt 7.2 enthalten. Aus den Benchmark-Modellen wurden keine Übergangsmatrizen vorgestellt (siehe Abschnitt 4.2).

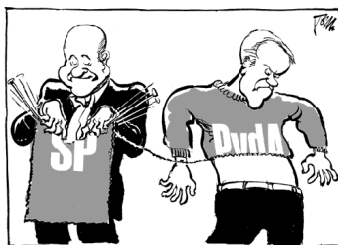
Tabelle 7. NKO-Übergangsmatrix 2003-2006 (diagonale Elemente fett gedruckt) (N = 2528) *

2003	2006									
	CDA	PVDA	VVD	SP	GL	D66	CU	PVV	Anders	NG
CDA	.7072	.0339	.0578	.0637	.0040	.0040	.0398	.0199	.0100	.0598
PVDA	.0294	.5910	.0157	.2035	.0274	.0078	.0078	.0117	.0137	.0920
VVD	.2303	.0284	.5521	.0347	.0032	.0095	.0063	.0473	.0189	.0694
SP	.0379	.1061		.6970	.0606		.0152	.0379	.0227	.0227
GL	.0316	.0737	.0105	.2526	.4632	.0105	.0421		.0211	.0947
D66	.0814	.1744	.1744	.1512	.1163	.2326	.0233		.0465	
CU	.0222	.0222		.0222			.9111	.0222		
LPF	.0706	.0353	.1765	.1412		.0118		.3647	.0471	.1529
Anders	.0952	.0238	.0238	.0952	.0238		.0714	.0238	.5714	.0714
NG	.0600	.0622	.0267	.0867	.0044	.0044	.0022	.0467	.0111	.6956

Quelle: (Aarts *et al.*, 2007: S. 224)

* Es gibt einige Abweichungen von den in Aarts angegebenen Werten *et al.*, (2007) aufgrund von Rundungsfehlern im Buch.

Diese gewichtete Matrix (Korrektur verschiedener Verzerrungen, siehe Abschnitt 2.3) zeigt, wohin die Wähler von 2003 im Jahr 2006 gegangen sind. Beispielsweise haben 71% der Wähler, die 2003 für die CDA gestimmt haben, dies auch im Jahr 2006 getan, während 3% dorthin gezogen sind die PvdA und 6% zum VVD usw. Auch 70% der SP-Wähler von 2003 blieben 2006 bei der SP, aber 11% stimmten stattdessen für die PvdA. Andererseits stimmten 22% der PvdA-Wähler von 2003 2006 für die SP, was mehr ist, als sie gewonnen haben, wie in der Karikatur dargestellt.



Aus den relativ hohen Werten in der Diagonale geht hervor, dass die meisten Wähler ihrer Partei treu waren. Sie sind sogenannte Steher. D'66 hat viel Unterstützung verloren und es gibt relativ wenig Steher in dieser Partei im Vergleich zu anderen Parteien. Es wird auch deutlich, dass viele ehemalige LPF-Wähler 2006 für den PVV gestimmt haben.

(Janssen, 2006)

Tabelle 8. Übergangsmatrix LCAmodel 2003-2006 (diagonale Elemente fett gedruckt)

2003	2006									
	CDA	PVDA	VVD	SP	GL	D66	CU	PVV	Anders	NG
CDA	.8304		.0028	.0664			.0093	.0400		.0511
PVDA		.7419		.2028			.0154			.0399
VVD	.1679		.7889				.0225			.0206
SP				1.0000						
GL					.8394	.1606				
D66		.1700	.1451		.0748	.2799	.0075		.1979	.1248
CU							1.0000			
LPF		.0047				.0017		.5587	.2224	.2125
Anders							.1453		.8547	
NG		.0158		.1145				.0645		.8051

Quelle: (Van der Ploeg *et al.*, 2008)

Die Überprüfung der Ökobilanz-Übergangsmatrix zeigt, dass die meisten Wähler Steher sind. Insbesondere SP, CDA, PvdA, CU, VDD und GL weisen hohe Steherraten auf. Jemand, der 2003 oft nicht gewählt hat, war 2006 auch kein Wähler (81%). Dies ist 6% höher als mit dem LPM-Modell geschätzt. Der Anteil der Nichtwähler ist selbst bei modellbasierten Techniken eine umstrittene Information. Sie können immer noch sehen, dass 11% der Nichtwähler im Jahr 2003 für die SP im Jahr 2006 gestimmt haben, was teilweise das Wachstum der SP erklärt. Die Wählerschaft von D'66 im Jahr 2003 war auf mindestens vier Parteien verteilt. Nach dem LCA-Modell stimmten 2006 über 55% der früheren LPF-Wähler für den PVV, und ein großer Teil der ehemaligen LPF-Wähler blieb zu Hause. SP und CU haben Werte von 1 in ihrer Diagonale, was darauf hinweist, dass sie eine sehr treue Unterstützung haben.

4.9 Sitzverteilung 2003-2006

Eine interessante Visualisierung der Leistung der Modelle kann durch Berechnung der Sitzverteilung in der niederländischen zweiten Kammer erhalten werden. Es sind 150 Plätze zu verteilen. Die Berechnung der Sitzverteilung erfolgt wie folgt:

Erstens, nachdem jede abgegebene Stimme gezählt wurde, die Quote, um einen Sitz zu bekommen (Q) ist berechnet. Alle Stimmen mit Ausnahme der ungültigen Stimmzettel werden nach folgender Formel berücksichtigt:

$$Q = \frac{V}{150}$$

Wo V ist die Gesamtzahl der gültigen Stimmen.

Es ist für Parteien möglich, sich zusammen mit einer anderen Partei zu registrieren und eine sogenannte Listenkombination zu bilden. Dies hat den Vorteil, dass die Anzahl der Stimmen pro Kombination gezählt wird und pro Liste berechnet wird, wie oft das Kontingent erreicht wird, um die Gesamtzahl der von diesen Parteien verdienten Sitze zu erhalten. Bei der Verteilung der verbleibenden Sitze wird die Listenkombination auch als eine Partei angesehen. Eine Listenkombination ist nur dann rentabel, wenn beide Parteien die Wahlschwelle erreichen.

In den Niederlanden entspricht die Wahlschwelle der Quote. Das heißt, je höher die Wahlbeteiligung, desto höher die Quote und damit die Wahlschwelle. Zunächst wird berechnet, wie oft eine Partei die Quote erfüllt, was zur Anzahl der pro Partei gesammelten ganzen Sitze führt. Die Zuordnung von

diese ganzen Sitze (S) an die einzelnen Parteien (ich_{ich}) wird wie folgt durchgeführt:

$$S = \sum_{i=1}^N \frac{ich_i}{Q}$$

Wo ich_i ist die Gesamtzahl der Stimmen für eine einzelne Partei oder Listenkombination i und N ist die Gesamtzahl der Parteien, die die Quote erfüllt haben und S ist eine ganze Anzahl von Sitzen.

Nach dieser Berechnung verbleibt eine begrenzte Anzahl von Sitzplätzen, die als verbleibende Sitzplätze bezeichnet werden. Sie werden durch Subtrahieren der Anzahl ganzer Sitze von insgesamt 150 Sitzen ermittelt. Die Anzahl der verbleibenden Plätze (R) variiert von 5 bis 12 Sitzplätzen. Die Zuweisung dieser verbleibenden Sitze erfolgt nach der Methode des größten Durchschnitts. Durchschnittswerte für verbleibende Sitzplätze (E/N) werden nach der Methode des größten Durchschnitts berechnet. Die Durchschnittswerte für die verbleibenden Plätze (E/N) werden dann berechnet, indem die Gesamtzahl der Stimmen pro Partei oder Listenkombination durch die Anzahl der ganzen Sitze + 1, + 2 und + 3 (verbleibende Sitze) dividiert wird:

$$A = \sum_{i=1}^N \sum_{j=1}^3 \frac{ich_i}{S_i + j}$$

Wo j ist verbleibende Anzahl von Sitzen und S_i ist die Gesamtzahl der Sitze pro Partei.

Die gesamte Liste der Durchschnittswerte wird dann eingestuft und ist die höchste R . Durchschnittswerte erhalten einen zusätzlichen Sitzplatz. Es ist daher möglich, dass eine Partei mehrere verbleibende Sitze erhält. Der letzte Schritt besteht darin, jeder einzelnen Partei innerhalb einer Listenkombination Sitzplätze zuzuweisen. Zu diesem Zweck wird ein Kontingent nur für diese Listenkombination berechnet. Die Gesamtzahl der Stimmen der Listenkombination wird durch die Gesamtzahl der zugewiesenen Sitze für diese Liste geteilt

Kombination. Auf der Grundlage der größten Reste werden die verbleibenden Sitze innerhalb der Listenkombination aufgeteilt.

Tabelle 9 zeigt die Übersetzung der Abstimmungsergebnisse in die Sitzverteilung der letzten vier Wahlen. Diese realen Ergebnisse von 2003 und 2006 werden mit den Schätzungen der Modelle verglichen, um eine weitere visuelle Anpassung zu zeigen.

Tabelle 9. Sitzverteilung 1998-aktuell

Party	1998	2002	2003	2006
Christdemokraten (CDA)	29	43	44	41
Labour Party (PvdA)	45	23	42	33
Sozialistische Partei (SP)	5	9	9	25
Liberale Partei (VVD)	38	24	28	22
Gruppe für die Freiheit (PVV)	--	--	--	9
Grüne Linke (GL)	11	10	8	7
Christliche Union (CU)		4	3	6
Demokraten 66 (D66)	14	7	6	3
Party für Tiere (PvdD)	--	--	--	2
Christlich-Reformierte Partei (SGP)	3	2	2	2
Liste Pim Fortuyn (LPF)	--	26	8	--
Reformierte Politische Union (GPV)	2	--	--	--
Reformational Political Federation (RPF) 3		--	--	--
Lebenswerte Niederlande (LN)	--	2	--	--
Gesamt	150	150	150	150

(Quelle: CBS: statline)

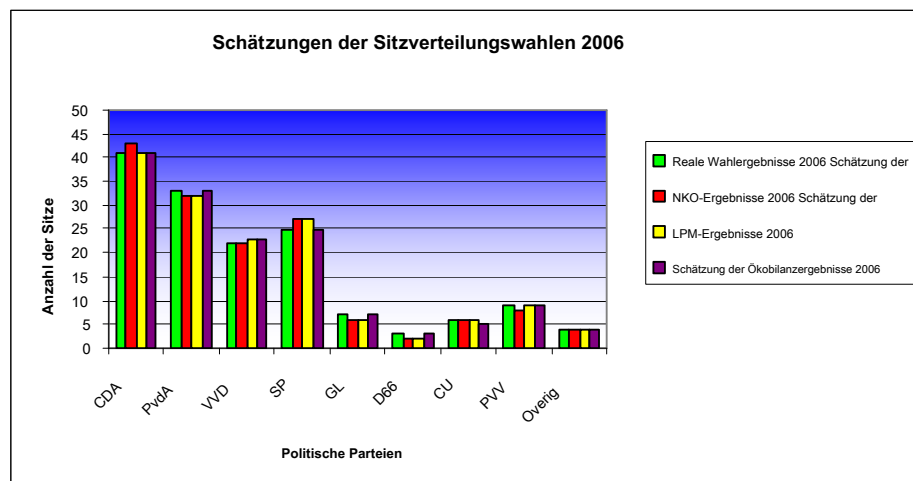


Abbildung 10. Schätzungen der Sitzverteilung im Jahr 2006

Abbildung 10 zeigt die tatsächliche Sitzverteilung im Vergleich zum Modell basierend auf den Umfrageergebnissen und den beiden leistungsstärksten Modellen für LPM und LCA. Es wird deutlich, dass die LCA-Schätzungen nur zweimal von den tatsächlichen Wahlergebnissen abweichen. Der VVD bekommt einen Sitz zu viel, während der CU einen Sitz zu wenig bekommt. Wir stellen fest, dass zur Berechnung dieser Schätzungen die „anderen“ Kategorien mit den tatsächlichen Ergebnissen und nicht mit den Schätzungen ausgeglichen wurden. Da in der Restkategorie alle anderen Parteien kategorisiert sind (siehe Abschnitt 7.1 in Anhang A), werden auch Parteien einbezogen, die die Wahlschwelle nicht überschritten haben. Die Zusammenfassung dieser Stimmen führt zu einer Ungenauigkeit und zu einer großen Anzahl von Stimmen, die bei der Sitzverteilung berücksichtigt werden müssen. Um die Sitzverteilung korrekt zu berechnen, Es sollten nur die Parteien berücksichtigt werden, die die Wahlschwelle überschritten haben. Um die Angelegenheit in dieser Analyse nicht zu kompliziert zu machen, entspricht die Restkategorie der Anzahl der Sitze, die von den Parteien in der Restkategorie gesammelt wurden. Im Jahr 2006 entspricht dies 4 Sitzen, da sowohl der SWP als auch der PvdD 2 Sitze im niederländischen Parlament erhalten haben.

5. Diskussion und weitere Forschung

Die Hauptforschungsfrage, die in diesem Artikel von zentraler Bedeutung war:

*Inwieweit können Abstimmungsübergänge in den Niederlanden auf Gemeindeebene sein?
genau beschrieben, wo die notwendigen Übergangsmatrizen fehlen?*

5.1 Ökologische Folgerung

Wie in der Einleitung erwähnt, besteht ein wichtiges Problem bei Abstimmungsdaten darin, dass sie sich auf aggregierter Ebene befinden. Anstatt die Wahlpräferenzen jedes Einzelnen zu kennen, kennen wir nur die Grenzsummen. Ziel der ökologischen Methodik ist es, die erwarteten Häufigkeiten auf individueller Ebene abzuschätzen. Die Forschung zur aggregierten statistischen Analyse hat lange Bestand. Soziologen waren die ersten, die dieses statistische Problem untersuchten (siehe z. B. Robinson, 1950; Goodman, 1953; Duncan & Davis, 1953). 1950 warf die heftige Kritik von Robinson ein ernstes Problem für diese Art der Analyse auf. Seine Analyse zeigte, dass die Korrelationen auf aggregierter Ebene viel höher waren als die tatsächlichen Korrelationen auf individueller Ebene (Robinson, 1950). Goodmans (1953) Artikel "Ökologische Regression und Verhalten von Individuen" widersprach diesem Gedanken. Mit einigen interessanten Techniken zeigte er, dass es unter bestimmten Umständen möglich ist, Rückschlüsse auf das individuelle Verhalten zu ziehen (Goodman, 1953). In den neunziger Jahren trat mit der Entwicklung neuer Modelle neue Aufmerksamkeit auf (siehe z. B. Freedman et al., 1991; Grofman, 1991; King, 1997; Cho, 1998; Cho & Gaines, 2004; King, 2004). Für einen umfassenden Überblick über die Methodik bis 1995 siehe Cleave et al. (1995). Der Durchbruch von Gary King (1997) hat diesem Forschungsbereich einen neuen Impuls gegeben.

Das größte Problem bei der Vorhersage auf individueller oder übergreifender Ebene unter Verwendung aggregierter Daten ist als Aggregationsbias bekannt (Robinson, 1950). Wenn die Daten nach abhängigen oder unabhängigen Variablen gruppiert werden, tritt eine gruppeninduzierte Korrelation auf. Um dieses Problem zu lösen, kann man davon ausgehen, dass für jeden Bezirk / jede Gemeinde usw. die Abstimmungsraten konstant sind. Dies wird auch als Konstanzannahme bezeichnet, und wenn dies zutrifft, ist die Analyse mit gewöhnlichen kleinsten Quadraten (OLS) unkompliziert. Die Parameter sind konstant, aber es ist nicht sehr wahrscheinlich, dass sie mit keinem Regressor korrelieren. Weil es unwahrscheinlich ist, dass alle Wähler unabhängig von demografischen Faktoren gleich wählen (Cho, 1998, S. 2).

5.1.1 EI-Schätzer

Gary King hat 1997 und 1999 zwei wichtige Bücher über ökologische Schlussfolgerungen geschrieben

2004. Laut King kann die Lösung des ökologischen Inferenzproblems im Zufallskoeffizientenmodell gefunden werden. Es verbessert den Goodman-Schätzer und konzentriert sich auf das sehr häufige Out-of-Bound-Problem. Um dieses Problem zu lösen, verwendet er die Informationen aus den Grenzen. Die erste Stufe eines zweistufigen Verfahrens ist die Bestimmung der Parameter, die das Wahlergebnis einer bestimmten Wahl über Bezirke / Gemeinden hinweg sind. Er geht davon aus, dass es keine Abhängigkeit zwischen dem statistischen Verhalten einer demografischen Gruppe und dem Bezirk gibt, in dem sie sich befinden (Konstanzannahme). In seinem Modell variieren die Parameter gemäß einer abgeschnittenen bivariaten Normalverteilung anstelle einer Normalverteilung, die in OLS verwendet wird. Diese sind so konditioniert, dass sie zwischen 0 und 1 liegen. Er verwendet die zugrunde liegende bivariate Normalverteilung, um lokale Schätzungen zu berechnen, die er dann kombiniert, um aggregierte Pegelschätzungen zu erstellen. Er tut dies durch Addition über die verschiedenen Bezirke (King, 1997). Dennoch funktioniert dieses Modell genauso schlecht wie OLS, wenn eine Aggregationsverzerrung vorliegt (Cho, 2001). Park schreibt 2004 eine interessante Arbeit mit dem Thomsen-Schätzer in einer 2-mal-2-Situation, in der er den EI-Schätzer, die ökologische Regression (Goodman-Schätzer) und die eingeschränkte Regression (Park, 2004) übertrifft. Das EI-Modell wird im Vergleich zu anderen Methoden am häufigsten als Benchmark-Modell verwendet. Park schreibt 2004 eine interessante Arbeit mit dem Thomsen-Schätzer in einer 2-mal-2-Situation, in der er den EI-Schätzer, die ökologische Regression (Goodman-Schätzer) und die eingeschränkte Regression (Park, 2004) übertrifft. Das EI-Modell wird im Vergleich zu anderen Methoden am häufigsten als Benchmark-Modell verwendet. Park schreibt 2004 eine interessante Arbeit mit dem Thomsen-Schätzer in einer 2-mal-2-Situation, in der er den EI-Schätzer, die ökologische Regression (Goodman-Schätzer) und die eingeschränkte Regression (Park, 2004) übertrifft. Das EI-Modell wird im Vergleich zu anderen Methoden am häufigsten als Benchmark-Modell verwendet. Ein Großteil der Forschung zur ökologischen Inferenz konzentriert sich auf 2-mal-2-Systeme, aber in letzter Zeit wurden mehr Methoden für allgemeinere Systeme wie Mehrparteiensysteme entwickelt.

Gibson und Cielecka (1995) schlagen für jede Partei mit demografischen Variablen separate OLS-Modelle vor. Sie behandeln die einzelnen Parteien als abhängige Variablen. Sie verwenden dieses Modell, um die ersten polnischen Wahlen nach der kommunistischen Zeit zu erklären. Sie vergleichen verschiedene demografische Variablen, um Korrelationen mit der Unterstützung der Partei zu finden und zufriedenstellende Erklärungen für die polnische Entwicklung zu finden (Gibson und Cielecka, 1995). Dieses Papier wird später von mehreren Autoren verwendet, um ihr Modell in einer Mehrparteien-Situation im Vergleich zu diesem Modell zu testen.

Katz und King (1999) entwickelten ein anderes Modell für Mehrparteiensysteme. Es eignet sich besonders für aggregierte Wahldaten auf Distriktebene. Es ist auch angebracht, Kontextvariablen einzubeziehen. Der größte Teil der aktuellen Literatur befasst sich mit Zweiparteiensystemen, was die Analyse erheblich erleichtert. In diesem Fall befassen sie sich speziell mit den britischen Wahlen von 1992, bei denen drei Parteien an diesen Wahlen teilgenommen haben. Sie führen einen Markov Chain Monte Carlo (MCMC) -Ansatz ein. Mit diesem Ansatz ist es möglich, bis zu 8 Parteien zu analysieren. Es ist auch anwendbar, Merkmale der aggregierten Bereiche einzubeziehen. Diese Methode eignet sich besser für fehlende Daten

dann Maximum-Likelihood-Schätzung. Die fehlenden Daten werden zunächst geschätzt und in die Daten eingefügt. Die Methode der MCMC schätzt die fehlenden Daten unter Verwendung der maximalen Wahrscheinlichkeit iterativ, bis die Gleichgewichtsverteilung und Konvergenz erreicht ist. Der MCMC-Ansatz ist immer noch recht begrenzt, und im Jahr 2000 wurde dieses Modell erweitert, um mehr Parteien einzubeziehen (Katz und King, 1999). Lewis (2003) verwendet denselben Algorithmus auch, um das Wahlverhalten in einem Zweiparteiensystem zu schätzen, jedoch über mehrere Wahlen hinweg. Die Implementierung des MCMC-Ansatzes auf diese Weise ermöglicht nicht nur eine Analyse über Bezirke hinweg wie in King (1997), sondern auch über Wahlen innerhalb von Bezirken hinweg. Er findet ähnliche Ergebnisse wie King (Lewis, 2003).

Im Jahr 2001 erweiterten Honaker, Katz und King das MCMC-Modell auf mehr als drei Parteien. Sie verwenden ein FIML-Modell (Full Information Maximum Likelihood), um Daten auf Distriktebene zu schätzen. Dieses Modell entspricht der Regression der kleinsten Quadrate in Zweiparteiensystemen und verwendet den EM-Algorithmus, um die Ergebnisse zu approximieren. Sie vergleichen ihr neues Modell mit ihrem eigenen Modell und mit dem Gibson- und Cielacka-Modell. Sie stellen fest, dass ihr Ansatz genauer und schneller ist und auf Mehrparteiensysteme skaliert werden kann (Honaker *et al.*, 2001).

Wellhofer (2001) wendet den EI-Schätzer erfolgreich im italienischen Mehrparteiensystem für die RxC-Kontingenztafel an (Wellhofer, 2001). Dies ist kurz gesagt ein Array, das durch den Schnittpunkt von zwei oder mehr Klassifizierungsvariablen gebildet wird. Die Felder im Array sind mit Beobachtungshäufigkeiten gefüllt (Gunst, 2004). Wellhofer verwendet die Umfrageergebnisse der italienischen Abstimmungsumfrage als Benchmark und kombiniert diese Ergebnisse mit einer politischen Analyse.

Das Problem der ökologischen Folgerung ist ein wichtiges Problem, und diese Studie kann die Diskussion darüber auf andere Weise ergänzen. Es ist jedoch nicht zentral für diese Studie, da wir nicht darauf abzielen, Aussagen über das individuelle Verhalten zu treffen, sondern nur über das Wahlverhalten auf Gemeindeebene.

5.2 Diskussion

Die wichtigste Annahme des Benchmark-Modells der Unabhängigkeit ist, dass zwischen zwei Wahlen eine perfekte Mobilität eines Wählers besteht. In diesem Modell ist die Wahrscheinlichkeit, im zweiten Moment für eine bestimmte Partei zu stimmen, unabhängig von der Wahl der Partei im ersten Moment. Die Analyse zeigt, dass mit diesem Modell die Wahlergebnisse auf Gemeindeebene mit diesem Modell nicht vorhergesagt werden können. Wie erwartet spiegelt auch das Modell, das davon ausgeht, dass alle Wähler völlig loyal sind, nicht die Realität wider. Die Verwendung einer Übergangsmatrix, die direkt aus NKO-Daten geschätzt werden kann, verbessert die Modellanpassung erheblich. Modellbasierte Schätztechniken verbessern diese Ergebnisse noch weiter. Die Ergebnisse zeigen, dass auf der Ebene der Abstimmung

Übergänge NKO und die Schätztechniken für die Modellbasis sind sehr ähnlich. Ein großer Vorteil von NKO ist, dass es die tatsächliche Mobilität der Wähler abbilden kann. Für jeden Befragten kann angezeigt werden, welche Partei er gewählt hat. Daher können Übergänge zwischen verschiedenen Parteien in beide Richtungen analysiert werden. Modellbasierte Schätztechniken spiegeln nur die Nettoübergänge zwischen Parteien wider. Auf diese Weise verringern sich Übergänge zwischen zwei Parteien häufig, und folglich enthält die Übergangsmatrix viele Nullen. Dies hängt eng mit dem Problem der ökologischen Folgerung zusammen (King,

1997, 2004).

Aggregierte Daten sind in der Regel einfacher zu erhalten als einzelne Daten und können wertvolle Hinweise auf das individuelle Verhalten geben. Ökologische Schlussfolgerungen werden daher weiterhin gezogen. Die Probleme der Verwirrung und der Aggregationsverzerrung werden wahrscheinlich in naher Zukunft nicht gelöst (Freedman, 1999). Umfrageforschung ist daher nach wie vor die einzige Möglichkeit, verlässliche Ergebnisse zum Verhalten von Personen zu erhalten.

Ein Vorteil der modellbasierten Forschung besteht darin, dass sie im Vergleich zur Umfrageforschung eine billige Technik ist. kommunale Ergebnisse sind immer verfügbar. Ein zweiter Vorteil besteht darin, dass mit modellbasierten Schätztechniken die Gesamtbevölkerung der Niederlande beobachtet werden kann, während die NKO-Forschung eine Stichprobenerhebung ist, bei der das Abstimmungsverhalten auf die Gesamtbevölkerung hochgerechnet wird. Auf diese Weise ermöglichte die modellbasierte Forschung die Beobachtung eines interessanten regionalen Verhaltens. In den Schätzungen wird deutlich, welchen Einfluss das Wahlverhalten bestimmter Politiker in den Regionen hat, aus denen sie stammen. Auch die Schwierigkeit, Übergänge auf den niederländischen Inseln zu erklären, wird sichtbar. Andere rein wahlspezifische Elemente werden besser visualisiert. So haben sich 2006 die Abstimmungsübergänge in Limburg deutlich von denen anderer Bundesländer unterschieden.

Die Ergebnisse der modellbasierten Techniken der quadratischen Programmierung und der Analyse latenter Klassen sind vergleichbar und es gibt nur einen geringen Unterschied in der Modellanpassung. Die Kombinationsmodelle bieten ebenfalls eine gute Passform, waren jedoch etwas schlechter und zeigten keine realistischeren Ergebnisse.

Die wichtigste Schlussfolgerung dieser Forschung ist, dass sich die verschiedenen Methoden (Stichprobenerhebung und modellbasiert) in gewisser Weise ergänzen. Die Ergebnisse der Methoden können sich gegenseitig bestätigen, wodurch die Validität der Ergebnisse verbessert wird. Darüber hinaus zeigt diese Untersuchung, dass die Annahme, dass es eine einzige Übergangsmatrix gibt, die die Grundlage für das Wahlverhalten aller niederländischen Wähler bildet, nicht nachhaltig genug ist. Zukünftige Forschungen müssen durchgeführt werden, um den Einfluss der Feinabstimmung der Modelle zum Erhalt der Übergangsmatrizen zu untersuchen. Mögliche Forschungsrichtungen sind: Erstens die Verwendung unterschiedlicher Matrizen für unterschiedliche Regionen oder für unterschiedliche Verstärkungsgrade. Zweitens andere Optionen wie selbstorganisierende Karten und Stochastic

Globale Suchmethoden, wie die Verwendung des Überlebens der Stärksten innerhalb genetischer Algorithmen, können sich als wertvolle Optimierungstechniken erweisen. Selbstorganisierende Karten sind eine Technik in künstlichen neuronalen Netzen. Diese Technik kann verwendet werden, um logische Gruppen von Gemeinden zu bilden, die verwendet werden können, um eine Aufteilung in Regionen zu verbessern und mehrere Übergangsmatrizen zu erstellen. Diese Techniken befinden sich auf dem Gebiet der künstlichen Intelligenz, einem anderen Forschungsgebiet. Drittens könnte es interessant sein, Cluster auf der Grundlage der Wahlunterstützung für eine bestimmte einflussreiche Partei zu definieren. Diese Technik ist möglicherweise besser für Systeme mit zwei oder drei Parteien anwendbar, kann jedoch auch in Systemen mit mehreren Parteien angewendet werden.

6. Bibliographie

- Aarts, K., H. van der Kolk und M. Rosema (2007). *Een verdeeld Electoraat De Tweede Kamerverkiezingen van 2006*, Spektrum.
- Aarts, K. & H. van der Kolk (2007). Die Parlamentswahlen in den Niederlanden, 22. November 2006, Anmerkungen zu den jüngsten Wahlen / Wahlstudien **26**: 797-837.
- Agresti, A. (1996). *Eine Einführung in die kategoriale Datenanalyse*, John Wiley & Söhne, New York.
- Barbosa, MF & amp; H. Goldstein (2000). Mehrstufige Modelle mit diskreter Antwort für Wiederholte Maßnahmen: Eine Anwendung auf Daten zu Abstimmungsabsichten, *Qualität & Quantität* **34**: 323-33.
- Bischof, YMM, SE Fienberg & PW Holland (1975). *Diskrete multivariate Analyse. Theorie und Praxis*. Cambridge: MIT Press.
- Betlehem JG & amp; HMP Kersten (1986). *Werken traf Nonresponse (Arbeiten mit Nichtbeantwortung)*. Voorburg: Centraal Bureau voor de Statistiek.
- Blumen, J., M. Kogan und PJ McCarthy (1955). *Die industrielle Mobilität der Arbeit als ein Wahrscheinlichkeitsprozess*, Cornell-Studien zu Arbeits- und Arbeitsbeziehungen (6). Ithaka, New York: Cornell University Press.
- Cameron, AC & FAG Windmeijer (1993). Abweichungsbasierte R-Quadrat-Messungen der Anpassungsgüte für verallgemeinerte lineare Modelle, Arbeitspapiere in Ökonomie und Ökonometrie.
- CBS: Statline (2008). Overheid en politiek, Verkiezingen en politiek, Tweede Kamerverkiezingen, URL: <http://statline.cbs.nl/StatWeb/dome/?LA=NL> (Zuletzt besucht am: 04. Juni 2008).
- Cho, WKT (1998). Wenn die Annahmen passen ...: ein Kommentar zum König ökologisch Inferenzlösung, *Politische Analyse* **7**: 143–163.
- Cho, WKT & AH Yoon (2001). Seltsame Bettgenossen: Politik, Gerichte und Statistik: Statistisches Expertenzeugnis in Fällen von Stimmrechten, *Cornell Journal of Law und öffentliche Ordnung*, **102**: 237-64.
- Cho, WKT & Gaines, BJ (2004). Die Grenzen der ökologischen Folgerung: Der Fall der Split-Ticket-Abstimmung, *Amerikanisches Journal für Politikwissenschaft*, **48** (1): 152–171.

- Cleave, N., PJ Brown und CD Payne (1995), Evaluation of Methods for
Ökologische Folgerung, *Zeitschrift der Royal Statistical Society. Serie A (Statistik in der Gesellschaft)*, **158**:
55-72.
- Clogg, CC & amp; LA Goodman (1985). Simultane latente Strukturanalyse in
mehrere Gruppen. In: NB Tuma, Hrsg., *Soziologische Methodik 1985*. San Francisco,
Jossey-Bass: 81-11.
- Cook, RJ, JD Kalbfleisch und GY Yi (2002). Ein verallgemeinerter Mover-Stayer
Modell für Paneldaten. *Bio-Statistiken*, **3**: 407-420.
- Dempster, AP, NM Laird und DB Rubin (1977). Maximale Wahrscheinlichkeit von
unvollständige Daten über den EM-Algorithmus (mit Diskussion), *Zeitschrift der Royal Statistical Society,
Reihe B, Methodological*, **39**: 1 - 37.
- Duncan, OD und Davis, B. (1953). Eine Alternative zur ökologischen Korrelation.
American Sociological Review, **18**: 665–666.
- Egmond, van M., ND de Graaf und C. van der Eijk (1998). Wahlbeteiligung an
Niederlande: Individuelle und kontextuelle Einflüsse, *Europäische Zeitschrift für politische Forschung* **34**:
281-300.
- DA Freedman, SP Klein, J. Sacks, CA Smyth & CG Everett (1991).
Ökologische Regression und Stimmrechte (mit Diskussion). *Bewertung Bewertung*,
15: 673–816.
- Freedman, DA (1999). Ökologische Folgerung und der ökologische Irrtum,
Internationale Enzyklopädie der Sozial- und Verhaltenswissenschaften, Ed. Smelser
Technischer Bericht von NJ & PB Baltes, **549 (6)**: 4027-403.
- Gibson, J. & amp; A. Cielecka (1995). Wirtschaftliche Einflüsse auf die politische Unterstützung für
Marktreformen bei postkommunistischen Übergängen: Einige Belege aus den polnischen
Parlamentswahlen 1993, *Europa-Asien-Studien*, **47 (5)**: 765-785.
- A. Greenwald, C. Carnot, R. Beach & B. Young (1987). Zunehmendes Wahlverhalten
indem man die Leute fragt, ob sie eine Abstimmung erwarten, *Zeitschrift für Angewandte Psychologie*, **72**:
315-318.
- Goodman, L. (1953). Ökologische Regressionen und Verhalten von Individuen. *amerikanisch
Soziologische Überprüfung*, **18**: 663–666.
- Goodman, LA (1961). Statistische Methoden für das Mover-Stayer-Modell, *Zeitschrift für
die American Statistical Association*, **56 (296)**: 841-868.

- Goodman, LA (1968). Die Analyse von Cross-Classified-Daten: Unabhängigkeit, Quasi Unabhängigkeit und Interaktionen in Kontingenztabellen mit oder ohne fehlende Werte, *Zeitschrift American Statistical Association*, **63**: 1091-1131.
- Grofman, B. (1991). Statistik ohne Substanz: Eine Kritik von Freedman et al. und Clark und Morrison, *Bewertung Bewertung*, **15 (6)**: 746-769.
- Gunst, MCM (2004). *Statistische Modelle*, Vrije Universiteit Amsterdam.
- Honaker, J., J. N. Katz & G. King (2001). Ein verbessertes statistisches Modell für Mehrparteien-Wahldaten, Working Papers 1111, California Institute of Technology, Abteilung für Geistes- und Sozialwissenschaften.
- Irwin, GA & DA Meeter (1969). Aufbau von Wählerübergangsmodellen aus Daten sammeln, *Midwest Journal Politikwissenschaft*, **13**: 545-566.
- Janssen, T. (2006). "Wouter Bos-PvdA", URL: <http://www.tomjanssen.net/> (Letzte besucht am: 23. April 2008).
- Richter, GG & Takayama, T. (1966). Ungleichheitsbeschränkungen in der Regressionsanalyse. *Journal American Statistical Association*., **61**: 166-181.
- Kampen, J. & M. Swyngedouw (2000). Die ordinale Kontroverse wurde erneut aufgegriffen, *Qualität und Menge* **34 (1)**: 87-102.
- Katz, JN & G. King (1999). Ein statistisches Modell für Mehrparteien-Wahldaten, *Das American Political Science Review*, **93 (1)**: 15-32.
- Kiesraad.nl (2008). Kiesraad.nl -- Verkiezingsuitslagen, URL: <http://www.verkiezingsuitslagen.nl> (Zuletzt besucht am: 17. Januar 2008).
- Keller, WJ & A. ten Cate (1977). De verschuiving van de kiezersvoorkeur, *Economische Statistische Berichten* 26-10-1977.
- King, G. (1997). *Eine Lösung für das ökologische Inferenzproblem. Rekonstruieren Individuelles Verhalten aus aggregierten Daten*, Cambridge MA, Cambridge University Press.
- King, G., O. Rosene & M. Tanner (2004). *Ökologische Folgerung. Neue Methodik Strategien*, Cambridge University Press, Cambridge.
- King, GO (1996). "Wahlen und die nationalen Wahlstudien", Papier vorbereitet für die Nationalen Wahlstudien, Kongresswahlkonferenz.
- Langeheime, R. & F. van de Pol (1990). Diskrete zeitlich gemischte Markov-Latentklasse Modelle, *Soziologische Methodik*, **20**: 213-247.

- Lewis, JB (2003). Erweiterung des ökologischen Inferenzmodells von King auf mehrere Wahlen mit Markov Chain Monte Carlo, Kapitel in Gary King, Ori Rosen, und Martin Tanner, Eds. *Ökologische Folgerung: Neue methodische Strategien*. Cambridge: Cambridge University Press. 2004.
- Little, RJA & DB Rubin (1987). *Statistische Inferenz mit fehlenden Daten*. New York: Wiley.
- Long, JS (1997). *Regressionsmodelle für kategoriale und begrenzte abhängige Variablen*. Thousand Oaks, LA: Salbei.
- McCarthy, C. & TM Ryan (1977). Schätzung der Wählerübergangswahrscheinlichkeiten aus die britischen Parlamentswahlen von 1974, *Zeitschrift der Royal Statistical Society Reihe A*. **140**: 78-85.
- McFadden, D. (1973). Bedingte Logit-Analyse des qualitativen Auswahlverhaltens, in Zarembka, P. (Hrsg.), *Frontiers in Econometrics*, S. 105-142, Academic Press, New York.
- Nationaal Kiezersonderzoek (2002/2003). Niederländische Parlamentswahlstudie 2002/2003 (NKO 2002/2003), DPES 2002/2003.
- Oosterhoff, J. & AW van der Vaart (2003). *Algemene Statistiek*, Vrije Universiteit Amsterdam.
- Park, WH. (2004). „Schätzung der Wählerübergangsraten und der ökologischen Inferenz“, vorgestellt auf der Jahreskonferenz 2003 der Midwest Political Science Association. Chicago, April 2003.
- Ploeg, CE van der, F. van de Pol und JK Kampen, (2008), De verschuivingen van de partijvoorkeur van de Nederlandse kiezers tussen de nationale verkiezingen van 2003 en 2006, Zur Veröffentlichung angeboten in: *Het Nationaal Kiezersonderzoek 2006: opzet, uitvoering en resultaten*, Schmeets, H. (Hrsg.), CBS Nederland.
- Pol, van de F. & R. Langeheime (2004). *Latente Übergangs- und Markov-Modelle*, Statistik Niederlande.
- Poulsen. CS (1982). "Latente Strukturanalyse mit Auswahlmodellierung Bewerbungen." Ph.D. Diss., Wharton School, Universität von Pennsylvania.
- Robinson, WS (1950). Ökologische Zusammenhänge und das Verhalten von Individuen. *American Sociological Review*, **15**: 351–357.
- Shachar, R. & M. Shamir, (1996), Estimating Vote Persistence Sources without Paneldaten, *Politische Analyse*, **6**: 107-124.

- Staatsblad (2003). "Wet op het Centraal Bureau voor de Statistiek" vom 20. November 2003: 516.
- Thomsen, SR (1987). *Dänische Wahlen 1920–79: Ein logistischer Ansatz für ökologische Analyse und Inferenz*. Aarhus: Politica.
- Thomsen, SR (2004). "Individuelles Wahlverhalten in Dänemark 1998-2001", Papier vorgestellt auf der Jahrestagung der Midwest Political Science Association, Chicago, April 2004.
- Tijms, HC & AAN Ridder, (2003), *Mathematisch Programmeren*, Vrije Universität Amsterdam.
- Quinn, KV, AD Martin & AB Whitford (1999). Wahl der Wähler in Multi-Party Demokratien: Ein Test konkurrierender Theorien und Modelle, *Amerikanisches Journal für Politikwissenschaft*, **43** (4): 1231-1247.
- Upton, GJG (1977). Ein Speichermodell für Abstimmungsübergänge bei britischen Wahlen, *Zeitschrift der Royal Statistical Society Reihe A*, **140**: 86-94.
- Veall, MR & KL Zimmermann (1996). Pseudo-R² Maßnahmen für einige gemeinsame Begrenzte abhängige variable Modelle, *Sonderforschungsbereich*, **386**, Papier-18, Institut für Statistik.
- Vermunt, JK, R. Langeheime und U. Bockenholt (1999). Diskreter zeitdiskreter Zustand latente Markov-Modelle mit zeitkonstanten und zeitvariablen Kovariaten, *Zeitschrift für Bildungs- und Verhaltensstatistik*, **24**: 178-205.
- Voogt, R. (2004). *Ich bin nicht interessiert - Nonresponsive Bias, Response Bias und Anregungseffekte in der Wahlforschung*, Universität van Amsterdam.
- Weir, BT (1975). Die Verzerrung des Wählerrückrufs, *American Journal of Political Wissenschaft*, **14** (1): 53-62.
- Wellhofer, ES (2001). Neuausrichtung der Partei und Wählerwechsel in Italien, 1987-1996, *Vergleichende politische Studien*, **34** (2): 156-86.
- Wiggins, LM (1955) 1973. Panel-Analyse, latente Wahrscheinlichkeitsmodelle für die Einstellung und Verhaltensprozesse, *Elsevier*, Amsterdam.
- Wikipedia: Bijbelgordel (2008). Bijbelgordel (Bibel Gürtel) URL: <http://nl.wikipedia.org/wiki/Bijbelgordel> (letzte Aktualisierung: 3. Mai 2008).

7. Anhang A.

7.1 Analyse der Parteien

Für 1998:

Cristian Democrats (CDA), Labour Party (PvdA), Liberale Partei (VVD), Sozialistische Partei (SP), Grüne Linke (GL), Demokraten 66 (D66), Reformierte Politische Union und Reformatorische Politische Föderation (GPV und RPF = CU), Christlich-Reformierte Partei (SGP), Andere (Zentrumsdemokraten (CD), Allgemeine Altenunion / Union 55+ (AOV / U55 +), Netherlands Mobile (NMob), Senioren 2000 (S2000), Niederländische Mittelklasse-Partei (NMP), Die Grünen (Groenen), die Naturrechtspartei (NWP), die katholische politische Partei (KPP), die Freie Indische Partei (VIP), die Neue Vereinigte Alte Union (NSOV), die Neue Kommunistische Partei (NCPN), Idealisten / Sie (IdeA / JIJ)), The Voters Collective (KColl)), nicht gewählt. Insgesamt sind 10 Parteien.

Für 2002:

Christdemokraten (CDA), Labour Party (PvdA), Liberale Partei (VVD), Sozialistische Partei (SP), Grüne Linke (GL), Demokraten 66 (D66), Christliche Union (CU), Liste Pim Fortuyn (LPF), Andere (Christlich-Reformierte Partei (SGP), Lebenswerte Niederlande (LN), Vereinigte Seniorenpartei (VSP), Freie Indische Partei & Alte-Union (VIP / OU), Dauerhafte Niederlande (DN), Partei der Zukunft (PvdT), Neues Zentrum Partei (NMP), Republikanische Volkspartei (RVP)), nicht gewählt. Insgesamt sind 10 Parteien.

Für 2003:

Christdemokraten (CDA), Labour Party (PvdA), Liberale Partei (VVD), Sozialistische Partei (SP), Grüne Linke (GL), Demokraten 66 (D66), Christliche Union (CU), Liste Pim Fortuyn (LPF), Andere (Partei für Tiere (PVDD), Christlich-Reformierte Partei (SGP), Allianz für Erneuerung und Demokratie (AVD), Conservatives.nl (Conservatieven), Durable Nederland (DN), Lebenswerte Niederlande (LN), List Veldhoen (Veldhoen), New Kommunistische Partei (NCPN), Partei der Zukunft (PvdT), List Ratelband (Ratelb), Progressive Integration Party (VIP)), nicht gewählt. Insgesamt sind 10 Parteien.

Für 2006:

Cristian Democrats (CDA), Labour Party (PvdA), Liberale Partei (VVD), Sozialistische Partei (SP), Grüne Linke (GL), Demokraten 66 (D66), Cristian Union (CU), List Pim Fortuyn (LPF), Andere (Partei für Tiere (PVDD), Christlich-Reformierte Partei (SGP), Liste Fünf Fortuyn (Fortuyn = LPF), Niederlande Transparent (NT), OneNL (EenNL), Liste Poortman, Partei für die Niederlande (PVN), Kontinuierliche Direkt demokratische Partei (CDDP), Liberaldemokratische Partei (LDP), Vereinigte Ältere Partei (VSP), Ad Bos

Kollektiv (Ad Bos), Grüne freie Internetpartei (GVIP), Liste Potmis = Islamdemokraten (ID), Tamaras Offene Partei (TOP), Solide Multikulturelle Partei (SMP), LRVP-Sitz im Parlament (LRVP)), Nicht gewählt. Insgesamt sind 10 Parteien.

7.2 Übergangsmatrizen 1998-2002, 2003-2006

Alle Übergangsmatrizen aller Modelle aus den Wahlen 1998-2002 und 2002-2003 sind nachstehend aufgeführt.

1998-2002

Tabelle 10. Übergangsmatrix NKOModel 1998-2002 (diagonale Elemente fett gedruckt)

1998	2002									
	CDA	PVDA	VVD	SP	GL	D66	CU	LPF	Anders	NG
CDA	.7602	.0045	.0317	.0090	.0090	.0136	.0136	.0860	.0090	.0633
PVDA	.1329	.4860	.0455	.0559	.0734	.0245		.0874	.0105	.0839
VVD	.1752	.0109	.5109	.0109	.0036	.0109	.0036	.2482	.0255	
SP	.0250	.0750		.5000	.0250			.2500		.1250
GL	.0400	.0800	.0267	.1333	.5067	.0533		.1067		.0533
D66	.1511	.1223	.1007	.0576	.0719	.2158		.0576	.0072	.2158
CU	.3409		.0227	.0227			.5455	.0682		
SGP	.0625								.9375	
Anders	.1628	.0698	.1395	.0465	.0930	.0233	.0233	.1395	.0930	.2093
NG	.0863	.0288	.0144	.0288	.0144	.0072	.0072	.1475	.0036	.6619

Quelle: (Nationaal Kiezersonderzoek, 2002/2003)

Tabelle 11. Übergangsmatrix LRmodel 1998-2002 (diagonale Elemente fett gedruckt)

1998	2002									
	CDA	PVDA	VVD	SP	GL	D66	CU	LPF	Anders	NG
CDA	1,3062	-, 1981	. 0431	. 0258	. 0238	. 0147	-, 0053	-, 3159	. 0093	. 0025
PVDA	. 0404	. 6233	. 0189	-, 0102	. 0199	-, 0007	. 0111	. 2130	. 0234	. 0438
VVD	. 2844	-, 1525	. 6047	-, 0381	. 0108	. 0342	-, 0164	. 6647	. 0646	-, 0789
SP	. 3817	-, 4568	-, 0864	. 6706	. 1180	. 0306	-, 0102	. 5405	. 0026	-, 0463
GL	-, 1002	. 1678	-, 0383	. 4966	1,0296	. 0684	. 0157	-, 2921	. 0924	. 0318
D66	-, 1055	. 4809	. 4405	. 2192	. 1002	. 5442	. 0222 -1,1844		-, 1227	-, 0203
CU	. 0333	. 2889	-, 0260	-, 0150	. 0115	-, 0233	. 8810	. 1450	. 0012	. 0431
SGP	. 1885	-, 1185	-, 1463	-, 0151	-, 0404	-, 0370	-, 0637	. 1089	1,0155	-, 1511
Anders	. 3863	-, 2061	-, 8131	. 4308	-, 5399	-, 2416	. 0711	. 7814	. 4833	. 5230
NG	-, 0708	. 0544	. 0042	-, 0553	-, 0175	-, 0012	-, 0083	. 2678	-, 0236	. 7840

Quelle: (Eigene Berechnungen des Autors)

Tabelle 12. Übergangsmatrix QPmodel 1998-2002 (diagonale Elemente fett gedruckt)

1998	2002									
	CDA	PVDA	VVD	SP	GL	D66	CU	LPF	Anders	NG
CDA	1.0000									
PVDA	. 0821	. 5237	. 025	. 1069	. 0615	. 044	. 0165	. 0261	. 0716	. 0427
VVD	. 2004		. 4768					. 30000	. 0229	
SP										1.0000
GL				. 3093	. 6907					
D66			. 5143		. 0312	. 4545				
CU	. 3398						. 6602			
SGP							. 0179	. 0289	. 9531	
Anders								1.0000		
NG	. 1332			. 0175				. 2003		. 6489

Quelle: (Eigene Berechnungen des Autors)

Tabelle 13. Übergangsmatrix LCAmodel 1998-2002 (diagonale Elemente fett gedruckt)

1998	2002									
	CDA	PVDA	VVD	SP	GL	D66	CU	LPF	Anders	NG
CDA	1.0000									
PVDA	.1668	.5291	.0917	.0314	.0279	.0230	.0014		.0242	.1046
VVD	.2045		.4997					.2733	.0226	
SP				.7349	.0830	.0074				.1747
GL		.1543		.1522	.6930		.0005			
D66		.0007	.1831	.1706	.1465	.4992				
CU	.1920						.7821		.0259	
SGP	.0548							.0363	.9089	
Anders								.6984	.3016	
NG	.0163			.0066		.0089	.0023	.2585	.0101	.6973

Quelle: (Private Mitteilung von F. van de Pol)

Tabelle 14. Übergangsmatrix IPF-Modell 1998-2002 (diagonale Elemente fett gedruckt)

1998	2002									
	CDA	PVDA	VVD	SP	GL	D66	CU	LPF	Anders	NG
CDA	.2765	.1078	.1252	.0396	.0446	.0360	.0242	.1295	.0375	.1789
PVDA	.2591	.1146	.1273	.0411	.0465	.0372	.0223	.1316	.0366	.1837
VVD	.2594	.1097	.1341	.0404	.0462	.0383	.0217	.1346	.0374	.1782
SP	.2571	.1130	.1250	.0439	.0472	.0372	.0202	.1328	.0341	.1895
GL	.2566	.1134	.1299	.0420	.0486	.0386	.0210	.1322	.0346	.1831
D66	.2565	.1121	.1338	.0415	.0479	.0391	.0213	.1331	.0355	.1792
CU	.2696	.1090	.1191	.0362	.0412	.0336	.0403	.1256	.0555	.1699
SGP	.2682	.1018	.1175	.0330	.0359	.0311	.0415	.1306	.0761	.1644
Anders	.2587	.1104	.1282	.0410	.0462	.0371	.0214	.1348	.0370	.1853
NG	.2593	.1110	.1261	.0410	.0461	.0369	.0219	.1332	.0368	.1878

Quelle: (Eigene Berechnungen des Autors)

Tabelle 15. Übergangsmatrix Kombination 1 Modell 1998-2002 (diagonale Elemente fett gedruckt)

1998	2002									
	CDA	PVDA	VVD	SP	GL	D66	CU	LPF	Anders	NG
CDA	.8806		.0199	.0080		.0059	.0411	.0174	.0272	
PVDA	.1511	.5236	.0074	.0758	.0724	.038		.0227	.0539	.0551
VVD	.2095		.4430			.0279		.2991	.0205	
SP				.4140						.5860
GL				.3108	.6210	.0681				
D66			.6274	.0061	.0445	.3220				
CU	.5060						.4940			
SGP							.0889		.9111	
Anders								.9797	.0203	
NG	.1284							.1958		.6758

Quelle: (Nationaal Kiezersonderzoek, 2002/2003 und eigene Berechnungen des Autors)

Tabelle 16. Übergangsmatrix Kombination 2 Modell 1998-2002 (diagonale Elemente fett gedruckt)

1998	2002									
	CDA	PVDA	VVD	SP	GL	D66	CU	LPF	Anders	NG
CDA	.8806		.0199	.0080		.0059	.0411	.0174	.0272	
PVDA	.1511	.5236	.0074	.0758	.0724	.038		.0227	.0539	.0551
VVD	.2095		.4430			.0279		.2991	.0205	
SP				.4140						.5860
GL				.3108	.6210	.0681				
D66			.6274	.0061	.0445	.3220				
CU	.5060						.4940			
SGP							.0889		.9111	
Anders								.9797	.0203	
NG	.1284							.1958		.6758

Quelle: (Nationaal Kiezersonderzoek, 2002/2003 und eigene Berechnungen des Autors)

Wie aus diesen Übergangsmatrizen ersichtlich ist, nehmen SGP und LPF die gleiche Position in der Matrix ein. Dies führt zu einem sehr niedrigen Wert oder Nullwert in der Diagonale für den Übergang zwischen SGP und LPF. Kleinere Probleme treten beim Stayer-Modell auf, bei dem davon ausgegangen wird, dass jeder Wähler, der zuvor SGP gewählt hat, dafür stimmen wird

wieder feiern. Da die LPF jedoch im Jahr 2002 eine neue Partei ist, kann das Stayer-Modell dies in keiner Weise berücksichtigen, was zu einer enormen Fehlanpassung führt.

Alle Modelle weisen einen großen Verlust für die PvdA auf, was mit den Wahlergebnissen übereinstimmt. Es ist interessant zu sehen, dass der andere Gewinner der Wahlen im Jahr 2002, der CDA, einen niedrigeren Wert auf der Diagonale der NKO-Matrix hat als die Werte auf der Diagonale der modellbasierten Matrizen. Dies kann wiederum durch die Fähigkeit der Umfrageforschung erklärt werden, Übergänge zwischen Parteien und nicht nur die Nettotransitionen darzustellen, wie dies bei modellbasierten Ansätzen der Fall ist. Da es schwierig ist, die Übergänge anders darzustellen, muss man die Unterstützung für die LPF aus der anderen Kategorie ableiten und in der vorherigen Wahlkategorie nicht gewählt haben. Es wird oft gesagt, dass Pim Fortuyn Nichtwähler zur Abstimmung angezogen hat, und dies ist in der Tat in den Übergangsmatrizen und in den Wahlbeteiligungsniveaus dieses Wahljahres sichtbar.

In der QP-Matrix mit dem SP zeigt sich etwas Seltsames, was ich nicht erklären kann. Auch in diesem Fall sind beide Kombinationsmatrizen gleich. Dies kann durch die Tatsache erklärt werden, dass der Optimierungsalgorithmus versucht, unter Berücksichtigung der Einschränkungen eine optimale Lösung zu finden. In diesem Fall sind diese Einschränkungen nicht restriktiv genug.

2002-2003

Tabelle 17. Übergangsmatrix NKO Modell 2002-2003 (diagonale Elemente fett gedruckt)

2002	2003									
	CDA	PVDA	VVD	SP	GL	D66	CU	LPF	Anders	NG
CDA	.7647	.0623	.0623	.0035	.0069	.0069	.0069	.0069	.0104	.0692
PVDA	.0415	.8238	.0104	.0259	.0207	.0104				.0674
VVD	.0476	.0595	.7321	.0060	.0060	.0060		.0060		.1369
SP		.1719		.6719	.0469	.0469				.0625
GL	.0135	.3649		.1351	.4324	.0135	.0135			.0270
D66	.0566	.2830	.0566		.0377	.5660				
CU	.2188	.0625			.0625		.6563			
LPF	.1227	.1104	.1411	.0123	.0123	.0061		.2822	.0307	.2822
Anders	.1333	.1000	.0667					.0667	.5667	.0667
NG	.0842	.0632	.0526	.0158	.0158	.0053	.0053	.0211	.0158	.7211

Quelle: (Nationaal kiezersonderzoek, 2002/2003)

Tabelle 18. Übergangsmatrix LRmodel 2002-2003 (diagonale Elemente fett gedruckt)

2002	2003									
	CDA	PVDA	VVD	SP	GL	D66	CU	LPF	Anders	NG
CDA	1,0338	. 1171	-. 0098	. 0487	-. 0362	-. 0322	. 0031	-. 0418	-. 0222	-. 0747
PVDA	-. 1405	1,4444	-. 0942	-. 0654	-. 0079	-. 0087	. 0005	. 0579	. 0098	-. 2459
VVD	. 1601	. 0283	1,1449	-. 0028	-. 0225	. 0043	-. 0016	-. 1260	-. 0546	-. 0150
SP	. 0799	. 0748	. 0028	1,0777	-. 0099	-. 0360	-. 0075	-. 0916	-. 0269	-. 1552
GL	. 1284	. 1669	-. 0243	. 0520	. 7543	. 0567	. 0226	-. 0270	-. 0386	. 1491
D66	-. 4574	-. 3375	. 1099	-. 2421	. 2289	. 9003	-. 0207	. 0795	. 1073	. 3610
CU	. 0114	-. 0587	. 0412	-. 0050	. 0311	. 0210	. 8825	. 0597	. 0273	. 0809
LPF	-. 0526	. 0774	. 1590	. 0279	. 0042	. 0185	. 0100	. 5654	. 0145	. 2131
Anders	. 2460	-. 0983	. 0305	-. 0666	. 0393	. 0145	-. 0463	-. 0371	. 9143	-. 0312
NG	. 0185	. 0605	-. 0410	. 0164	-. 0005	-. 0046	-. 0048	-. 0449	. 0028	. 9753

Quelle: (Eigene Berechnungen des Autors)

Tabelle 19. Übergangsmatrix QPmodel 2002-2003 (diagonale Elemente fett gedruckt)

2002	2003									
	CDA	PVDA	VVD	SP	GL	D66	CU	LPF	Anders	NG
CDA	1.0000									
PVDA		1.0000								
VVD			1.0000							
SP		. 3637		. 566						. 0703
GL		. 7137			. 2863					
D66					. 3237	. 6763				
CU	. 1433	. 0828					. 7738			
LPF	. 0348		. 1685	. 1059	. 0623	. 0304	. 0068	. 3222		. 2692
Anders	. 2206		. 0101				. 0127	. 0117	. 7449	
NG		. 2334		. 0434						. 7232

Quelle: (Eigene Berechnungen des Autors)

Tabelle 20. Übergangsmatrix LCAmodel 2002-2003 (diagonale Elemente fett gedruckt)

2002	2003									
	CDA	PVDA	VVD	SP	GL	D66	CU	LPF	Anders	NG
CDA	.9960			.0040						
PVDA		1.0000								
VVD			1.0000							
SP		.0970		.9020						.0010
GL		.2940			.6640					.0420
D66		.1010			.0960	.8040				
CU	.0480	.0930			.0080		.8510			
LPF		.2120	.1570	.0090	.0040	.0020	.0010	.3430		.2720
Anders	.2500		.0220						.7280	
NG		.1990		.0300						.7700

Quelle: (Private Mitteilung von F. van de Pol)

Tabelle 21. Übergangsmatrix IPF-Modell 2002-2003 (diagonale Elemente fett gedruckt)

2002	2003									
	CDA	PVDA	VVD	SP	GL	D66	CU	LPF	Anders	NG
CDA	.2900	.1978	.1486	.0442	.0327	.0285	.0198	.0424	.0285	.1675
PVDA	.2680	.2130	.1498	.0454	.0344	.0295	.0193	.0426	.0270	.1711
VVD	.2738	.2018	.1587	.0444	.0343	.0308	.0179	.0437	.0269	.1677
SP	.2709	.2073	.1510	.0473	.0348	.0300	.0175	.0430	.0253	.1728
GL	.2694	.2084	.1525	.0463	.0358	.0307	.0177	.0426	.0247	.1718
D66	.2709	.2048	.1566	.0454	.0352	.0312	.0176	.0433	.0255	.1694
CU	.2891	.1972	.1400	.0398	.0311	.0267	.0332	.0411	.0434	.1586
LPF	.2749	.2013	.1534	.0447	.0334	.0295	.0182	.0447	.0282	.1716
Anders	.2833	.1920	.1460	.0404	.0304	.0270	.0268	.0437	.0482	.1621
NG	.2738	.2059	.1478	.0458	.0336	.0288	.0182	.0433	.0268	.1759

Quelle: (Eigene Berechnungen des Autors)

Tabelle 22. Übergangsmatrix Kombination 1 Modell 2002-2003 (diagonale Elemente fett gedruckt)

2002	2003									
	CDA	PVDA	VVD	SP	GL	D66	CU	LPF	Anders	NG
CDA	.8858	.0157	.0179	.0349	.0153	.0036	.0145	.0107	.0015	
PVDA		.9462		.0173	.0192					.0173
VVD			.8529		.0672	.0800				
SP		.4086		.5914						
GL		.6553			.3447					
D66		.013	.4296		.0748	.4825				
CU	.3460						.5755		.0785	
LPF	.2273		.154	.0329				.3037		.2821
Anders	.2370	.0146	.0117				.0534		.6832	
NG		.2518		.0291						.719

Quelle: (Nationaal Kiezersonderzoek, 2002/2003 und eigene Berechnungen des Autors)

Tabelle 23. Modell 2002-2003 der Übergangsmatrixkombination 2 (diagonale Elemente fett gedruckt)

2002	2003									
	CDA	PVDA	VVD	SP	GL	D66	CU	LPF	Anders	NG
CDA	.8858	.0157	.0179	.0349	.0153	.0036	.0145	.0107	.0015	
PVDA		.9462		.0173	.0192					.0173
VVD			.8529		.0672	.08				
SP		.4086		.5914						
GL		.6553			.3447					
D66		.013	.4296		.0748	.4825				
CU	.3460						.5755		.0785	
LPF	.2273		.1540	.0329				.3037		.2821
Anders	.2370	.0146	.0117				.0534		.6832	
NG		.2518		.0291						.719

Quelle: (Nationaal Kiezersonderzoek, 2002/2003 und eigene Berechnungen des Autors)

Ganz offensichtlich in diesen Übergangsmatrizen ist der offensichtliche Verlust des LPF. Es ist auch sichtbar, dass der CDA stabil geblieben ist und der PvdA viel Unterstützung zurückgewonnen hat. Aufgrund der kurzen Zeitspanne zwischen den Wahlen scheint es noch mehr Steher zu geben als normalerweise, und dies zeigt sich an den hohen Werten auf der

Diagonale in der NKO-Matrix, aber auch in den modellbasierten Matrizen. Es ist klar, dass in allen Matrizen die lila Parteien, die bei den vorherigen Wahlen so stark verloren haben, einen Teil ihrer Unterstützung wiedererlangt haben, dass sie gegen die LPF verloren hatten. Auch hier sind beide Kombinationsmatrizen gleich, der Grund für diesen Effekt ist im vorherigen Absatz angegeben.

2003-2006

Für 2003-2006 sind die Übergangsmatrizen von NKO und LCA bereits in Kapitel 4 vertreten.

Tabelle 24. Übergangsmatrix LRmodel 2003-2006 (diagonale Elemente fett gedruckt)

2003	2006									
	CDA	PVDA	VVD	SP	GL	D66	CU	PVV	Anders	NG
CDA	.8430	-.0318	-.0022	.0765	-.0162	-.0075	.0172	.0564	-.0172	.0027
PVDA	-.0748	.7416	-.0224	.2252	-.0044	-.0184	.0109	-.0164	.0052	.0128
VVD	.2548	.0850	.8329	.2721	-.0945	-.0447	.0368	.1887	-.0164	.1924
SP	.1417	-.2749	-.1320	1,6291	-.1073	-.0930	-.0578	-.0123	-.0922	.0544
GL	.3436	.2122	.2608	.0798	.9923	.2024	.0934	.1932	.0411	.3042
D66	-.2413	-.1210	.0920	-1,361	.5959	.4550	-.0819	-.7486	.2387	-.4962
CU	.0849	.1137	-.0278	-.1720	-.0847	.0257	1,3768	-.1866	.0361	.2518
LPF	.2513	-.1456	-.0593	-.0533	-.0825	.0699	.0198	.2922	.2566	.1648
Anders	-.0549	-.0465	.0181	-.1415	.0910	-.0069	.0861	-.0259	1,0380	-.1088
NG	-.1050	.1060	.0050	-.0085	.0146	.0318	-.0158	.0946	.0098	.8106

Quelle: (Eigene Berechnungen des Autors)

Tabelle 25. Übergangsmatrix QPmodel 2003-2006 (diagonale Elemente fett gedruckt)

2003	2006									
	CDA	PVDA	VVD	SP	GL	D66	CU	PVV	Anders	NG
CDA	.8396		.0014	.0318			.0316	.0392	.0016	.0547
PVDA		.7627		.1822			.0122			.0429
VVD	.1496		.8218						.0286	
SP				1.0000						
GL		.0439		.1366	.5604					.2591
D66					.3551	.4029	.0764		.1656	
CU							1.0000			
LPF								.5608	.1138	.3254
Anders							.0993		.9007	
NG				.1811				.0761		.7428

Quelle: (Eigene Berechnungen des Autors)

Tabelle 26. Übergangsmatrix IPF-Modell 2003-2006 (diagonale Elemente fett gedruckt)

2003	2006									
	CDA	PVDA	VVD	SP	GL	D66	CU	PVV	Anders	NG
CDA	.2668	.1511	.1227	.1237	.0281	.0113	.0368	.0480	.0407	.1708
PVDA	.2477	.1635	.1229	.1285	.0300	.0120	.0352	.0472	.0384	.1745
VVD	.2549	.1545	.1316	.1238	.0302	.0125	.0340	.0477	.0398	.1709
SP	.2515	.1590	.1228	.1308	.0300	.0121	.0327	.0480	.0368	.1762
GL	.2490	.1599	.1268	.1275	.0319	.0129	.0340	.0471	.0375	.1733
D66	.2509	.1572	.1314	.1252	.0317	.0131	.0336	.0472	.0385	.1714
CU	.2626	.1534	.1155	.1139	.0273	.0108	.0555	.0430	.0541	.1637
LPF	.2534	.1545	.1267	.1250	.0290	.0119	.0346	.0494	.0411	.1743
Anders	.2614	.1454	.1185	.1120	.0256	.0105	.0502	.0465	.0657	.1642
NG	.2507	.1582	.1221	.1285	.0292	.0118	.0338	.0492	.0385	.1778

Quelle: (Eigene Berechnungen des Autors)

Tabelle 27. Übergangsmatrix Kombination 1 Modell 2003-2006 (diagonale Elemente fett gedruckt)

2003	2006									
	CDA	PVDA	VVD	SP	GL	D66	CU	PVV	Anders	NG
CDA	.8267		.0176	.0346			.0223	.0421	.0072	.0495
PVDA		.7072		.2112	.0019		.0182			.0615
VVD	.1678		.6680		.0175	.0427	.0141		.0899	
SP				.8154						.1846
GL		.3809		.0058	.5765	.0369				
D66			.628		.2327	.1393				
CU							.8374		.1626	
LPF								.4749	.0387	.4864
Anders	.0338		.0116				.2668		.6877	
NG				.2161				.093		.6909

Quelle: (Aarts *et al.* 2007: S. 224 und eigene Berechnungen der Autoren)

Tabelle 28. Übergangsmatrix Kombination 2 Modell 2003-2006 (diagonale Elemente fett gedruckt)

2003	2006									
	CDA	PVDA	VVD	SP	GL	D66	CU	PVV	Anders	NG
CDA	.8380		.0008	.0295			.0328	.0404	.0006	.058
PVDA		.7529		.1965			.0134			.0372
VVD	.1505		.7825						.0669	
SP		.0245		.9755						
GL				.0178	.5978	.0164	.0054			.3626
D66		.0811	.0811	.0579	.312	.3841	.0588		.025	
CU							1.0000			
LPF			.0867	.0514				.5333	.0903	.2383
Anders	.0115			.0115			.0896		.8873	
NG				.1729				.0811		.7460

Quelle: (Aarts *et al.* 2007: S. 224 und eigene Berechnungen der Autoren)

Die lineare Regressionsübergangsmatrix zeigt viele Ergebnisse unter 0 und über 1. Der Wert dieser Matrix ist daher aufgrund der unlogischen Übergänge nicht sehr hoch.

Die Überprüfung der LPM-Übergangsmatrix zeigt, dass die meisten Wähler Steher sind. Insbesondere SP, CDA, PvdA, CU, VDD und GL weisen hohe Steherraten auf. Jemand, der 2003 oft nicht gewählt hat, war 2006 auch kein Wähler (74%). Aber Sie können immer noch sehen, dass 18% der Nichtwähler im Jahr 2003 für die SP im Jahr 2006 gestimmt haben, was teilweise das Wachstum der SP erklärt. Die Wählerschaft von D'66 im Jahr 2003 war auf mindestens vier Parteien verteilt. Nach dem LPM-Modell stimmten 2006 über 56% der früheren LPF-Wähler für den PVV, und ein großer Teil der ehemaligen LPF-Wähler blieb zu Hause. SP und CU haben Werte von 1 in ihrer Diagonale, was darauf hinweist, dass sie eine sehr treue Unterstützung haben.

Aufgrund der starken Einschränkungen des IPF-Modells wird eine ganz andere Übergangsmatrix dargestellt. Die Grenzergebnisse müssen bei diesem Ansatz gleich sein, was eine sehr strenge Einschränkung darstellt. Dieses Modell verwendet die Anfangsmatrix als Startmatrix. Es ist interessant, die niedrigen Werte auf der Diagonale zu beobachten. Man kann sehen, dass diese niedrigen Werte auf der Diagonale nicht realistisch sind, was zu einer mittelmäßigen Anpassung führt.

Wie bei der ersten Kombinationsmatrix gezeigt, enthält diese Kombinationsmatrix viele Nullen. Die Werte von 1 sind aufgrund der Einschränkung des Konfidenzintervalls verschwunden. Aufgrund der anderen Einschränkung der auf 1 summierenden Zeilen sind einige andere Nullen in der Matrix verschwunden. Die zweite Kombinationsmatrix enthält weniger Nullen als die erste Kombinationsmatrix, enthält jedoch einige Nullwerte, die nicht wahrscheinlich erscheinen. Es ist ziemlich logisch, dass die Werte in dieser Matrix immer noch so nahe wie möglich an der optimalen Lösung berechnet werden, wobei wiederum Nullen erzeugt werden. Es werden jedoch keine Übergänge zwischen PvdA und CDA erkannt, was überhaupt nicht wahrscheinlich erscheint.

7.3 Vergleich aller Modelle pro Wahljahr

1998-2002



Abbildung 11a. 100% Steher Modell 98-02

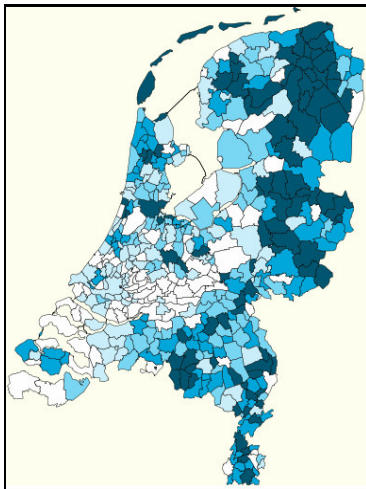


Abbildung 11b. NKO-Modell 98-02

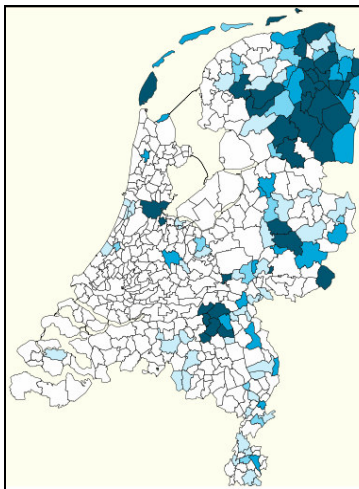


Abbildung 11c. LPM-Modell 98-02

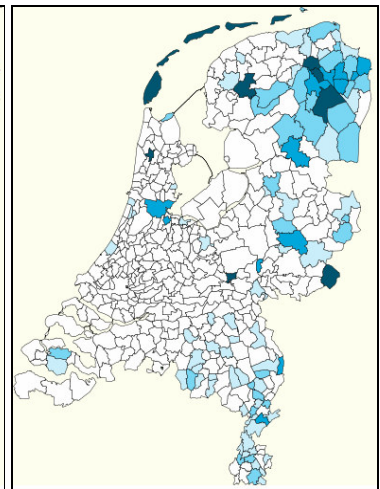


Abbildung 11d. Ökobilanzmodell 98-02

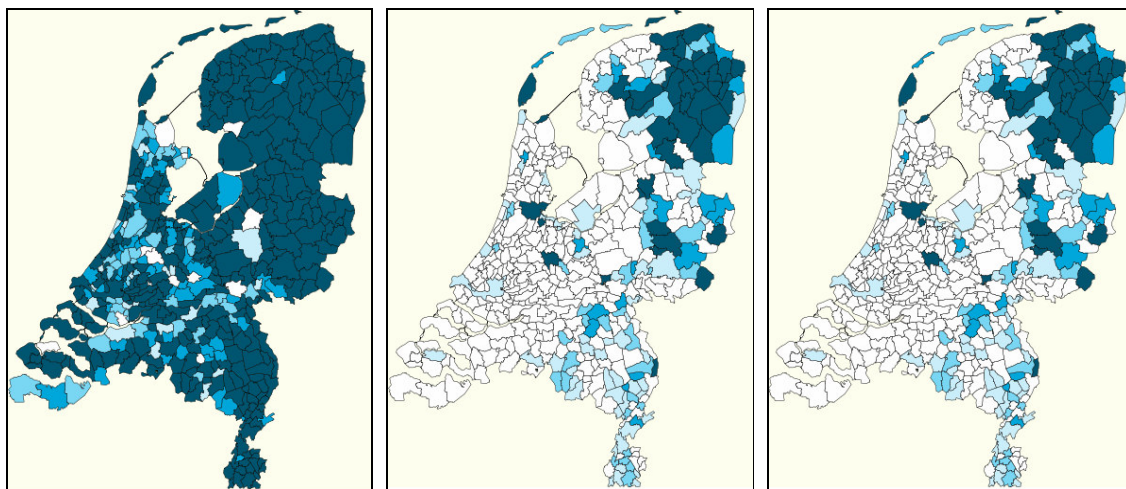







Abbildung 11e. IPF-Modell 98-02

Abbildung 11f. Kombimodell 98-02

Abbildung 11g. Combo2 Modell 98-02

Legende 1998-2002:

	$R^2 < 0.9598$		$R^2 < 0.9702$		$R^2 < 0.9778$		$R^2 < 0.9839$		$R^2 < 1$
---	----------------	---	----------------	---	----------------	--	----------------	---	-----------

Diese Bilder, basierend auf der Perzentilklassifizierung des NKO 2002, zeigen, dass die Passform der Modelle sehr hoch ist. Die meisten Gemeinden im LCA-Modell haben einen McF. Pseudo R^2 zwischen 0,9839 und 1. Die Passform des 100% Steher-Modells ist schlechter als die des Unabhängigkeitsmodells. Alle Modelle scheinen wieder Probleme mit den Inseln zu haben. Auch die Regionen / Provinzen Groningen und Overijssel weisen eine schlechte Passform auf. Da die allgemeine Anpassung der Modelle so hoch ist, ist es schwierig, eine statistisch signifikante Erklärung zu finden. Die einzige sichtbare Erklärung in den Ergebnissen ist, dass der VVD in diesen Teilen der Niederlande leicht unterschätzt wird. Wenn man sich die Vorzugsstimmen für einige wichtige Kandidaten ansieht, führt dies zu Spitzenwerten bei der Anzahl der Stimmen für eine bestimmte Partei in mehreren Gemeinden. Dies führt in den meisten dieser Gemeinden zu einer etwas geringeren Passform. Wir können Außenseiter für die SP in Oss, Außenseiter für die LPF in Rotterdam und Außenseiter für die D'66 in Leiden finden. Diese Analogie gilt auch für andere Gemeinden.

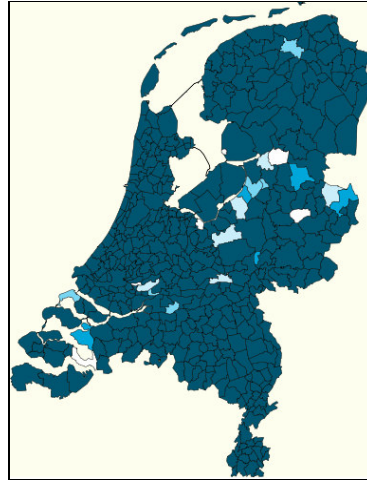


Abbildung 12a. 100% Steher Modell 02-03

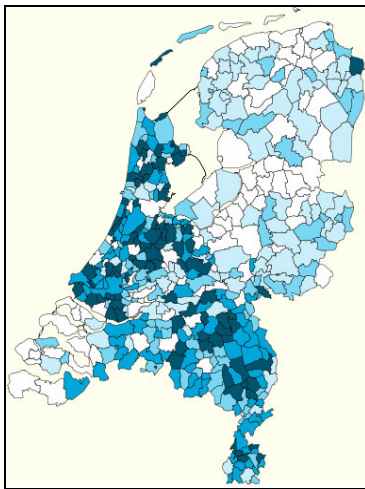


Abbildung 12b. NKO Modell 02-03

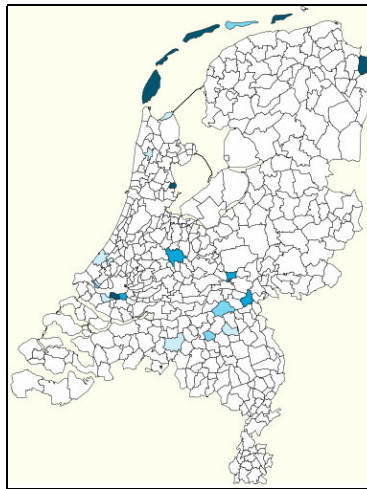


Abbildung 12c. LPM-Modell 02-03

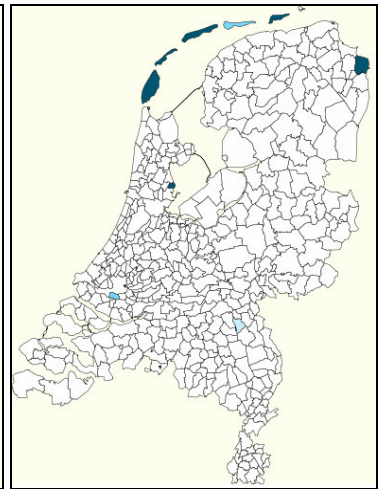


Abbildung 12d. Ökobilanzmodell 02-03

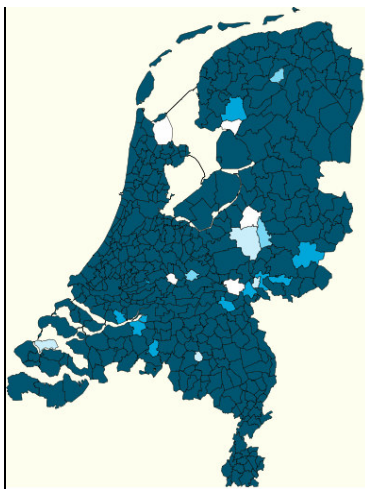


Abbildung 12e. IPF-Modell 02-03

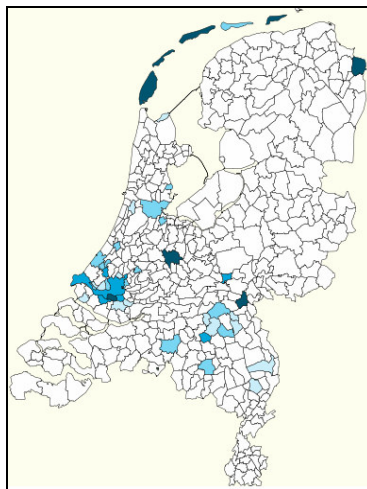


Abbildung 12f. Kombimodell 02-03

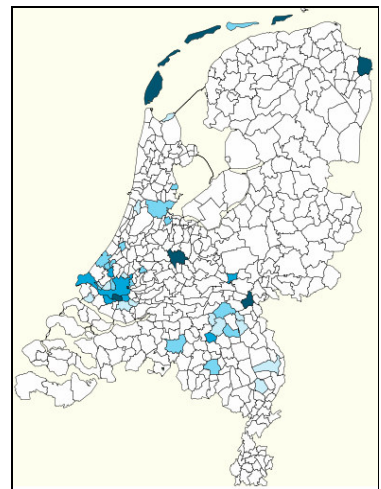
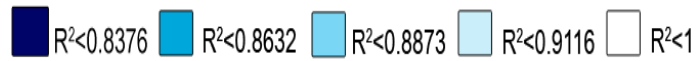


Abbildung 12g. Combo2 Modell 02-03

Legende 2002-2003:



Diese Bilder zeigen wenig Unterscheidungskraft zwischen den Modellanpassungen, die Schätzungen aller Modelle liegen nahe beieinander. Die Wahlen waren nur 17 Monate voneinander entfernt. Die meisten Wähler sind bei ihren Parteien geblieben, was zu einem großen Anteil von Stehern geführt hat (siehe die Übergangsmatrizen in Anhang A Abschnitt 7.2). Auch hier sind die Schätzungen für die Inseln nicht gut. Interessanterweise ist die Gemeinde Reiderland im Norden von Groningen etwas schlechter. Da die New Communist Party (NCPN) 2002 nicht teilnahm, stimmte ein relativ großer Teil der Wähler für die LPF. Im Jahr 2003 wechselten sie jedoch zu einem kleinen Teil zurück zum NCPN und zu anderen Parteien, was zu einer etwas schlechteren Passform führte.

2003-2006

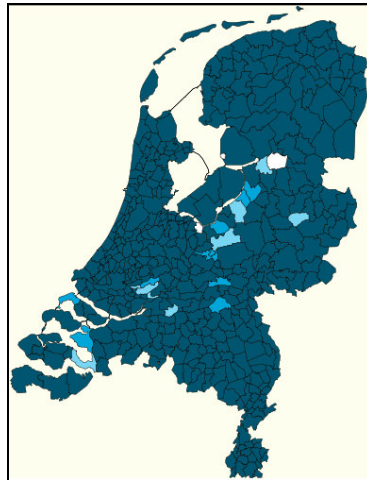


Abbildung 13a. 100% Steher Modell

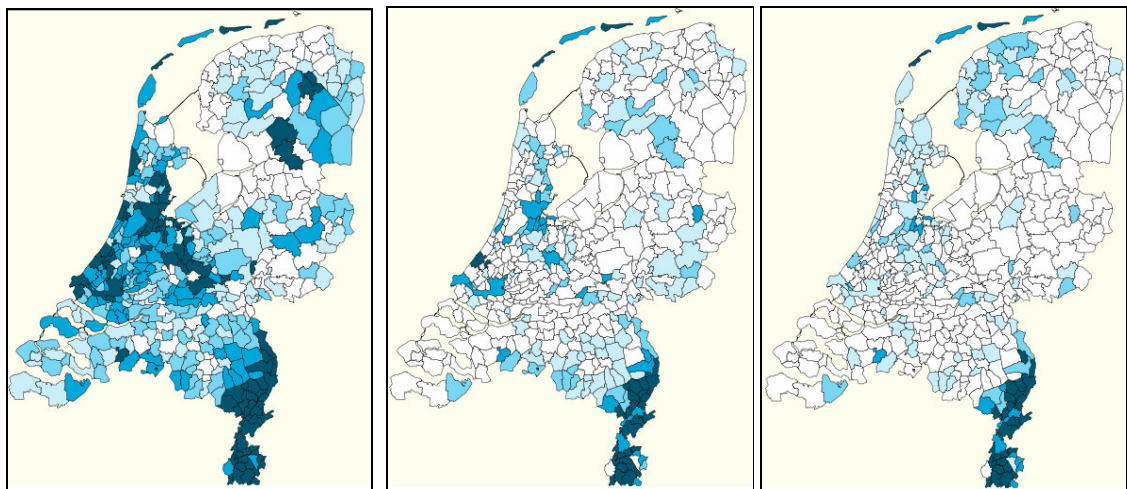


Abbildung 13b. NKO-Modell 03-06 Abbildung 13c. LPM-Modell 03-06 Abbildung 13d. Ökobilanzmodell 03-06

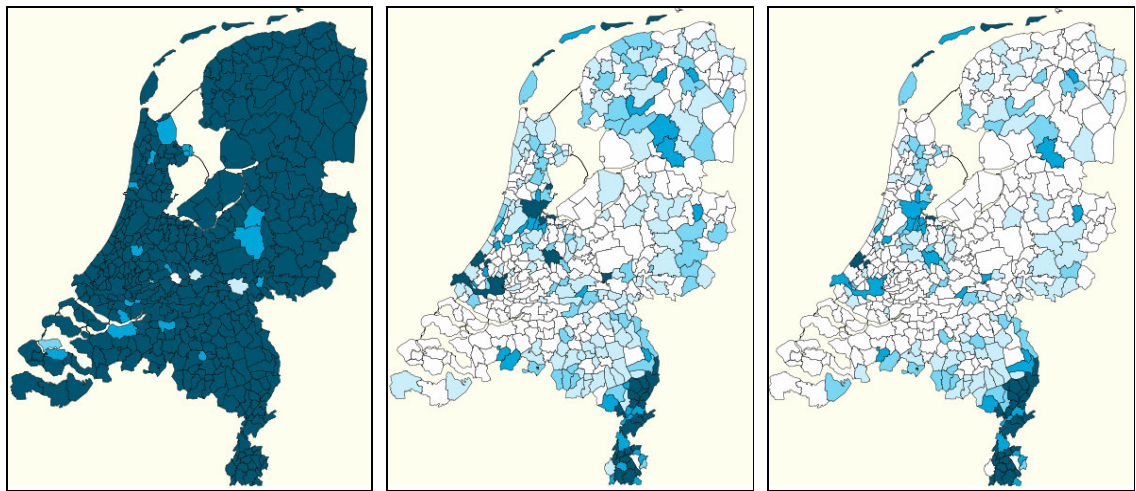
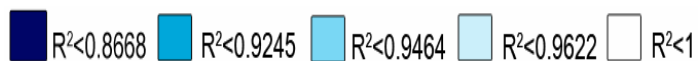


Abbildung 13e. IPF-Modell 03-06

Abbildung 13f. Kombimodell 03-06

Abbildung 13g. Combo2 Modell 03-06

Legende 2003-2006:



Hier können wir sehen, dass sich bei der Analyse aller Modelle nicht viel ändert
2006, bereits in Kapitel 4 vorgestellt.

7.4 Erweiterungen des Mover-Stayer-Modells

Es gibt verschiedene Erweiterungen des Mover-Stayer-Modells. Die Hawkes-Mehrfachübergangsmatrix verwendet eine lineare Regression. Diese Methode ermöglicht es, Wahlkreise in kleinere homogene Einheiten aufzuteilen. Dieses einfache Modell übertrifft das ursprüngliche Mover-Stayer-Modell (Hawkes, 1969).

Eine weitere Erweiterung des Mover-Stayer-Modells ist das Speichermodell von Upton, 1977. Anstatt bei jeder Wahl gleiche Wahrscheinlichkeiten zu haben, führt er Konstanten ein

$ein_{k-1}, ein_{k-2}, \dots, ein_{k-k}$, die mit der Wahl der einzelnen Partei verbunden sind ein in dem

Bisherige / Wahlen. Mit $x_{ij} = 1$ oder 0 die Wahrscheinlichkeit der Abstimmung durch eine Person für

Party j in dem k Die Wahl ist gegeben durch $(\sum_{i=1}^k x_{ij} A_i)$. Die Anziehungskraft eines

Einzelperson bei der Wahl k ohne vorherige Abstimmungshistorie für die Partei j wird dargestellt als

p_j nimmt dann eine Einzelperson auf / eigene bisherige Abstimmungsgeschichte in die

Modell. Seine Forschung liefert keine endgültigen Ergebnisse, aber motivierende Anreize für weitere Studien (Upton, 1977, S. 87).

Das verallgemeinerte Mover-Stayer-Modell von Cook *et al.* ((2002) beschreibt eine Erweiterung des latenten

Mover-Stayer-Modells. Sie definieren eine Mischung aus verschachtelten zeitkontinuierlichen Markov-Prozessen. Anstatt die

Steyer in ihrem anfänglichen stationären Zustand zu halten, kann jedes Individuum einen oder mehrere absorbierende

Zustände haben, aber wenn dieser Zustand eingegeben wird, finden keine Übergänge aus diesem Zustand mehr statt. Vor

diesem Endzustand ist es jedoch

Eingetragen können Steher Übergänge in eine Reihe anderer Staaten vornehmen (Cook *et al.*, 2002, S. 2).

Insbesondere die in den Wahldaten festgestellte Typheterogenität kann auch besser berücksichtigt werden. Da es Wähler gibt, die sehr lange im selben Staat bleiben, ist der ursprüngliche Mover-Stayer geeignet. Zum Beispiel sind die Wähler in ihrem Wahlverhalten relativ stabil (Egmond *et al.*, 1998). Sie können jedoch einige Parteientscheidungen treffen, bevor sie sich für ihre endgültige Entscheidung entscheiden. Dies kann unter Verwendung der verschiedenen Absorptionszustände modelliert werden und kann daher ein geeigneteres Modell sein als das ursprüngliche Mover-Stayer-Modell (Cook *et al.*, 2002, S. 4).

7.5 Verwendete Programme und Code

Die in diesem Projekt verwendeten Programme sind vielfältig. Für die Zusammenstellung der Datensätze und die NKO-Analyse wurde das Statistikprogramm SPSS verwendet.

Für die Berechnung der Latent Class Analysis wurde das von F. van de Pol geschriebene Programm Panmark verwendet.

Für die vergleichende Analyse der Schätzmethoden wurden die Programme R und Matlab verwendet. Der Code für das kostenlose Statistikprogramm R ist unten angegeben.

Code R:

```
getwd ()
kamer0306n <- as.matrix (read.table ("kamer0306n.dat"))
# kamerverkiezingen 2003 en 2006
kamer0203n <- as.matrix (read.table ("kamer0203n.dat"))
# kamerverkiezingen 2002 en 2003
kamer9802n <- as.matrix (read.table ("kamer9802n.dat"))
# kamerverkiezingen 1998 en 2002
NKO2006 <- as.matrix (read.table ("NKO2006.dat")) #NKO transitie matrix NKO2003 <- as.matrix (read.table
("NKO2003.dat")) #NKO transitie matrix NKO2002 <- as.matrix (read .table ("NKO2002.dat")) #NKO transitie matrix
frank2006 <- as.matrix (read.table ("frank2006.dat")) # franks matrix frank2003 <- as.matrix (read.table ("frank2003.dat")
)) # franks matrix frank2002 <- as.matrix (read.table ("frank2002.dat")) # franks matrix
```

```
# combo2 <- as.matrix (read.table ("combo.dat")) # Combo Matrix Stayer <- as.matrix (read.table ("stayer.dat")) #
Stayer Matrix Pseudo4 <- as.matrix ( read.table ("PSEUDO4.dat")) # Pseudo-R LPM en gemcodes
```

```
eps <- .Machine $ double.eps
```

```
Code <- Pseudo4 [, 1] #Gemeentecodes Pseudo4 <-
Pseudo4 [, 2] # Pseudo-R
```

```
gem <- kamer0306n [, 1] #Gemeentes genummerd 1-459 gemcode <- kamer0306n [,
2] #Gemeentecodes
t <- kamer0306n [, 3] #verkiezingen genummerd 1,2 p <- kamer0306n [, 4] #Partijen
genummerd 1-10 f <- kamer0306n [, 5] #Frequentie
```

```
# gem <- kamer0203n [, 1] #Gemeentes genummerd 1-459
# gemcode <- kamer0203n [, 2] #Gemeentecodes
# t <- kamer0203n [, 3] #verkiezingen genummerd 1,2
```

```

# p <- kamer0203n [, 4] #Partijen genummerd 1-10
# f <- kamer0203n [, 5] #Frequentie

# gem <- kamer9802n [, 1] #Gemeentes genummerd 1-459
# gemcode <- kamer9802n [, 2] #Gemeentecodes
# t <- kamer9802n [, 3] #verkiezingen genummerd 1,2
# p <- kamer9802n [, 4] #Partijen genummerd 1-10
# f <- kamer9802n [, 5] #Frequentie

p.kij <- Array (0, c (max (gem), max (p), max (p)))
P.kij <- Array (0, c (max (Edelstein), max (p), max (p)))
f.ki <- Matrix (0, max (gem), max (p))
f.kj <- Matrix (0, max (gem), max (p))
f.ij <- Matrix (0, max (p), max (p))
fk <- Array (0, max (Edelstein)) f1.k <- Matrix (0,
max (Edelstein), 1) f2.k <- Matrix (0, max
(Edelstein), 1) p2.k <- Matrix (0, max (Edelstein), 1)

fi <- Array (0, max (p))
fj <- Array (0, max (p))

p.ki <- Matrix (0, max (gem), max (p))
p.kj <- Matrix (0, max (gem), max (p))
p.ij <- Matrix (0, max (p), max (p)) Inip.ij <- Matrix (0, max (p),
max (p))
P.ki <- Matrix (0, max (Edelstein), max (p))
P.kj <- Matrix (0, max (gem), max (p))
P.ij <- Matrix (0, max (p), max (p))
pk <- Matrix (0, max (Edelstein), 1)
pi <- Matrix (0, max (p), 1)
pj <- Matrix (0, max (p), 1)

für (x in 1: Länge (Edelstein)) {
    k <- gem [x]
    i <- p [x]
    s <- t [x]
    wenn (s <2) f.ki [k, i] <- f [x]
    if (s > 1) f.kj [k, i] <- f [x] #Marginalen gemeenten x
verkiezing
für (i in 1: max (p)) {
    für (k in 1: max (gem)) {
        fi [i] <- fi [i] + f.ki [k, i]
        f1.k [k] <- f1.k [k] + f.ki [k, i]] #Marginalen
    }
}
voor eerste verkiezingen
pi <- fi / sum (fi) # Marginale kansen voor eerste verkiezingen

für (j in 1: max (p)) {
    für (k in 1: max (gem)) {
        fj [j] <- fj [j] + f.kj [k, j]
        f2.k [k] <- f2.k [k] + f.kj [k, j]] #Marginalen
    }
}
voor tweede verkiezingen
pj <- fj / sum (fj) # Marginale kansen für Tweede verkiezingen p2.k <- f2.k / sum (f2.k) # Marginale kansen für
gemeentes bij tweede verkiezingen

für (k in 1: max (gem)) {
    für (i in 1: max (p)) {
        p.ki [k, i] <- f.ki [k, i] /f1.k [k]]
    }
}
für (k in 1: max (gem)) {
    für (j in 1: max (p)) {
        p.kj [k, j] <- f.kj [k, j] /f2.k [k]]
    }
}

für (i in 1: max (p)) {
    für (j in 1: max (p)) {

```

```

Inip.ij [i, j] <- pi [ij]] #De initiële TRANSITIEMATRIX !!

für (k in 1: max (gem)) {
  für (i in 1: max (p)) {
    für (j in 1: max (p)) {
      P.kij [k, i, j] <- p2.k [k] * pi [i] * pj [j]]}}

für (i in 1: max (p)) {
  für (j in 1: max (p)) {
    für (k in 1: max (gem)) {
      P.ij [i, j] <- P.ij [i, j] + P.kij [k, i, j]]}}

#####
# Test traf onafhankelijkheid
#####
phat1.kj <- p.ki% %% Inip.ij
fhat1.kj <- Matrix (0, max (gem), max (p))

pSums.k <- rowSums (phat1.kj)
für (k in 1: max (gem)) {
  für (j in 1: max (p)) {
    fhat1.kj [k, j] <- phat1.kj [k, j] * f2.k [k] /pSums.k [k]]}
LRX2.1 <- 2 * sum (f.kj * log (f.kj / fhat1.kj)) Inip.ij

#####
# Testen Sie alleen maar Steher
#####
phat2.kj <- p.ki% %% Steher
fhat2.kj <- Matrix (0, max (gem), max (p))

pSums.k <- rowSums (phat2.kj)
für (k in 1: max (gem)) {
  für (j in 1: max (p)) {
    fhat2.kj [k, j] <- phat2.kj [k, j] * f2.k [k] /pSums.k [k]]}
LRX2.2 <- 2 * Summe (f.kj * log (f.kj / fhat2.kj)) Steher

#####
# Test traf NKO
#####
phat3.kj <- p.ki% %% NKO2006
fhat3.kj <- Matrix (0, max (gem), max (p))

pSums.k <- rowSums (phat3.kj)
für (k in 1: max (gem)) {
  für (j in 1: max (p)) {
    fhat3.kj [k, j] <- phat3.kj [k, j] * f2.k [k] /pSums.k [k]]}
LRX2.3 <- 2 * Summe (f.kj * log (f.kj / fhat3.kj)) NKO2006

#####
# Regressie op frequenties (Keller & Ten Cate, 1977) Lineaire regressie
#####
B <- lösen (t (f.ki)% %% f.ki)% %% t (f.ki)% %% f.kj phat4.kj <- p.ki% %% B.

fhat4.kj <- Matrix (0, max (gem), max (p))

pSums.k <- rowSums (phat4.kj)
für (k in 1: max (gem)) {
  für (j in 1: max (p)) {
    fhat4.kj [k, j] <- phat4.kj [k, j] * f2.k [k] /pSums.k [k]]}
LRX2.4 <- 2 * Summe (f.kj * log (f.kj / fhat4.kj))

```

B.

```
#####
# Regressie op frequenties (Keller & Ten Cate, 1977) Kwadratisch Programmeren

#####
# http://cran.r-project.org/web/packages/quadprog/index.html
# hier functie alle waarden corrigeren zodat de rijen optellen tot 1
# Quadratisch programmeren oplossing met voorwaarden kansverdeling library (quadprog)

# Rijen en kolomrestricties tegelijk
#####
meq <- max (p)
XtX <- t (p.ki)% %% p.ki.
Dmat <- Matrix (0, max (p) ≤ 2, max (p) ≤ 2) dvec <- Matrix (0,
max (p) ≤ 2, 1)
Amat <- Matrix (0, max (p)~ 2, max (p)) für (i in 1: max (p))

{
  Xty <- t (p.ki)% %% p.kj [, i]
  Bereich <- ((i-1) * max (p) +1) :( i * max (p))
  Dmat [Bereich, Bereich] <- XtX% %% XtX
  dvec [Bereich] <- XtX% %% Xty #te minimaliseren Vektor
}
Amat <- cbind (
  t (Matrix (diag (max (p)), max (p), max (p) ≤ 2)),

  diag (max (p)~ 2)
)
bvec <- rbind (
  Matrix (1, max (p), 1),
  Matrix (0, max (p) ^ 2, 1)
)
sol <- solve.QP (Dmat, dvec, Amat, bvec, meq) LPM <- Matrix (sol $
Lösung, max (p), max (p))

phat5.kj <- p.ki% %% LPM
fhat5.kj <- Matrix (0, max (gem), max (p))

pSums.k <- rowSums (phat5.kj)
für (k in 1: max (gem)) {
  für (j in 1: max (p)) {
    fhat5.kj [k, j] <- phat5.kj [k, j] * f2.k [k] /pSums.k [k]}
LRX2.5 <- 2 * Summe (f.kj * log (f.kj / fhat5.kj))
print (formatC (abs (LPM), dig = 4, format = "f"), quote = FALSE)

#####
# Test erfüllt Ökobilanz
#####
phat6.kj <- p.ki% %% frank2006
fhat6.kj <- Matrix (0, max (gem), max (p))

pSums.k <- rowSums (phat6.kj)
für (k in 1: max (gem)) {
  für (j in 1: max (p)) {
    fhat6.kj [k, j] <- phat6.kj [k, j] * f2.k [k] /pSums.k [k]}
LRX2.6 <- 2 * sum (f.kj * log (f.kj / fhat6.kj)) frank2006

#####
# IPF
```

```
#####
iter <- 0
P.kij.size.k <- dim (P.kij) [1]
P.kij.size.i <- dim (P.kij) [2]
P.kij.size.j <- dim (P.kij) [3]

während (abs (p.ij [1,1] - P.ij [1,1])> 10E-6) {
  iter <- iter +1
  p.ij <- P.ij

  # Quasi-marginalen bijstellen
  P.ki <- anwenden (P.kij, c (1,2), Summe)

  # Sjtapf einzahlen
  P.kij <- P.kij * ((p.ki/P.ki)% o% Array (1, P.kij.size.j))

  # Quasi-marginalen bijstellen
  P.kj <- anwenden (P.kij, c (1,3), Summe)

  # Sjtapf zwei
  P.kij <- P.kij * aperm ((p.kj / P.kj)% o% Array (1, P.kij.size.i),
c (1,3,2))

  # Quasi-marginalen bijstellen
  P.ij <- anwenden (P.kij, c (2,3), Summe)

  # Sjtapf drei
  P.kij <- P.kij * (Array (1, P.kij.size.k)% o% (p.ij / P.ij))

} #End while

Pi <- Matrix (0, max (p), 1) für (i in 1: max (p))
{
  für (j in 1: max (p)) {
    Pi [i] <- Pi [i] + P.ij [i, j]}
für (i in 1: max (p)) {
  für (j in 1: max (p)) {
    P.ij [i, j] <- P.ij [i, j] / Pi [i]}

phat7.kj <- p.ki% %% P.ij.
fhat7.kj <- Matrix (0, max (gem), max (p))

pSums.k <- rowSums (phat7.kj)
für (k in 1: max (gem)) {
  für (j in 1: max (p)) {
    fhat7.kj [k, j] <- phat7.kj [k, j] * f2.k [k] /pSums.k [k]}

LRX2.7 <- 2 * Summe (f.kj * log (f.kj / fhat7.kj))
P.ij.

#####
# COMBO K & T und NKO
#####
# http://cran.r-project.org/web/packages/quadprog/index.html
# hier functie alle waarden corrigeren zodat de rijen optellen tot 1
# Quadratisch programmeren oplossing met voorwaarden kansverdeling library (quadprog)

# bereking BTI
#####
SD <- sd (NKO2006, na.rm = FALSE) / sqrt (459) #Sx
#  $S_x s / \sqrt{n}$  s =  $\sqrt{1 / N - 1} \sum_{i=1}^N (x_i - x_{\text{streep}})^2$  LCFI <- Mittelwert (phat3.kj) -2 *
SD # Stichprobenmittelwert - / + 2 * Sx
```

```

HCFI <- Mittelwert (phat3.kj) + 2 * SD
# marge <- 0,025 # 10%
Diagmin <- Matrix (Diag (NKO2006) - LCFI, max (p), 1) Diagmax <- Matrix (Diag
(NKO2006) + HCFI, Max (p), 1)

# nieuwe berekening
meq <- max (p)
XtX <- t (p.ki)% **% p.ki.
DiagHelper <- Matrix (0, max (p) ≤ 2, max (p)) Dmat <- Matrix (0, max (p)
≤ 2, max (p) ≤ 2) dvec <- Matrix (0, max (p) 2, 1)

für (i in 1: max (p))
{
  DiagHelper [(i - 1) * max (p) + i, i] <- t (p.ki)% **% p.kj [, i]

  Bereich <- ((i-1) * max (p) + 1) : (i * max (p))
  Dmat [Bereich, Bereich] <- XtX% **% XtX
  dvec [Bereich] <- XtX% **% Xty #te minimaliseren Vektor
}}
Amat <- cbind (
  t (Matrix (diag (max (p)), max (p), max (p)2)), DiagHelper,

  - DiagHelper,
  diag (max (p)2)
)
bvec <- rbind (
  Matrix (1, max (p), 1),
  diagmin,
  - Diagmax,
  Matrix (0, max (p) ^ 2, 1)
)
Sol <- solve.QP (Dmat, dvec, Amat, bvec, meq) Combo <- Matrix (Sol $ Lösung,
max (p), max (p))

phat8.kj <- p.ki% **% combo
fhat8.kj <- Matrix (0, max (gem), max (p))

pSums.k <- rowSums (phat8.kj)
für (k in 1: max (gem)) {
  für (j in 1: max (p)) {
    fhat8.kj [k, j] <- phat8.kj [k, j] * f2.k [k] / pSums.k [k]}
LRX2.8 <- 2 * Summe (f.kj * log (f.kj / fhat8.kj))
print (formatC (abs (Combo), dig = 4, format = "f"), quote = FALSE)

#####
# COMBO 2 0 waarden vervangen traf NKO-waarden
#####
SD <- sd (NKO2006, na.rm = FALSE) / sqrt (459) #Sx
# Sx s / sqrt (n) s = sqrt 1 / N-1 sum_i = 1 N (x_i-xstreep)2 2 LCFI <- Mittelwert (phat3.kj) -2 *
SD # Stichprobenmittelwert - / + 2 * Sx HCFI <- Mittelwert (phat3.kj) + 2 * SD

Marge <- 0,025 # 10%
Grenswaarde <- 1e-4
BTIMIN <- (LPM < grenswaarde) * (NKO2006-LCFI) - (LPM > = grenswaarde) * (1E99) BTIMIN <- Matrix (BTIMIN, max
(p) ^ 2, 1)
BTIMAX <- (LPM < grenswaarde) * (NKO2006 + HCFI) + (LPM > = grenswaarde) * (1E99) BTIMAX <- Matrix (BTIMAX,
max (p) ^ 2, 1)

# nieuwe berekening
meq <- max (p)
XtX <- t (p.ki)% **% p.ki.
Dmat <- Matrix (0, max (p) ≤ 2, max (p) ≤ 2) dvec <- Matrix (0,
max (p) ≤ 2, 1)

```

```

für (i in 1: max (p))
{
  Xty <- t (p.ki)% **% p.kj [, i]
  Bereich <- ((i-1) * max (p) +1) :( i * max (p))
  Dmat [Bereich, Bereich] <- XtX% **% XtX
  dvec [Bereich] <- XtX% **% Xty #te minimalisieren Vektor
}}
Amat <- cbind (
  t (Matrix (diag (max (p)), max (p), max (p) ≤ 2)), diag (max (p) ≤ 2),

  - diag (max (p) ^ 2),
  diag (max (p) ^ 2)
)
bvec <- rbind (
  Matrix (1, max (p), 1),
  BTIMIN,
  - BTIMAX,
  Matrix (0, max (p) ^ 2, 1)
)
sol <-solve.QP (Dmat, dvec, Amat, bvec, meq) combo2 <- Matrix (sol $ Lösung,
max (p), max (p))

phat9.kj <- p.ki% **% combo2
fhat9.kj <- Matrix (0, max (gem), max (p))

pSums.k <- rowSums (phat9.kj)
für (k in 1: max (gem)) {
  für (j in 1: max (p)) {
    fhat9.kj [k, j] <- phat9.kj [k, j] * f2.k [k] /pSums.k [k]}
LRX2.9 <- 2 * Summe (f.kj * log (f.kj / fhat9.kj))
print (formatC (abs (combo2), dig = 4, format = "f"), quote = FALSE)

#####
# CHI-QUADRAT-TEST PRO GEMEINDE
#####
res1.kj <-Matrix (0, max (gem), max (p))
res2.kj <-Matrix (0, max (gem), max (p))
res3.kj <-Matrix (0, max (gem), max (p))
res4.kj <-Matrix (0, max (gem), max (p))
res5.kj <-Matrix (0, max (Edelstein), max (p))
res6.kj <-Matrix (0, max (gem), max (p))
res7.kj <-Matrix (0, max (gem), max (p))
res8.kj <-Matrix (0, max (gem), max (p))
res9.kj <-Matrix (0, max (Edelstein), max (p))
enen <- Matrix (1,10,1)
für (k in 1: max (gem)) {
  für (j in 1: max (p)) {
    res1.kj [k, j] <- ((fhat1.kj [k, j] -f.kj [k, j])^ 2) /fhat1.kj [k, j] res2.kj [k, j] <- ((fhat2.kj [k, j] -f.kj [k, j])^ 2)
    /fhat2.kj [k, j]
    res3.kj [k, j] <- ((fhat3.kj [k, j] -f.kj [k, j])^ 2) /fhat3.kj [k, j] res4.kj [k, j] <- ((fhat4.kj [k, j] -f.kj [k, j])^ 2)
    /fhat4.kj [k, j] res5.kj [k, j] <- ((fhat5.kj [k, j] -f.kj [k, j])^ 2) /fhat5.kj [k, j] res6.kj [k, j] <- ((fhat6.kj [k, j] -f.kj
    [k, j])^ 2) /fhat6.kj [k, j] res7.kj [k, j] <- ((fhat7.kj [k, j] -f.kj [k, j])^ 2) /fhat7.kj [k, j] res8.kj [k, j] <- ((fhat8.kj [k,
    j] -f.kj [k, j])^ 2) /fhat8.kj [k, j] res9.kj [k, j] <- ((fhat9.kj [k, j] -f.kj [k, j])^ 2) /fhat9.kj [k, j]

  }}
}}
res1.k <- res1.kj% **% enen
res2.k <- res2.kj% **% enen
res3.k <- res3.kj% **% enen
res4.k <- res4.kj% **% enen

```

```

res5.k <- res5.kj% %% enen
res6.k <- res6.kj% %% enen
res7.k <- res7.kj% %% enen
res8.k <- res8.kj% %% enen
res9.k <- res9.kj% %% enen

#####
# Pseudo R pro gemeente
#####
PR2.2.k <- 1-res2.k / res1.k.
PR2.3.k <- 1-res3.k / res1.k.
PR2.4.k <- 1-res4.k / res1.k.
PR2.5.k <- 1-res5.k / res1.k.
PR2.6.k <- 1-res6.k / res1.k.
PR2.7.k <- 1-res7.k / res1.k.
PR2.8.k <- 1-res8.k / res1.k.
PR2.9.k <- 1-res9.k / res1.k.

#####
# Schatting Zetelverdeling pro Modell
#####
sumtot <- colSums (f.kj)
sumkiesdeler <- sum (sumtot) 1: 9)
kiesdeler <- 9838683/150
sumfecht <- colSums (f.kj)
sumfONAF <- colSums (fhat1.kj)
sumfSTAYER <- colSums (fhat2.kj)
sumfNKO <- colSums (fhat3.kj)
sumfLPM <- colSums (fhat5.kj)
sumfLCA <- colSums (fhat6.kj)
sumfIPF <- colSums (fhat7.kj)
sumfCOMBO <- colSums (fhat8.kj)
sumfCOMBO2 <- colSums (fhat9.kj)

sumfONAF
sumfSTAYER
sumfNKO
sumfLPM
sumfLCA
sumfIPF
sumfCOMBO
sumfCOMBO2
sumfecht
# 2003-2006
Herverdeling <-Funktion (Fractioneel)
{
  direkter <-Boden (Fractioneel)
  rest <-sum (fraktional) -summe (direkt)
  # - voor aflopend
  klingelte <-Rang (-c (Bruchteil / (direkt + 1),
    Fraktional / (direkt + 2)))
  extra <-Matrix (0, Länge (Bruchteil), 2)
  extra [klingelte [1: Pause]] <- extra [klingelte [1: Pause]] + 1 extra <-rowSums
    (extra)
  direkt + extra
}}

#####
# STATISTIEKEN
#####
GFX2.1 <-sum (res1.k) #Onafhankelijkheid
GFX2.2 <-sum (res2.k) #STAYER

```


GFX2.3 <-sum (res3.k) #NKO
 GFX2.4 <-sum (res4.k) #LR
 GFX2.5 <-sum (res5.k) #LP
 GFX2.6 <-sum (res6.k) #LCA
 GFX2.7 <-sum (res7.k) #IPF
 GFX2.8 <-sum (res8.k) #COMBO
 GFX2.9 <-sum (res9.k) # COMBO2

PR2.2 <- 1-GFX2.2 / GFX2.1
 PR2.3 <- 1-GFX2.3 / GFX2.1
 PR2.4 <- 1-GFX2.4 / GFX2.1
 PR2.5 <- 1-GFX2.5 / GFX2.1
 PR2.6 <- 1-GFX2.6 / GFX2.1
 PR2.7 <- 1-GFX2.7 / GFX2.1
 PR2.8 <- 1-GFX2.8 / GFX2.1
 PR2.9 <- 1-GFX2.9 / GFX2.1

GFX2.1 #Onafhankelijkheid
 GFX2.2 #STAYER
 GFX2.3 #NKO
 GFX2.4 #LR
 GFX2.5 #LP
 GFX2.6 #LCA
 GFX2.7 #IPF
 GFX2.8 #COMBO
 GFX2.9 # COMBO2

LRX2.1 #Onafhankelijkheid
 LRX2.2 #STAYER
 LRX2.3 #NKO
 LRX2.4 #LR
 LRX2.5 #LP
 LRX2.6 #LC
 LRX2.7 #IPF
 LRX2.8 #COMBO
 LRX2.9 # COMBO2

PR2.2 #STAYER
 PR2.3 #NKO
 PR2.4 #LR
 PR2.5 #LP
 PR2.6 #LC
 PR2.7 #IPF
 PR2.8 #COMBO
 PR2.9 # COMBO2

welche (PR2.2.k [, 1]> 3) #STAYER
 welche (PR2.3.k [, 1]> 3) #NKO
 welche (PR2.4.k [, 1]> 3) #LR
 welche (PR2.5.k [, 1]> 3) #LP
 welche (PR2.6.k [, 1]> 3) #LC
 welche (PR2.7.k [, 1]> 3) #IPF
 welche (PR2.8.k [, 1]> 3) #COMBO
 welche (PR2.9.k [, 1]> 3) # COMBO2

```
#####
# Misfit
#####
misfitLCA2006 <- (fhat6.kj-f.kj) /fhat6.kj
write.csv (misfitLCA2006, "misfitLCAenEcht2006.txt", row.names = F, quote = FALSE)

#####
#GRUNDSTÜCKE
#####
par (mfrow = c (3,3))
# chi square tegen gemeenten uitgezet plot (PR2.2.k)

Handlung (PR2.3.k)
Handlung (PR2.4.k)
Handlung (PR2.5.k)
Handlung (PR2.6.k)
Handlung (PR2.7.k)
Handlung (PR2.8.k)
Handlung (PR2.9.k)

write.csv (res1.k, "ONAF2006.txt", row.names = F, quote = FALSE) write.csv (res2.k, "STAYER2006.txt",
row.names = F, quote = FALSE) write.csv (res3.k, "NKO2006.txt", row.names = F, quote = FALSE)
write.csv (res4.k, "LR2006.txt", row.names = F, quote = FALSE) write.csv (res5.k, "LPM2006.txt",
row.names = F, quote = FALSE) write.csv (res6.k, "LCA2006.txt", row.names = F, quote = FALSE) write.csv
(res7.k, "IPF2006.txt", row.names = F, quote = FALSE) write.csv (res8.k, "COMBO2006.txt", row.names =
F, quote = FALSE) write.csv (res9.k, "COMBO22006.txt", row.names = F, quote = FALSE)

write.csv (PR2.2.k, "pr22006.txt", row.names = F, quote = FALSE) write.csv (PR2.3.k, "pr32006.txt",
row.names = F, quote = FALSE) write.csv (PR2.4.k, "pr42006.txt", row.names = F, quote = FALSE)
write.csv (PR2.5.k, "pr52006.txt", row.names = F, quote = FALSE) write.csv (PR2.6.k, "pr62006.txt",
row.names = F, quote = FALSE) write.csv (PR2.7.k, "pr72006.txt", row.names = F, quote = FALSE)
write.csv (PR2.8.k, "pr82006.txt", row.names = F, quote = FALSE) write.csv (PR2.9.k, "pr92006.txt",
row.names = F, quote = FALSE)

#####
# Plaatje Nederland
#####
Codenum <- as.numeric (Code)
ExportSaeToSvg = Funktion (IDs, Daten, Ebenen, Methode = "Bereich", Grenzen = "", Ausgabedatei)

{
  # ids: ID-Codes van gemeenten
  # data: Vektor von Daten mit der gleichen Länge wie IDs
  # Ebenen: Die Anzahl der Ebenen, in die quantisiert werden soll (jede Ebene entspricht einer Farbe).

  # Im Moment ist dies unabhängig davon auf 5 gesetzt.
  # outputFile: Dateiname für die exportierte SVG-Datei
  # (fest codierte) templateFile: Dies ist eine SVG-Datei, in der die Farbcodes ersetzt werden müssen.

  # es sollte die kleinen Bereiche von enthalten
  Interesse (dh 'gemeenten' für jetzt)

  Ebenen = 5

  # Definieren Sie Farbcodes (Tabelle mit hexadezimalen Einträgen) und weisen Sie jedem kleinen Bereich einen Index
  zu
}
```

```

# colorCodes = c ("# F8FE00", "# FED400", "# FEA100", "# FE6200", "# FF0000")

colorCodes = c ("# 005670", "# 00A7DA", "# 79D7F4", "#CBEFFA", "#FFFFFF")

dataCodes = 1: Länge (Daten)

if (method == "range") {
  # Ebenengrenzen berechnen
  minData <- min (Daten)
  maxData <- max (Daten)
  binSize <- (maxData - minData) / Ebenen für (i in 1: Ebenen) {

    minVal = minData + (i-1) * binSize maxVal = minVal +
    binSize
    dataCodes [data>= minVal & data <= maxVal] = i

  }}

if (Methode == "Quantil") {
  q = Quantil (Daten, c (.20, .40, .60, .80))
}
if (method == "manual") {
  q = Grenzen
}
dataCodes [data < q [1]] = 1
Datencodes [Daten>= q [1] & Daten < q [2]] = 2 Datencodes [Daten>= q
[2] & Daten < q [3]] = 3 Datencodes [Daten>= q [3] & Daten < q [4]] = 4
dataCodes [data>= q [4]] = 5

# Vorlagendatei analysieren:
# ID für kleine Bereiche extrahieren: Wenn gefunden, ersetzen Sie den Farbcode

# ID-Zeichenfolgen erstellen, die aus 4 Zeichen bestehen idstr <- as.character
(idstr)

idstrlen <- nchar (idstr)
nzeros <- 4 - idstrlen
idstr [nzeros == 0] = Einfügen ("GM", idstr [nzeros == 0], sep = "") idstr [nzeros == 1] = Einfügen ("GM0", idstr
[nzeros == 1], sep = "") idstr [nzeros == 2] = Einfügen ("GM00", idstr [nzeros == 2], sep = "") idstr [nzeros == 3]
= Einfügen ("GM000", idstr [nzeros == 3], sep = "")

# Vorlagendatei lesen
templateFile = "C: \ Programme \ R \ R-
2.6.0 \ Gemeenten2006grijs.svg "
conIn <- Datei (templateFile, "r") allLines <- readLines
(conIn, -1) close (conIn)

# Temporärer Code, damit alle Polygone dieselbe Farbe (grau) für (i in 1: Länge (allLines)) haben
# {
#   thisLine <- allLines [i]
#   if (Teilzeichenfolge (thisLine, 1, 9) == "<g id = " GM ") {
#     # Teilzeichenfolge (thisLine, 28, 34) <- "#cccccc" thisLine = paste
#     (Teilzeichenfolge (thisLine, 1, 27),
# "cccccc", Teilzeichenfolge (thisLine, 33, 35), sep = "")
#     allLines [i] <- thisLine
#   }}
# }
#   outputFile = "C: \ Programme \ R \ R-
2.6.0 \ Gemeenten2006grijsv2.svg "
#   conOut <- Datei (outputFile, "w")

```

```

#       writeLines (allLines, conOut)
#       schließen (conOut)
#       Rückkehr()
# END - Temporärer Code, damit alle Polygone dieselbe Farbe haben

# Ersetzen Sie jetzt Farbcodes im entsprechenden Zeilenbereich <- 1: Länge
(idstr)

iCountGMLinesNotReplaced <- 0
iCountGMLinesReplaced <- 0
für (i in 1: Länge (allLines)) {
    thisLine <- allLines [i]
    testCode <- Teilzeichenfolge (thisLine, 8, 13) thisIndex <- Bereich
[idstr == testCode] if (Länge (thisIndex)> 0) {

        newCode <- colorCodes [dataCodes [thisIndex]]
        Teilzeichenfolge (thisLine, 28, 34) <- newCode allLines [i] <-
        thisLine
        iCountGMLinesReplaced <- iCountGMLinesReplaced + 1} else {

        if (Teilzeichenfolge (thisLine, 1, 9) == "<g id =" GM ") {
            iCountGMLinesNotReplaced <-
iCountGMLinesNotReplaced +1
        }}

    }}

conOut <- Datei (outputFile, "w") writeLines (allLines,
conOut)
schließen (conOut)

# if (method == "range") {
#     return (Liste (minData, maxData))
#}
unsichtbar()
return (Liste (iCountGMLinesReplaced, iCountGMLinesNotReplaced))
}}

# ExportSaeToSvg = Funktion (IDs, Daten, Ebenen, Methode = "Bereich", Grenzen = "", Ausgabedatei))

# q = c (0,8668, 0,9245, 0,9464, 0,9622) q = c (0,8650, 0,9163, 0,9522,
0,9728) outputPath = "C: \\\ Programme \\\ R \\\ R-2.6.0 \\"

ExportSaeToSvg (Codenum, PR2.2.k, 5, Methode = "manuell", Grenzen = q, Ausgabedatei = Einfügen (outPath,
"stayer2006b.svg", sep = "")) ExportSaeToSvg (Codenum, PR2.3.k, 5, method = "manual", bounds = q, outputFile =
paste (outPath, "NKO2006b.svg", sep = "")) ExportSaeToSvg (codenum, PR2.4.k, 5, method = "manual", bounds = q,
outputFile = paste (outPath, "LR2006b.svg", sep = "")) ExportSaeToSvg (Codenum, PR2.5.k, 5, method = "manual",
bounds = q, outputFile = paste (outPath, "LPM2006b .svg ", sep = " ")) ExportSaeToSvg (codenum, PR2.6.k, 5,
method =" manual ", bounds = q, outputFile = paste (outPath," LCA2006b.svg ", sep = " ")) ExportSaeToSvg
(Codenum, PR2.7.k, 5, Methode = "manuell", Grenzen = q, outputFile = paste (outPath, "IPF2006b.svg", sep = ""))
ExportSaeToSvg (codenum, PR2.8.k, 5, method = "manual", bounds = q, outputFile = paste (outPath,
"COMBO2006b.svg") ", sep = " ") ExportSaeToSvg (codenum, PR2.9.k, 5, method =" manual ", bounds = q,
outputFile = paste (outPath," COMBO22006b.svg ", sep = " "))

```