

Anhang zu „Erste Schritte auf kaum erschlossenem Terrain. Zur Stabilität der Partei-identifikation in Deutschland“

Bitte beachten Sie: Es handelt sich hier um ein Manuskript. Bitte zitieren Sie uns nur nach der gedruckten Fassung in der PVS 4/2005.

Kai Arzheimer / Harald Schoen

[Zum Artikel](#)

[Publikationsliste von Kai Arzheimer](#)

[Homepage von Kai Arzheimer](#)

1. Das Sozio-ökonomische Panel als Datenquelle für die Erforschung von Parteiidentifikationen

Die im Folgenden vorgestellten Analysen basieren auf den Daten des Sozio-ökonomischen Panels (SOEP, <http://www.diw.de/deutsch/sop/>). Das SOEP umfasste ursprünglich eine Stichprobe von 4.528 westdeutschen Haushalten, in denen der Haushaltsvorstand die deutsche Staatsangehörigkeit besaß („Sample A“), sowie von 1.326 Haushalten, in denen der Haushaltsvorstand einer der großen Ausländergruppen (Türken, Griechen, Spanier, Italiener, Jugoslawen) angehörte („Sample B“). Seit 1984 werden alle Mitglieder dieser Haushalte, die 16 Jahre oder älter sind, einmal jährlich befragt. Neue Mitglieder eines Haushalts werden ebenfalls in das Panel aufgenommen und auch dann weiterbefragt, wenn sie den Haushalt verlassen. Im Juni 1990 wurde das Panel durch 2.179 ostdeutsche Haushalte ergänzt, deren Vorfstände die deutsche Staatsangehörigkeit besaßen („Sample C“). 1994/95, 1998, 2000 und 2002 wurden außerdem zusätzliche Stichproben gezogen, um bestimmte Bevölkerungssegmente besser untersuchen zu können (Frick et al. 2003: 17-21). Derzeit umfasst das Panel rund 12.000 Haushalte mit etwa 24.000 Personen, die zu einem beträchtlichen Teil seit rund zwei Jahrzehnten regelmäßig befragt werden. Neben den Fragen etwa zur Erwerbs- und Familienbiographie, zur Erwerbsbeteiligung oder zur beruflichen Mobilität, die im Zentrum der Untersuchung stehen, enthält der SOEP-Datensatz auch eine Reihe von Items, die sich auf politische Einstellungen beziehen, darunter die Parteiidentifikation (PI).

Im SOEP wird dabei der klassische Indikator verwendet.¹ Dieser unterstellt, dass die Identifikation mit einer Partei eine Identifikation mit einer anderen Partei ausschließt. Auch wenn etwa für die Niederlande Indizien vorliegen, die die vollständige wechselseitige Ausschließlichkeit von Parteisymphathien nicht unterstützen (siehe Tillie 1995), sprechen doch etliche empirische Befunde für diese Annahme (vgl. etwa Green et al. 2002). Für die Untersu-

¹ „Viele Leute neigen in der Bundesrepublik längere Zeit einer bestimmten Partei zu, obwohl sie auch ab und zu eine andere Partei wählen. Wie ist das bei Ihnen: Neigen Sie – ganz allgemein gesprochen – einer bestimmten Partei zu? Wenn ja, welcher?“

chung der Stabilität von Parteiidentifikationen stellt das SOEP damit eine Datenquelle von unschätzbarem Wert dar, weil es die Überprüfung der fast ausschließlich aus Querschnittsdaten gewonnenen Ergebnisse der deutschen PI-Forschung mit einem Langfristpanel ermöglicht.

Für die vorliegende Studie wurden nur die Samples A (Westdeutsche) und C (Ostdeutsche) verwendet. Gebürtige Ostdeutsche, die im Gebiet der alten Bundesrepublik leben, und gebürtige Westdeutsche, die in die neuen Länder verzogen sind, wurden aus der Analyse ausgeschlossen, um eine Überlagerung etwaiger regionaler Sozialisierungseffekte mit Kontexteffekten der jeweils anderen Region ausschließen zu können. Damit stehen 13.169 Befragte aus den alten und 6.131 Befragte aus den neuen Ländern zur Verfügung. Für die Auswertung der Daten wurde ein „unbalanciertes“ Paneldesign zugrunde gelegt, d.h. es wurden auch solche Befragte berücksichtigt, die an einer oder mehrerer der 19 (West) bzw. 13 (Ost) Befragungswellen nicht teilgenommen bzw. einzelne Fragen nicht beantwortet haben. Anderenfalls hätte sich die Fallzahl auf eine sehr kleine und für die Gesamtbevölkerung mit Sicherheit nicht repräsentative Personengruppe reduziert. Für die Datenanalyse sind diese Ausfälle vergleichsweise unproblematisch, da das von uns verwendete Schätzverfahren wie unten dargelegt das Fehlen einzelner Beobachtungswerte berücksichtigen kann.

Aus forschungspraktischen Gründen wurden der ost- und der westdeutsche Datensatz anschließend in kleinere Teildatensätze zerlegt, die jeweils drei oder vier Panelwellen umfassen. Dies ist deshalb erforderlich, weil der numerische Aufwand mit jeder zusätzlichen Panelwelle um mehr als das Zehnfache ansteigt, sodass selbst moderne Computer rasch an ihre Grenzen stoßen,² und das von uns verwendete Programm WinLTA (<http://methodology.psu.edu/downloads/winlta.html>) maximal fünf Panelwellen verarbeiten kann. Diese Einteilung wurde so getroffen, dass jeweils die letzte Panelwelle in das Jahr einer Bundestagswahl fällt. Mit Blick auf unsere Fragestellung ist diese Einteilung konservativ, da die höchsten Stabilitätskoeffizienten jeweils zwischen der Befragung im Wahljahr und der Befragung im darauf folgenden Jahr zu erwarten sind.³

Das DIW betreibt eine ungewöhnlich intensive Panelpflege. Dennoch sind bei einer derart langen Laufzeit selektive Ausfälle unvermeidlich. Die Teildatensätze wurden deshalb abschließend nach der von Frick et al. (2003) vorgeschlagenen Prozedur gewichtet. Diese be-

² Bei der von uns gewählten Codierung der Variablen ergeben sich für eine Panelwelle $3 \cdot 4 = 12$ (West) bzw. $4 \cdot 4 = 16$ (Ost) mögliche Antwortmuster. Bei drei Panelwellen müssen $12^3 = 1.728$ bzw. $16^3 = 4.096$ mögliche Antwortmuster berücksichtigt werden, bei vier Panelwellen sind es dementsprechend schon $12^4 = 20.736$ bzw. $16^4 = 65.536$ Antwortmuster. Zusätzliche Probleme für die Parameterschätzung ergeben sich daraus, dass aufgrund der beschränkten Fallzahl viele Zellen der zugrundeliegenden Tabelle gar nicht oder sehr schwach besetzt sind.

³ Etwaige durch den Wahlkampf zu erklärende Kristallisationseffekte müssten zu einem relativ *schwachen* Zusammenhang zwischen der PI im Jahr vor der Wahl und der PI im Wahljahr führen.

rücksichtigt sowohl die Wahrscheinlichkeit, dass eine Person in der ersten Welle des untersuchten Zeitraums befragt wurde, als auch die Wahrscheinlichkeiten dafür, dass diese Person sich an den jeweils darauf folgenden Befragungswellen beteiligt hat (für Details zur Gewichtung des SOEP siehe Pannenberg et al. 2003).⁴ Die gewichteten Daten sollten deshalb für den jeweiligen Untersuchungszeitraum repräsentativ sein.

2. Eine Einführung in das Verfahren der Latent Transition Analysis (LTA)

Im Zentrum unseres Forschungsinteresses steht die Frage nach der Stabilität von Parteiidentifikationen, d.h. nach der Entwicklung einer nominalskalierten, nicht direkt beobachtbaren Variablen auf der Individualebene. Zur Modellierung solcher Veränderungen ist das von uns angewandte Verfahren der Latent Transition Analysis in besonderer Weise geeignet, da es eigens zur Untersuchung derartiger Prozesse entwickelt wurde.

Bei der LTA handelt es sich um eine Weiterentwicklung der latenten Klassenanalyse (LCA). Obwohl die LTA in der politischen Soziologie bislang kaum bekannt ist, sind ihre Grundannahmen leicht nachzuvollziehen. Ausgangspunkt der Modellierung ist die Überlegung, dass die Befragten auf der Grundlage ihrer Parteiidentifikation in eine von drei bzw. vier Gruppen eingeordnet werden können. Sie identifizieren sich entweder mit der Union, mit der SPD, der PDS⁵ oder mit einer anderen bzw. gar keiner Partei.⁶ Da es sich bei der PI um eine Einstellungsvariable handelt, die nicht direkt beobachtet, sondern nur mit Hilfe eines fehlerbehafteten Indikators (nämlich der gängigen PI-Frage) gemessen werden kann, sind diese Gruppen „latent“. In der Terminologie der LTA wird die Mitgliedschaft in einer dieser Gruppen als „latenter Status“ bezeichnet.

Anders als im Falle der LCA können mit der LTA jedoch *dynamische* latente Variablen modelliert werden. Ein Befragter muss also nicht notwendigerweise zeitlebens in derselben latenten Gruppe verbleibt. Vielmehr ist es mit Hilfe der LTA möglich, auf der Grundlage em-

⁴ Die vom DIW bereitgestellten Faktoren bewirken eine implizite Hochrechnung der Ergebnisse auf die Gesamtbevölkerung. In einem letzten Schritt haben wir die Gewichte deshalb mit einem Korrekturfaktor multipliziert, um die Fallzahlen, die unseren Berechnungen zugrunde liegen, an die Zahl der tatsächlich befragten Personen anzunähern.

⁵ Aus Fallzahlgründen konnten die PDS-Anhänger nur in den neuen Ländern als eigenständige Gruppe analysiert werden.

⁶ Aufgrund der geringen Zahl von Befragten, die sich mit der FDP, den Grünen oder einer der kleineren Parteien (zu denen in den alten Ländern auch die PDS zählt) identifizieren, wurden alle Personen, die sich *nicht* mit Union, PDS (nur in Ostdeutschland) oder SPD identifizieren, in dieser Gruppe zusammengefasst. Aus der Perspektive unserer Fragestellung ist dies relativ unproblematisch, da sich nur ein vergleichsweise geringer Anteil der

pirischer Daten die Wahrscheinlichkeit dafür zu schätzen, dass die Angehörigen einer gegebenen latenten Gruppe zwischen zwei Messzeitpunkten in eine der anderen Gruppen wechseln, d.h. ihre PI ändern. In diesem Fall ändert sich der latente Status des Befragten.⁷

Abbildung A-1 etwa hier

Die Schätzung dieser Übergangswahrscheinlichkeiten hat dem Verfahren seinen Namen gegeben. Im Unterschied zu anderen Maßzahlen, die zur Analyse von Paneldaten verwendet werden, haben Übergangs- bzw. Bleibewahrscheinlichkeiten eine denkbar klare und anschauliche Interpretation: Sie entsprechen dem geschätzten Anteil derjenigen Gruppenmitglieder, die bis zur nächsten Befragungswelle in derselben latenten Gruppe verbleiben bzw. in eine andere Gruppe überwechseln. Mit ihrer Hilfe lässt sich bereits ein Großteil der Fragen beantworten, die sich Politikwissenschaftler typischerweise stellen, wenn sie kategoriale Paneldaten analysieren.

Abbildung A-1 verdeutlicht dies durch einige fiktive Zahlenwerte. Aus ihr geht hervor, dass 90 Prozent der in der ersten Welle befragten Unionsanhänger sich auch zum Zeitpunkt der zweiten Befragungswelle mit den Christdemokraten identifizieren. Lediglich fünf Prozent identifizieren sich nun mit der SPD, weitere fünf Prozent haben sich einer dritten Partei zugewandt oder identifizieren sich mit keiner Partei. Ähnlich hoch ist die Haltequote der CDU/CSU zwischen der zweiten und der dritten Befragungswelle: Von denjenigen, die sich während der zweiten Befragungswelle mit der Union identifizieren, neigten zum Zeitpunkt der dritten Welle immer noch 85 Prozent den Christdemokraten zu. Multipliziert man beide Größen miteinander, ergibt sich der Anteil derjenigen unter den ursprünglichen Anhängern von CDU und CSU, die sich zu allen drei Zeitpunkten mit der Union identifizierten (ca. 77 Prozent). Die PI mit der CDU/CSU ist also über den hier betrachteten Zeitraum hinweg recht stabil. In der Residualgruppe sind hingegen relativ große Fluktuationen zu verzeichnen: Nur 40 bzw. 45 Prozent der Befragten in dieser Gruppe behalten diesen latenten Status bei. Dabei

Wähler kleinerer Parteien mit diesen identifiziert, und wir in erster Linie an der Stabilität der „klassischen“ Identifikationen interessiert sind, die in der Vergangenheit oft sozialstrukturell vermittelt waren.

⁷ Wir gehen hier von einem Modell „erster Ordnung“ aus, d.h. nach unserer Spezifikation hat die Parteiidentifikation zum Zeitpunkt $t-2$ keinen Einfluss auf die Wahrscheinlichkeit eines Übergangs zwischen den Zeitpunkten $t-1$ und t . Mit einem Modell „zweiter Ordnung“ könnte überdies überprüft werden, ob ältere Identifikationen in Übereinstimmung mit den theoretischen Erwartungen tatsächlich stabiler sind. Von einer solchen Spezifikation haben wir aber Abstand genommen, da wir auf diese Weise lediglich ein Befragungsjahr mehr berücksichtigen könnten, was aus theoretischer Perspektive von untergeordnetem Interesse ist. Dieser geringe Vorteil würde mit zwei gravierenden Problemen erkaufte: Die ohnehin bereits sehr hohe Zahl der zu schätzenden Koeffizienten würde sich dramatisch erhöhen, was die Interpretation erschwert. Zudem wären gerade in den neuen Ländern,

sind Wechsel *innerhalb* der Residualgruppe noch nicht berücksichtigt, die durch die gewählte Form der Modellierung verdeckt werden.

An den Übergangswahrscheinlichkeiten lässt sich also unmittelbar ablesen, wie stabil Parteiidentifikationen tatsächlich sind, und nach welchem Muster die Bürger ihre Orientierungen verändern. Zugleich ist in Abbildung A-1 aber auch das zentrale Problem bei der Interpretation einer LTA zu erkennen: Selbst bei diesem relativ einfachen Design werden bereits 18 Wahrscheinlichkeiten geschätzt. Erweitert man das Modell um eine mögliche PI mit der PDS sowie eine weitere Panelwelle, ergeben sich bereits 64 Parameter. Die Darstellung und Interpretation der Ergebnisse gestaltet sich infolgedessen ausgesprochen schwierig. Wir konzentrieren uns deshalb auf die Betrachtung der Bleibewahrscheinlichkeiten für die zwei bzw. drei großen Parteien und sprechen Übergangswahrscheinlichkeiten nur insoweit an, als sie für unsere Hypothesen relevant sind.

In der Terminologie der LTA werden die Übergangswahrscheinlichkeiten als „t-Parameter“ bezeichnet. Neben den t-Parametern existieren noch vier weitere Gruppen von Koeffizienten. Diese zeichnen sich ebenfalls dadurch aus, dass ihre Interpretation sehr anschaulich ist, während ihre schiere Zahl bisweilen verwirrend sein kann. Unter diesen sind zunächst die so genannten P-Parameter⁸ von Interesse, die den Zusammenhang zwischen der dynamischen latenten Variable (PI) und den beobachtbaren Indikatoren (PI-Frage) beschreiben und somit über die Qualität des Messmodells informieren.⁹

Im Gegensatz zu den vergleichbaren Koeffizienten in einem Strukturgleichungsmodell oder einer Faktoranalyse können auch diese Werte schlicht als Wahrscheinlichkeiten bzw. relative Häufigkeiten interpretiert werden: Ein Wert von 0,95 für die Anhänger der Union und die manifeste Antwort „Union“ zum Zeitpunkt t etwa würde bedeuten, dass die übergroße Mehrheit der Unionsanhänger in dieser Panelwelle mit großer Sicherheit die betreffende Ant-

wo ein Unterschied von einem Jahr im Alter der PI am ehesten von Bedeutung sein könnte, die Zellen der Tabelle, die der LTA letztlich zugrunde liegen, so schwach besetzt, dass präzise Schätzungen unmöglich werden.

⁸ Das Symbol ρ steht hier für den griechischen Großbuchstaben Rho.

⁹ In Panel-Analysen für kontinuierliche Variablen, die auf Regressionsrechnungen bzw. Strukturgleichungsmodellen basieren, versucht man, die Varianz der beobachteten Variablen in zwei Komponenten – Messfehler und „wahrer“ Wert – zu zerlegen. Dabei müssen häufig mehr oder minder restriktive Annahmen über die Fehlervarianz der Indikatoren bzw. die Stabilität und Varianz der latenten Konstrukte getroffen werden, um die Modelle zu identifizieren (McAdams 1986). Bei der Schätzung von LTA-Modellen stellt sich im Grunde ein ähnliches Problem. Hier geht man davon aus, dass ein Befragter mit einem bestimmten latenten Status nicht deterministisch die „richtige“ Antwortkategorie wählt, sondern sich mit einer zu bestimmenden Wahrscheinlichkeit auch für eine „falsche“ Kategorie entscheiden kann. Obwohl die Identifikation von LTA-Modellen häufig unproblematisch ist, kann es auch hier notwendig sein, zusätzliche Annahmen in das Modell aufzunehmen. Sinnvoll ist dabei besonders die Annahme stabiler Reliabilitäten (Collins/Wugalter 1992: 137-138). Im Falle der von uns berechneten Modelle konnten jedoch alle P-Parameter frei geschätzt werden (vgl. Abschnitt 3). Dass sich dabei weitgehend konstante Reliabilitäten ergeben, betrachten wir als einen weiteren Hinweis darauf, dass unser Modell den Daten angemessen ist.

wortkategorie wählen würde. Daraus ließe sich schließen, dass dieser Indikator sehr gut zur Messung der PI geeignet ist.¹⁰

Für *jede* Kombination von latentem Status, Kategorien der Indikatorvariable(n) und Panelwelle wird jeweils ein Parameter geschätzt. Probleme ergeben sich also auch hier nicht aus der Interpretation, sondern vielmehr aus der enormen Zahl der Koeffizienten. So würden für das in Abbildung A-1 skizzierte Design bereits 27 Parameter benötigt. Wird das Design wiederum um die Identifikation mit der PDS und eine zusätzliche Welle erweitert, ergeben sich 64 Parameter – wohlgermerkt bei der Berücksichtigung eines einzelnen Indikators. Durch einen zweiten Indikator mit ebenfalls vier Antwortvorgaben würde sich diese Zahl verdoppeln. Die γ -Parameter lassen also eine sehr differenzierte Beurteilung der Messgüte zu, sind aber schon bei relativ einfachen Modellen nur schwer überschaubar. Bei der empirischen Analyse konzentrieren wir uns deshalb auf die für unser Forschungsproblem wichtigste Frage: Wie viel Prozent der Unions- und SPD-Identifizierer geben sich tatsächlich als solche zu erkennen?

Vergleichsweise gering ist hingegen die Zahl der d -Parameter. Diese entsprechen den geschätzten Anteilen der jeweiligen latenten Gruppen an der Gesamtpopulation zum Zeitpunkt der ersten Panelwelle.

Mit Hilfe von Übergangswahrscheinlichkeiten, Antwortwahrscheinlichkeiten und Anteilswerten lässt sich bereits ein einfaches Modell der Messung und Stabilität von PI über die Zeit formulieren. Allerdings ist aufgrund der oben vorgestellten theoretischen Überlegungen zu erwarten, dass die Stabilität von PI mit dem politischen Interesse variiert. Zudem ist zu erwarten, dass das politische Interesse einen Einfluss darauf hat, ob sich ein Bürger überhaupt mit einer Partei identifiziert, und wenn ja, mit welcher. Deshalb nutzen wir die Möglichkeit, das Modell um zwei *latente Klassen* zu ergänzen, die durch das politische Interesse (hoch vs. mittel und niedrig) definiert werden. Ähnlich wie bei der PI handelt es sich beim Interesse um eine Variable, die nicht direkt beobachtet werden kann, sondern durch entsprechende Interviewfragen erschlossen werden muss. Im Gegensatz zu PI fassen wir diese Variable jedoch als statisch auf, d.h. wir gehen davon aus, dass Menschen, die zu Beginn der Untersuchung zur Gruppe der Hochinteressierten zählen, in dieser Gruppe verbleiben, und dass während des Untersuchungszeitraumes niemand in diese Gruppe hineinwechselt. Dementsprechend betrachten wir die jährlichen Messungen des politischen Interesses als voneinander unabhängige, multiple Indikatoren für denselben Sachverhalt.

¹⁰ Präziser müsste man formulieren: Aus dem Wert von 0,95 ergibt sich, dass die *Antwortvorgabe* „Union“ sehr gut zur Messung von Identifikationen mit den Christdemokraten geeignet ist.

Diese Annahme ist weniger restriktiv, als es zunächst den Anschein hat. Zum einen scheint sie durch die Empirie gedeckt: Selbst wenn man die Antworten der Befragten zum Nennwert nimmt, d.h. die Tatsache ignoriert, dass Messungen immer fehlerbehaftet sind, fällt auf, dass die große Mehrheit der Menschen, die sich in einer Befragung als „politisch interessiert“ oder „stark interessiert“ bezeichnen, auch in den folgenden Wellen eine dieser beiden Antwortkategorien wählt. Gleiches gilt mit umgekehrtem Vorzeichen für diejenigen, die „nicht so stark“ oder „überhaupt nicht“ interessiert sind. Zum anderen schließt das Modell nicht aus, dass sich beispielsweise einzelne eigentlich hochinteressierte Bürger bei einer der drei bzw. vier Befragungswellen aus Ärger oder Unaufmerksamkeit als desinteressiert bezeichnen oder politisch apathische Menschen aus ähnlichen Gründen die Kategorie „stark interessiert“ wählen. Auch hier gehen wir nicht von einer perfekten Messung des Konstrukts durch die Indikatoren aus, sondern bestimmen deren Reliabilität vielmehr empirisch.

Die Koeffizienten, die den Zusammenhang zwischen Indikatorvariablen und latenten Klassen beschreiben, werden im Kontext der LTA als γ -Parameter bezeichnet und sind analog zu den γ -Parametern als Wahrscheinlichkeit dafür zu interpretieren, dass ein Mitglied einer gegebenen Klasse eine der (vier) vorgegebenen Antwortmöglichkeiten wählt. Dementsprechend gilt auch hier, dass das eigentliche Problem bei der Interpretation dieser Schätzungen in der großen Zahl der Parameter liegt, die betrachtet werden müssen. Da wir weniger am politischen Interesse selbst als vielmehr an dessen Wirkung auf die Stabilität und Verbreitung von Parteiidentifikationen interessiert sind, konzentrieren wir uns in diesem Zusammenhang auf die globale Frage nach der Reliabilität der Indikatoren, ohne näher auf Einzelbefunde einzugehen.

Eine letzte Gruppe von Koeffizienten wird schließlich unter der Bezeichnung δ -Parameter zusammengefasst. Analog zu den d -Parametern stellen diese eine Schätzung für den Anteil der betreffenden Klasse an der Gesamtpopulation dar. Da das politische Interesse von uns als über die Zeit stabil modelliert wird, muss für jeden Untersuchungszeitraum nur ein Set dieser Parameter geschätzt werden.

Die eigentliche Kovariation zwischen dem politischen Interesse und der Häufigkeit und Stabilität von Parteiidentifikationen wird bei der LTA ebenfalls in sehr anschaulicher, aber numerisch aufwendiger Weise modelliert: Für jede der latenten Klassen werden eigene t -, P - und d -Parameter geschätzt, sodass beispielsweise die Bleibewahrscheinlichkeiten für politisch hochinteressierte und politisch weniger interessierte SPD-Anhänger direkt miteinander verglichen werden können. Dadurch verdoppelt sich allerdings die Zahl dieser Parameter, was nicht gerade zur Übersichtlichkeit beiträgt.

3. Die Schätzung der Modellparameter

Da im Rahmen der LTA ausschließlich kategoriale Variablen betrachtet werden, basiert die Schätzung der Parameter letztlich auf einer Analyse der Muster, die sich aus den Antworten der Befragten auf die Items zur Parteiidentifikation und zum politischen Interesse ergeben. Tabelle A-1 zeigt zur Illustration für den Zeitraum von 1995-98 (alte Länder) Ausprägung und Häufigkeit von drei der 20.736 theoretisch möglichen Antwortmuster.

- Tabelle A-1 etwa hier -

Bei allen fünf zu bestimmenden Gruppen von Modellparametern handelt es sich um Wahrscheinlichkeiten. Um diese schätzen zu können, wird das Modell so umformuliert, dass sich die erwartete relative Häufigkeit jedes Antwortmusters aus den Summen über die Produkte dieser Wahrscheinlichkeiten ergibt. So hängt die Häufigkeit des Antwortmusters 319 davon ab, wie hoch der Anteil von stark und weniger stark interessierten Bürgern ist, mit welchen Parteien sich die Bürger zu welchem Zeitpunkt tatsächlich identifizieren, und wie sich diese Eigenschaften auf die Wahl der Antwortkategorien auswirkt.

Die eigentliche Schätzung erfolgt in WinLTA mit Hilfe des EM-Algorithmus¹¹ (Gill 2002: Kapitel 8.6). Dabei handelt es sich um ein alternatives Verfahren zur Bestimmung von Maximum Likelihood Schätzungen. Ohne näher auf die technischen Einzelheiten einzugehen, lässt sich festhalten, dass auch hier die Parameter, d.h. die geschätzten Wahrscheinlichkeiten, so lange iterativ variiert werden, bis eine möglichst gute Anpassung der erwarteten Häufigkeiten an die tatsächliche Verteilung der Antwortmuster erzielt ist.

Vom „normalen“ Maximum Likelihood-Verfahren unterscheidet sich der EM-Algorithmus dadurch, dass eine Schätzung auch dann durchgeführt werden kann, wenn ein Respondent eine oder mehrere Fragen nicht beantwortet hat („item missing“). Gerade im Kontext einer Panel-Studie ist dies ein nicht zu überschätzender Vorteil, da eine Beschränkung der Analyse auf die (wenigen) Befragten, die in jeweils drei bzw. vier Befragungswellen *alle* relevanten Items vollständig beantwortet haben („listwise deletion“), nicht nur die Fallzahlen reduzieren, sondern mit Sicherheit auch zu substantziellen Verzerrungen der Parameterschätzungen und zu einer Unterschätzung der Standardfehler führen würde (King et al.

¹¹ EM steht für „Expectation-Maximization“.

2001: 51-52).¹² Im Gegensatz zu den gängigen ML-Prozeduren produziert der EM-Algorithmus allerdings keine Schätzungen für die Variabilität der Parameter. Mit Hilfe des DA-Algorithmus (Gill 2002: 325-327)¹³, bei dem es sich im Grunde um eine stochastische Variante des EM-Algorithmus handelt, lassen sich jedoch Standardfehler generieren, die auch die zusätzliche Unsicherheit durch das Fehlen einzelner Antworten berücksichtigen und deshalb als besonders konservativ gelten können. Problematisch beim Einsatz von EM- und DA-Algorithmus ist in erster Linie der immense numerische Aufwand: Für jede der hier präsentierten Schätzungen wurden insgesamt einige tausend Iterationen benötigt, was selbst auf einem modernen PC mehreren Stunden Rechenzeit entspricht.

Als Grundlage zur Beurteilung der Anpassung des Modells an die empirischen Daten errechnet WinLTA die G^2 -Statistik (Likelihood-Ratio χ^2), die asymptotisch χ^2 -verteilt ist und deshalb zum Test der Nullhypothese verwendet werden kann, dass etwaige Abweichungen zwischen Modell und empirisch beobachteten Daten auf zufällige Einflüsse zurückgehen. Dieser Test ist jedoch in mehrfacher Hinsicht problematisch. Denn erstens wird die G^2 -Statistik auch durch das Fehlen einzelner Messwerte beeinflusst, d.h. die Aussagen über die Anpassung des Modells sind zu pessimistisch. Dies lässt sich allerdings relativ leicht durch die von WinLTA standardmäßig vorgenommene Berechnung eines Korrekturfaktors kompensieren (Collins et al. 2002).¹⁴ Zweitens ist die χ^2 -Verteilung bekanntermaßen keine gute Approximation der tatsächlichen Verteilung von G^2 , wenn die zugrunde liegende mehrdimensionale Häufigkeitstabelle wie im Falle unserer Analysen viele schwach besetzte Zellen aufweist. In diesem Fall können theoretische und empirische Verteilung der Prüfgröße dramatisch voneinander abweichen, sodass der Signifikanztest nicht anwendbar ist (Langeheine et al. 1996). Drittens schließlich stellt sich auch und gerade bei der Beurteilung der Modellanpassung grundsätzlich die Frage nach der Sinnhaftigkeit der gängigen Signifikanztests (Gill 1999).

Denn ein Anpassungstest prüft nicht, ob sich geschätzte und beobachtete Häufigkeiten in inhaltlich bedeutsamer Weise unterscheiden. Errechnet wird vielmehr die Wahrscheinlichkeit p , die in der Stichprobe auftretenden Diskrepanzen zwischen Modell und Daten zu beobachten, wenn das Modell die Daten „in Wirklichkeit“, d.h. in der Grundgesamtheit perfekt repro-

¹² Erfreulicherweise ist der Anteil der item missings in unseren Analysen vergleichsweise niedrig. Beispielsweise haben für den Zeitraum von 1984-87 fast 94 Prozent der Panelteilnehmer in allen vier Wellen alle für uns interessanten Fragen beantwortet. Im Sinne einer konservativen Analysestrategie haben wir dennoch auf ein listenweises Löschen der unvollständigen Fälle verzichtet.

¹³ DA steht für Data Augmentation. Die EM-Schätzung dient üblicherweise als Ausgangspunkt für eine Serie von DA-Schätzungen.

¹⁴ Dieser Korrekturfaktor kann seinerseits wieder verwendet werden, um zu testen, ob die Annahmen über die Antwortausfälle mit den Daten vereinbar sind. In unserem Fall gab es dabei keinerlei Hinweise auf etwaige Probleme.

duziert, also *keinerlei Abweichungen* zwischen Modell und Daten auftreten (Nullhypothese). Ist diese errechnete Wahrscheinlichkeit kleiner als ein vorab festgelegter Schwellenwert von bspw. fünf Prozent, dann gelten die Abweichungen zwischen Modell und Daten als „signifikant“.

Die Vorstellung, dass irgendein statistisches oder verbal formuliertes Modell in der Lage sein könnte, die Wirklichkeit perfekt zu erklären, ist aber offensichtlich unrealistisch. Deshalb ist immer davon auszugehen, dass auch in der Grundgesamtheit Abweichungen zwischen den realen und den vom Modell implizierten Häufigkeiten auftreten. Ob diese vom Signifikanztest „entdeckt“ werden, hängt neben der Komplexität des Modells einerseits von der Größe dieser Abweichungen, andererseits aber auch von der Zahl der Fälle in der Stichprobe ab, da mit der Fallzahl die „Stärke“ des Tests, d.h. die Fähigkeit, auch kleinere Abweichungen zwischen Modell und Daten sicher nachweisen zu können, zunimmt. Mit sehr großen Datensätzen wie dem unseren sprechen Abweichungstests auch auf triviale Unterschiede zwischen beobachteten und erwarteten Häufigkeiten an.¹⁵ [Im Kontext](#) der Interpretation von Strukturgleichungsmodellen wird der Nutzen von Abweichungstests deshalb häufig in Frage gestellt (vgl. z.B. Hu/Bentler 1995).

Gelöscht: Im Kontext

Bei den von uns geschätzten Modellen kommt es in den alten Ländern in zwei Fällen zu „signifikanten“ Abweichungen zwischen den beobachteten und den vom Modell implizierten Häufigkeiten. Betroffen sind die Modelle für die Zeiträume 1984-1987 und 1988-1990. Von den übrigen Modellen für das Gebiet der alten Bundesrepublik unterscheiden sich diese beiden nur insofern, als sie bei vergleichbar großer Fallzahl weniger komplex sind, sodass der Anpassungstest auf einer erheblich geringeren Zahl an Freiheitsgraden basiert.¹⁶ Da die Schätzung der Parameter über die Zeit hinweg nahezu identisch sind, gehen wir vor dem Hintergrund der oben skizzierten Überlegungen zur Bedeutung von Anpassungstests davon aus, dass die beiden signifikanten Abweichungen von einem perfekten Fit auf die Kombination von vergleichsweise einfachen Modellen und sehr großen Fallzahlen zurückzuführen sind, weshalb wir sie als unproblematisch betrachten.

Ein letzter Punkt betrifft schließlich die Wahl der Startwerte. Wie bei allen iterativen Verfahren muss auch beim EM-Algorithmus eine Ausgangslösung vorgegeben werden, die dann schrittweise verbessert wird. Dabei besteht grundsätzlich die Möglichkeit, dass der Al-

¹⁵ Dies ergibt sich unmittelbar aus der Konstruktion von G^2 , dessen Wert ebenso wie der von χ^2 eine Funktion der Fallzahl ist.

¹⁶ Das erste Modell ist deshalb weniger komplex, weil es zwar vier Panelwellen, aber nur drei Messungen des politischen Interesses beinhaltet; das zweite Modell beschränkt sich ohnehin auf drei Panelwellen. Daraus ergeben sich 5.076 bzw. 1.644 Freiheitsgrade. Die drei Modelle für den Zeitraum von 1991 bis 2002 weisen hingegen jeweils 20.622 Freiheitsgrade auf.

gorithmus nicht das globale, sondern nur ein lokales Maximum der Likelihood-Funktion findet. Deshalb ist es wichtig, Startwerte zu verwenden, die bereits möglichst dicht bei den wahren, aber unbekanntem Werten der Koeffizienten liegen.

Die von uns verwendeten Startwerte spiegeln unsere theoriegeleiteten Erwartungen wider: Den Anteil der politisch Hochinteressierten (γ -Werte) haben wir auf 20 Prozent festgesetzt, bei den ρ -Werten sind wir davon ausgegangen, dass die politisch Hochinteressierten mit einer Wahrscheinlichkeit von 60 Prozent für die Kategorie „sehr stark interessiert“ und mit einer Wahrscheinlichkeit von 20 Prozent für die Kategorie „stark interessiert“ entscheiden, während die Kategorien „schwach interessiert“ und „sehr schwach interessiert“ mit einer Wahrscheinlichkeit von je 10 Prozent gewählt werden.¹⁷ Als Startwerte für die Ausgangsverteilung der Parteiidentifikation (d -Wert) haben wir in den alten Ländern unabhängig vom politischen Interesse einen relativ niedrigen Anteilswert von 20 Prozent für den latenten Status „keine/andere PI“ und einen Anteilswert von je 40 Prozent für SPD und Union gewählt. In den neuen Ländern haben wir als Startwert für die Residualgruppe einen Anteil von 70 Prozent festgelegt und jeder der drei großen Parteien einen initialen Anteil von 10 Prozent zugeordnet. Bei den Vorgaben für die t -Matrizen haben wir im Westen eine Bleibewahrscheinlichkeit von 80 Prozent für alle Gruppen festgelegt, während wir in den neuen Ländern von einer Bleibewahrscheinlichkeit von 70 Prozent ausgehen. Als Startwert für die Wahrscheinlichkeit des Übergangs in eine beliebige andere Gruppe haben wir jeweils einen Wert von 10 Prozent angesetzt. Bei der Wahrscheinlichkeit, die PI-Frage „richtig“ zu beantworten schließlich gehen wir in den alten Ländern von einem Wert von 80, in den neuen Ländern von einem Wert von 70 Prozent aus. Als Startwert für die Wahrscheinlichkeit einer „falschen“ Antwort haben wir jeweils 10 Prozent angesetzt.

Mit diesen Startwerten konvergiert der EM-Algorithmus innerhalb von einigen Dutzend bis einigen hundert Iterationen, ohne dass das Iterationsprotokoll Hinweise auf mögliche Probleme geben würde. Um abschätzen zu können, ob es dennoch Einflüsse der von uns gewählten Startwerte auf das substanzielle Ergebnis unserer Analysen gibt, haben wir bei dem einfachsten von uns geschätzten Modelle (alte Länder 1988-1990) zunächst einige willkürliche Änderungen vorgenommen, nämlich den Anteil der politisch Hochinteressierten auf 50 Prozent festgelegt, die Bleibewahrscheinlichkeit der SPD-Anhänger in allen t -Matrizen auf 50 Prozent herabgesetzt¹⁸ und die P -Werte für die politisch in hohem Maße interessierten Mit-

¹⁷ Die Startwerte für die Antwortwahrscheinlichkeiten der politisch weniger interessierten haben wir spiegelbildlich dazu gewählt.

¹⁸ Die „überzähligen“ 30 Prozentpunkte haben wir abwechselnd der Übergangswahrscheinlichkeit zur Union bzw. zur Residualgruppe zugeschlagen.

glieder der Residualgruppe verändert.¹⁹ Trotz dieser vergleichsweise unsinnigen Startwerte wurde hier innerhalb von relativ wenigen Iterationen das Konvergenzkriterium erreicht. Der resultierende G^2 -Wert ist mit dem der ursprünglichen Lösung identisch. Dies gilt auch für die Mehrzahl der geschätzten Koeffizienten. Dort wo sich Abweichungen zeigen, liegen diese im Bereich von maximal zwei bis drei Prozentpunkten. Wichtig ist aus unserer Sicht vor allem, dass die Wahl sehr niedriger Startwerte in der t- und der ρ -Matrix keinen substanziellen Einfluss auf die endgültigen Schätzungen hat.

Um uns zusätzlich abzusichern, haben wir in einem zweiten Experiment einen möglichen Einwand gegen unsere Ergebnisse vorweggenommen und die Startwerte der d- und der t-Matrix der manifesten Einstellung der PI im Jahr 1988 angeglichen. Diese Konfiguration entspricht einer Extremposition in der Diskussion um die PI, nämlich der Hypothese, dass die Identifikationen als solche gar nicht existent sind, sondern lediglich kurzfristigen Wahlabsichten entsprechen und deshalb über einen längeren Zeitraum von einem Jahr hinweg rein zufällig, wenn auch mit unterschiedlichen Wahrscheinlichkeiten fluktuieren. Die Wahrscheinlichkeit, dass eine solche Pseudoidentifikation „beibehalten“ wird, entspricht aus dieser Perspektive schlicht dem Zuspruch, den die entsprechende Partei im Mittel in der Bevölkerung findet.

Auch hier konvergiert der EM-Algorithmus innerhalb relativ weniger Iterationen; und wiederum ist der resultierende G^2 -Wert mit dem der ursprünglichen Lösung identisch. Trotz der extrem unterschiedlichen Startwerte unterscheiden sich die geschätzten Werte der d-Matrix auch hier wieder um maximal zwei bis drei Prozentpunkte. Etwas größere Abweichungen zeigen sich in der t-Matrix. Hier tritt bei einem der 36 Koeffizienten (Bleibewahrscheinlichkeit der politisch in hohem Maße interessierten Mitglieder der Residualgruppe von 1988 nach 1989) eine Differenz von sieben Prozentpunkten auf, bei einem zweiten Koeffizienten (Bleibewahrscheinlichkeit der weniger stark interessierten Mitglieder der Residualgruppe von 1988 nach 1989) ergibt sich eine Abweichung von knapp fünf Prozentpunkten. Ansonsten sind alle geschätzten Werte nahezu identisch. Aufgrund dieser Ergebnisse können wir mit großer Sicherheit davon ausgehen, dass die im Folgenden vorgestellten Ergebnisse *kein* Artefakt unserer Startwerte sind, d.h. dass wir nicht etwa deshalb relativ hohe Schätzungen für die Stabilität der PI erhalten, weil der Algorithmus in der Nähe der von uns gewählten Werte in einem lokalen Maximum „stecken bliebe“.

¹⁹ Die Wahrscheinlichkeit einer „richtigen“ Antwort wurde auf 60 Prozent reduziert, die Wahrscheinlichkeit, dass ein Mitglied dieser Gruppe behauptet, Anhänger der Union oder der SPD zu sein, auf jeweils 20 Prozent heraufgesetzt.

Tabelle A-1: Häufigkeit von drei empirischen Antwortmustern, alte Länder 1995-1998

Nr.	PI 1995	PI 1996	PI 1997	PI 1998	Interesse 1995	Interesse 1996	Interesse 1997	Interesse 1998	N
319	SPD	SPD	SPD	SPD	schwach	schwach	schwach	schwach	336
320	SPD	SPD	SPD	Union	s. stark	s. stark	s. stark	s. stark	40
321	SPD	SPD	SPD	Union	s. stark	s. stark	s. stark	stark	9

Quelle: SOEP 1995-1998, Sample A und C.

Abbildung A-1: Hypothetische Übergangs- und Bleibewahrscheinlichkeiten

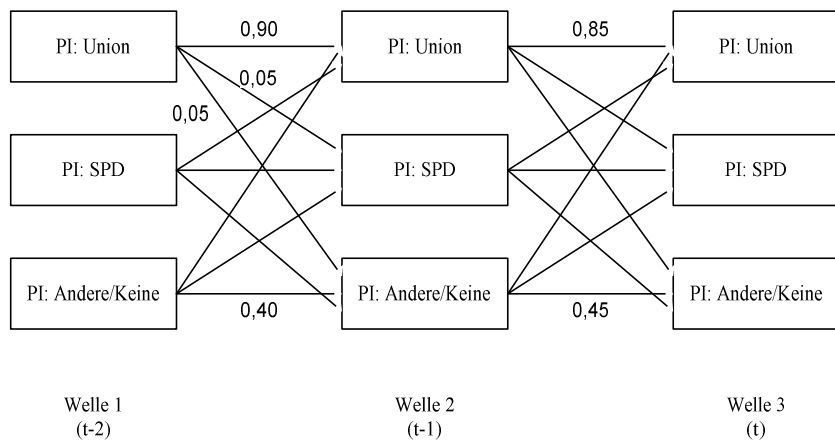
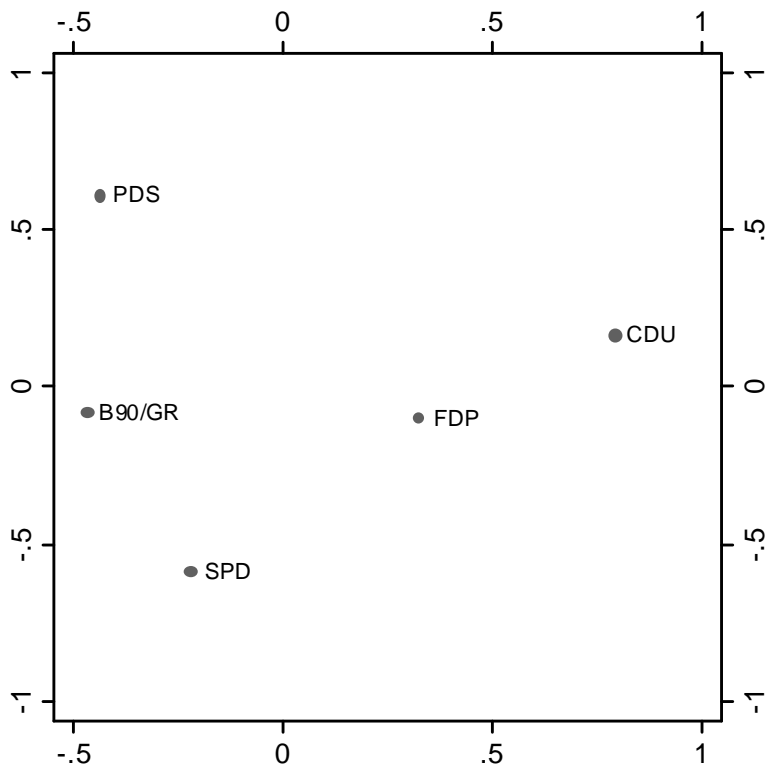


Abbildung A-2: Die räumliche Struktur des Parteienwettbewerbs in Ostdeutschland (2002) –
Multidimensionale Skalierung



Quelle: Politische Einstellungen, politische Partizipation und Wählerverhalten im vereinigten Deutschland 2002, Querschnitt, ZA-Nr. 3681. Zu sehr ähnlichen Ergebnissen kommen aber auch Arzheimer/Klein (1997: 49-50) mit Daten aus dem Jahr 1994 (ZA-Nr. 3065).

Literatur

Arzheimer, Kai/Markus Klein, 1997: Die Wähler der REP und der PDS in West- und Ostdeutschland, in: Backes, Uwe/Jesse, Eckhard (Hrsg.), Jahrbuch Extremismus und Demokratie. Baden-Baden, 39-63.

Collins, Linda M./Lanza, Stephanie T./Schafer, Joseph L./Flaherty, Brian P., 2002: WinLTA User's Guide. Version 3.0.
<http://methcenter.psu.edu/downloads/winlta30/mainmanual.pdf> (03.08.05).

Collins, Linda M./Wugalter Stuart E., 1992: Latent Class Models for Stage-Sequential Dynamic Latent Variables, in: Multivariate Behavioral Research 27, 131-157.

- Frick, Joachim R./ Haisken-DeNew, John P./Spiess, Martin/Wagner, Gert G.*, 2003: Overview of the SOEP, in: *Haisken-DeNew, John P./Frick, Joachim R.* (Hrsg.), DTC. Desktop Companion to the German Socio-Economic Panel Study (SOEP). Berlin. <http://www.diw.de/deutsch/sop/service/dtc/dtc.pdf> (03.08.05), 13-52.
- Gill, Jeff*, 1999: The Insignificance of Null Hypothesis Significance Testing, in: *Political Research Quarterly* 52, 647-674.
- Gill, Jeff*, 2002: *Bayesian Methods. A Social and Behavioral Sciences Approach*. Boca Raton et al.
- Green, Donald/Palmquist, Bradley/Schickler, Eric*, 2002: *Partisan Hearts and Minds. Political Parties and the Social Identities of Voters*. New Haven/London.
- Hu, Li-Tze/Bentler, Peter M.*, 1995: Evaluating Model Fit, in: *Hoyle, Rick H.* (Hrsg.), *Structural Equation Modeling. Concepts, Issues, and Applications*. Thousand Oaks/London/New Delhi, 76-99.
- King, Gary/Honaker, James/Joseph, Anne/Scheve, Kenneth*, 2001: Analyzing Incomplete Political Science Data. An Alternative Algorithm for Multiple Imputation, in: *American Political Science Review* 95, 49-69.
- Langeheine, Rolf/Pannekoek, Jeroen/van de Pol, Frank*, 1996: Bootstrapping Goodness-of-Fit Measures in Categorical Data Analysis, in: *Sociological Methods & Research* 24, 492-516.
- McAdams, John*, 1986: Alternatives for Dealing with Errors in the Variables. An Example Using Panel Data, in: *American Journal of Political Science* 30, 256-278.
- Pannenberg, Markus/Pischner, Rainer /Rendtel, Ulrich/Spiess, Martin/Wagner, Gert G.*, 2003: Sampling and Weighting, in: *Haisken-DeNew, John P./Frick, Joachim R.* (Hrsg.), DTC. Desktop Companion to the German Socio-Economic Panel Study (SOEP). Berlin. <http://www.diw.de/deutsch/sop/service/dtc/dtc.pdf> (03.08.05), 137-169.
- Tillie, Jean*, 1995: *Party Utility and Voting Behavior*. Amsterdam.