

Die Kontrolle von Drittvariablen bei der Messung von Segregation

Ein Vorschlag am Beispiel der familialen Assimilation von Migranten

Controlling for Independent Variables in Measures of Segregation

A Suggestion Exemplified by Family-Type Assimilation on the Part of Immigrants

Frank Kalter*

Universität Mannheim, Fakultät für Sozialwissenschaften, A 5, D-68131 Mannheim

Zusammenfassung: In diesem Beitrag wird ein Verfahren vorgeschlagen, durch das eine Drittvariablenkontrolle bei der Verwendung von Segregationsindizes möglich wird. Es verknüpft das Multinomiale Logitmodell und den Dissimilaritätsindex. Diese Verbindung des Regressions- und Segregationsansatzes erweist sich als ein äußerst nützliches Instrument zur Analyse von sozialen Ungleichheitsstrukturen. Dies wird am Beispiel der Frage nach der familialen Assimilation von Migranten in Deutschland und mit Daten der Volkszählung 1970 und des Mikrozensus 1995 ausführlich demonstriert. Isoliert verwendet, erlaubt der Regressionsansatz keine eindeutige Einschätzung des grundsätzlichen Trends, und der Segregationsansatz kommt zu einem wenig sinnvollen Gesamtbild, da er die demographischen Veränderungen nicht berücksichtigt. Die vorgeschlagene Kombination beider Verfahren ermöglicht es jedoch, strukturelle Änderungen aus den Ungleichheitsstrukturen herausrechnen und somit die Grundthese einer tendenziellen Angleichung der Familientypen zwischen Deutschen und Migranten zu belegen.

1. Hintergrund: Die ‚methodische Lücke‘ der Ungleichheitsforschung

Die empirische Analyse sozialer Ungleichheit stützt sich heutzutage überwiegend auf die Regressionsanalyse. In einem weiten Sinne sind darunter diverse Verfahren von der einfachen linearen Regression über Pfadmodelle, Logit- und Tobitmodelle bis hin zur Ereignisdatenanalyse zu verstehen (Brüderl 2000). Das prinzipielle Vorgehen besteht dabei darin, Dummy-Variablen – z. B. weiblich vs. männlich – für eine bestimmte Gruppenzugehörigkeit zu definieren und den Effekt dieser Variablen auf bestimmte abhängige Merkmale zu vergleichen, entweder über die Zeit oder zwischen verschiedenen Kontexten.

Die Popularität dieser Strategie begründet sich zum einen durch ihre ungeheure Flexibilität. Schon in der Aufzählung der einzelnen Varianten wird deutlich, dass sich sehr verschiedene Besonderheiten im Hinblick auf die abhängige Variable behandeln lassen. Zum anderen liegt eine besondere Stärke des Regressionsansatzes im Allgemeinen darin, dass sich

strukturelle Veränderungen auf sehr bequeme Weise in einer Beurteilung von Ungleichheiten berücksichtigen lassen. In Bezug auf Veränderungen in unabhängigen Randbedingungen kann dies einfach durch die Kontrolle von entsprechenden Drittvariablen geschehen. Vergleicht man etwa den Effekt der Variable ‚weiblich‘ (vs. männlich) auf das Einkommen zu zwei verschiedenen Zeitpunkten und bezieht außerdem die Schulbildung mit in die Analyse ein, so wird die Bildungsexpansion im Prinzip aus einer eventuellen Verschiebung der Einkommensungleichheit zwischen den Geschlechtern ‚herausgerechnet‘. Da ferner zeitspezifische Konstanten geschätzt werden können, ist es auch leicht, einen strukturellen Wandel in der abhängigen Variable, im Beispielfall also eine allgemeine Einkommenserhöhung, zu berücksichtigen. In vielen Anwendungsfällen würde eine mangelnde Kontrolle solcher unabhängigen oder abhängigen Strukturveränderungen zu einem verzerrten Bild führen und gerade deshalb ist die Regressionsanalyse bei der Analyse sozialer Ungleichheit oft unverzichtbar.

Vor dem Hintergrund dieser enormen Leistungsfähigkeit stellt sich die Frage, ob denn überhaupt noch andere Werkzeuge als die Regressionsverfahren notwendig sind. Und in der Tat lässt sich in bestimmten Anwendungsfällen eine zumindest kleine Schwäche dieses Ansatzes erkennen. Es kann

* Für wertvolle Hinweise zu früheren Versionen danke ich Josef Brüderl, Stephan Ganter, Nadia Granato, Johann Handl und Douglas Massey.

manchmal schwierig werden, aus dem komplexen Bild vieler Effekte bzw. Koeffizienten zu einem kurzen Urteil über die grundlegende Veränderung der Ungleichheitsstruktur zu gelangen. Dies gilt insbesondere dann, wenn die abhängige Variable ein nominalskaliertes Merkmal mit relativ vielen Ausprägungen ist. Betrachtet man beispielsweise die geschlechtsspezifische Verteilung auf verschiedene Branchen, so kann sich im zeitlichen Vergleich für manche Branchen eine Zunahme, für andere hingegen eine Abnahme der geschlechtsspezifischen Konzentration ergeben. Wie setzt man diese Befunde nun sinnvoll zu einem generellen Trend zusammen?

Gerade wenn es darum geht, die ungleiche Verteilung zweier Gruppen auf relativ viele Kategorien eines eventuell nur nominalskalierten Merkmals zu erfassen, hält die Ungleichheitsforschung jedoch noch ein anderes Set von Werkzeugen bereit: so genannte Segregationsindizes. Diese Maße sind gewissermaßen Versuche, auch komplexe Muster der Ungleichheit durch eine einzige Zahl zu beschreiben. Die Werte der Indizes lassen sich dann bequem über die Zeit oder zwischen verschiedenen Kontexten vergleichen. Mittlerweile wurden eine Vielzahl von Varianten vorgeschlagen (Massey et al. 1996), die bekanntesten und gebräuchlichsten dürften der Dissimilaritätsindex und der Gini-Index sein (Duncan/Duncan 1955, James/Taeuber 1985). Ihr Hauptanwendungsgebiet ist eben die Segregation auf dem Arbeitsmarkt (z. B. Hakim 1981, Handl 1984, Siltanen 1990, Watts 1992) oder auch die räumliche Segregation von ethnischen Gruppen (z. B. Massey/Denton 1987, 1988).

Der Segregationsansatz hat also gerade dort seine Vorzüge, wo der Regressionsansatz gewisse Schwächen aufweist – leider gilt dies jedoch auch in der umgekehrten Richtung. Ein nicht unwesentliches Problem der meisten Segregationsindizes liegt nämlich darin, dass sie – im Gegensatz zu Regressionskoeffizienten – durch strukturelle Veränderungen beeinflusst werden (Blackburn et al. 1993, Cortese et al. 1976: 631, Duncan/Duncan 1955: 216, Hakim 1993: 295, 1996: 69, Massey 1977: 587). Mittlerweile liegen jedoch Vorschläge vor, wie Strukturverschiebungen im Hinblick auf die abhängige Variable auch im Rahmen des Segregationsansatzes behandelt werden können, nämlich entweder durch so genannte ‚randverteilungsfreie‘ (margin free) Indizes (z. B.: Gibbs 1965, Charles/Grusky 1995) oder durch geeignete Dekompositionsverfahren (z. B.: Blau/Hendricks 1979, Handl 1984, Karmel/McLachlan 1988).¹

Im Gegensatz dazu wurde dem Umgang mit Strukturverschiebungen in unabhängigen Variablen im Rahmen von Segregationsindizes bisher nur wenig Aufmerksamkeit geschenkt. In vielen Anwendungsfällen muss jedoch gerade ein ‚unabhängiger‘ Wandel kontrolliert werden, um ein sinnvolles Urteil über die Veränderung von Ungleichheitsstrukturen abgeben zu können. Die vorhandenen Methoden zur empirischen Analyse von sozialer Ungleichheit weisen also eine gewisse Lücke auf, wenn man kategoriale abhängige Variablen untersuchen will und eine Einschätzung von Ungleichheitstrends ohne eine gewisse Mindestkontrolle von unabhängigen Variablen wenig sinnvoll erscheint.

Mit diesem Beitrag soll ein Vorschlag gemacht werden, diese Lücke zu schließen. Es wird sich zeigen, dass somit ein sehr nützliches Verfahren für die Analyse von Ungleichheitsstrukturen bereitgestellt werden kann. Die Lösung ist naheliegend: Wenn die jeweiligen Vor- und Nachteile der Regressionsanalyse und der Segregationsindizes komplementär sind, wäre eine geeignete Verbindung zwischen beiden Ansätzen wünschenswert. Eine solche Verbindung ist nun relativ einfach über die Verknüpfung des Multinomialen Logitmodells (MNL), dem angebrachten Regressionsverfahren bei nominalskalierten abhängigen Variablen, und dem Dissimilaritätsindex (D), dem wohl gebräuchlichsten aller Segregationsindizes, möglich. Während sich D leicht aus den Spaltenprozenten einer Kreuztabelle berechnen lässt, kann das MNL diese Spaltenprozentre reproduzieren und gleichzeitig unabhängige Variablen berücksichtigen.

Im Folgenden sollen die technischen Einzelheiten dieses Verfahrens erläutert werden. Um die Darstellung möglichst anschaulich zu gestalten, werden die hier angerissenen Argumente und Schritte in einem konkreten Anwendungsfeld aufgebaut. Es geht dabei um die Frage, ob sich in der Zeit zwischen 1970 und 1995 eine Angleichung im familialen Verhalten zwischen den in Deutschland lebenden Migranten und den einheimischen Deutschen feststellen lässt, wenn man Daten der amtlichen Statistik zu Rate zieht. Dieses Ausgangsproblem hat die Beschäftigung mit den näheren methodischen Detailfragen ursprünglich motiviert. Wie sich im Laufe der Darstellung zeigen wird, bewegt sich das Problem genau in der aufgezeigten Lücke zwischen Regressions- und Segregationsanalyse. Einerseits stehen Informationen über das familiale Verhalten in Form einer Verteilung über mehrere Familientypen zur

Strukturverschiebungen in der abhängigen Variable werden unten (6.) noch einmal aufgegriffen.

¹ Die beiden generellen Strategien im Zusammenhang mit

Verfügung, andererseits muss bei der Einschätzung der Entwicklung notwendigerweise berücksichtigt werden, dass sich die demographische Struktur von Deutschen und Migranten im betrachteten Zeitraum fundamental gewandelt hat.

Zunächst werden die Ausgangsfrage und die zur Verfügung stehenden Daten noch einmal ausführlich beschrieben (2.). Danach (3.) wird der Dissimilaritätsindex vorgestellt und diskutiert. Sein Gebrauch wirkt im Anwendungsfall jedoch das Problem mangelnder Drittvariablenkontrolle auf (4.). Deshalb wird eine Verbindung mit dem Multinomialen Logitmodell vorgeschlagen, durch die dieses Problem gelöst werden kann (5.). Außerdem bietet dieses Verfahren die zusätzliche Möglichkeit, einen strukturellen Wandel in der abhängigen Variable bei der Beurteilung einer Veränderung von Ungleichheitsstrukturen zu berücksichtigen (6.). Abschließend erfolgt dann ein Resümee im Hinblick auf die inhaltliche Fragestellung und auf die generelle Nützlichkeit einer Verbindung von Regressions- und Segregationsansatz (7.).

2. Anwendungsbeispiel: Die Assimilation im familialen Verhalten zwischen Deutschen und Migranten

Die Frage, ob sich die Lebensumstände von Einheimischen und Migranten im Laufe der Zeit einander annähern oder nicht, steht im Zentrum vieler Beiträge zur Soziologie ethnischer Minderheiten und wird in der Regel unter dem Begriff der ‚Assimilation‘ diskutiert. In einem nüchternen Sinne ist damit lediglich die Gleichheit der Verteilungen zweier Gruppen über einen bestimmten Merkmalsraum gemeint. Assimilation ist in diesem Verständnis letztlich eine Makroeigenschaft, das Gegenteil von Segregation in einem weiteren Sinne.²

In diesem Beitrag steht eine bestimmte Dimension der Assimilation,³ das familiäre Verhalten im Vordergrund. Es geht also um Eheschließungen und

Scheidungen, um Auszüge aus dem Elternhaus und Fertilität. Aus einer bestimmten theoretischen Perspektive lassen sich nun eine Reihe von guten Gründen dafür ableiten, warum man auf Dauer eine Angleichung des Verhaltens von Zugewanderten und Einheimischen erwarten würde. Ohne auf die entsprechenden Begründungen näher einzugehen zu können, lautet das Kernargument: Familiales Verhalten kann als eine Art von Produktionsstrategie aufgefasst werden, durch die erstrebenswerte Güter bereitgestellt werden. Die Eignung bestimmter Strategien ist unter anderem von gesellschaftlichen Randbedingungen abhängig. *Ceteris paribus* (!) wird eine Angleichung der gesellschaftlichen Randbedingungen (wie durch den Akt der Migration) somit zu einer tendenziellen Angleichung im familialen Verhalten führen (Kalter/Granato 2001: 9, 2002). Eine solche Sichtweise ist unter anderem in der ökonomischen Theorie der Familie (Becker et al. 1977, Becker 1981) oder im Value-of-children-Konzept (Fawcett 1972, Kagitcibasi 1982, Nauck 1997) zu finden.

Diese Assimilations-Erwartung soll nun empirisch überprüft werden. Die Daten der amtlichen Statistik stellen eine breite Basis für entsprechende Analysen dar. Hier soll auf eine 1 %-Stichprobe der Volkszählung 1970 und auf die 70 %-Unterstichprobe des Mikrozensus 1995 (ZUMA-Files) zurückgegriffen werden, der selbst wiederum eine 1 %-Stichprobe der Bevölkerung darstellt (Lüttinger/Riede 1997).⁴ Beide Datensätze enthalten relativ viele vergleichbare Variablen, da die zugrunde liegenden Fragen weitgehend identisch sind. Unter anderem lässt sich jeweils die gleiche ‚Familientypologie‘ entwickeln. Damit ist die Art der Familie gemeint, in der eine Befragungsperson zum Zeitpunkt der jeweiligen Erhebung wohnt. Insgesamt lassen sich neun verschiedene Typen unterscheiden. Die Fallzahlen für die einzelnen Kategorien und die Randverteilung sind für die beiden Jahre und getrennt für Personen mit und ohne deutschen Pass in Tabelle 1 zusammengestellt.

Man erkennt, dass sich im betrachteten Vierteljahrhundert innerhalb jeder Gruppe deutliche Veränderungen abzeichnen. Bei den Deutschen ist die Lebensform der klassischen Kernfamilie deutlich zurückgegangen. Dies erkennt man in Tabelle 1 sowohl am Typ 2 als auch am Typ 9. Deutlich zugenommen haben hingegen die verheirateten Paare ohne Kinder (Typ 1) und die alleinlebenden Ledigen.

² Einerseits impliziert der Begriff damit, dass es jeweils auch eine Binnengruppenvarianz geben kann, andererseits ist in keiner Weise vorweggenommen, dass eine Konvergenz der Verteilungen nur in einer Richtung stattfinden kann (vgl. Alba/Nee 1997). Schon aus diesen Gründen hat ‚Assimilation‘ nichts mit dem Begriff der ‚Leitkultur‘ zu tun.

³ Seit dem Werk von Milton Gordon (1964) wird Assimilation in der Regel als multidimensionales Konzept begriffen, d. h. die Angleichung kann im Hinblick auf verschiedene Aspekte des Lebens durchaus unterschiedlich weit voranschreiten.

⁴ Befragte im Mikrozensus 1995, die im Gebiet der ehemaligen DDR wohnen, werden in den Analysen nicht berücksichtigt.

Tabelle 1 Familientypen von Deutschen und Migranten 1970 und 1995 (absolute Fallzahlen und Spaltenprozentage)

	1970		1995	
	Deutsche	Migranten	Deutsche	Migranten
1 verheiratetes Paar, ohne Kind	101675 17,5 %	3565 17,6 %	88565 23,4 %	4159 13,3 %
2 verheiratetes Paar, mit Kind(ern)	182270 31,3 %	5134 25,4 %	94766 25,1 %	11010 35,3 %
3 geschieden oder verwitwet, ohne Kind	48391 8,3 %	617 3,0 %	39465 10,4 %	1000 3,2 %
4 geschieden oder verwitwet, mit Kind(ern)	12544 2,2 %	138 0,7 %	7996 2,1 %	486 1,6 %
5 nie verheiratet, mit Kind(ern)	1331 0,2 %	49 0,2 %	2276 0,6 %	159 0,5 %
6 verheiratet, getrennt lebend, ohne Kind	5725 1,0 %	2012 9,9 %	3729 1,0 %	763 2,4 %
7 verheiratet, getrennt lebend, mit Kind(ern)	1474 0,3 %	174 0,9 %	1378 0,4 %	192 0,6 %
8 nie verheiratet, ohne Kind, nicht bei den Eltern lebend	27137 4,7 %	2919 14,4 %	41906 11,1 %	265 28,5 %
9 nie verheiratet, ohne Kind, bei den Eltern lebend	201393 34,6 %	5641 27,9 %	97827 25,9 %	10741 34,5 %
insgesamt	581940	20249	377908	31162

Quelle: Volkszählung 1970 (1 %-Stichprobe), Mikrozensus 1995 (ZUMA-File), eigene Berechnungen

gen (Typ 8). Bei den Migranten verhält es sich bei den gerade genannten Kategorien jeweils genau umgekehrt. Bei den Typen 3 und 5 wiederum gehen die Trends für beide Gruppen in die gleiche Richtung.

Was lässt sich nun über die Veränderung der Ungleichheit *zwischen* den Gruppen sagen? Nimmt man die Prozentsatzdifferenz als Indikator, so hat sich die Ungleichheit im Hinblick auf einige Kategorien vergrößert (Typ 1, 2, 3, 5 und 9). Bei anderen Kategorien (z. B. Typ 4, 6, 7 und 8) sind die Abstände dagegen kleiner geworden. Ohne ein Maß, das die Gleichheit in Bezug auf alle Kategorien umfasst, lässt sich also nur schwer etwas über den allgemeinen Trend aussagen. Segregationsindizes stellen gerade solche Maße dar; sie sind somit im vorliegenden Anwendungsfall ein nahe liegendes Analysemittel.

3. Ausgangspunkt: Der Dissimilaritätsindex zur Messung von Segregation

Spätestens seit dem klassischen Aufsatz „A methodological analysis of segregation indexes“ von Otis Dudley Duncan und Beverly Duncan (1955) wird intensiv und sehr kontrovers diskutiert, welche

Maßzahl zur Kennzeichnung von Segregationstendenzen zu verwenden ist. Die Durchsicht der unzähligen Beiträge zu dieser ‚Indexdebatte‘ ist jedoch eher verwirrend denn erhellend. Von einem tendenziellen Konsens über das geeignetste Segregationsmaß kann nicht die Rede sein, vor allem, wenn es um den intertemporalen und interkontextuellen Vergleich von Segregationsstrukturen geht. Immer wieder werden vorhandene Indizes (oft zurecht) kritisiert, neue Indizes vorgeschlagen, die dann (nicht weniger zurecht) wieder kritisiert bzw. verworfen werden.⁵ Die Vielzahl von Vorschlägen ist nicht zuletzt darauf zurückzuführen, dass kein klarer Kriterienkatalog für eine Beurteilung der Maßzahlen vorliegt, der sich unmittelbar aus einer umfassenden Definition der Segregation ergibt (James/Taeuber 1985: 2), bzw. dass kein Konsens über solche Kriterien besteht. Vor dem Hintergrund oftmals sehr spezifischer Anwendungen und Fragestellungen ist es auch nicht verwunderlich, dass

⁵ So berichten Douglas S. Massey und Nancy A. Denton (1987: 805), dass sie der Literatur nicht weniger als 19 Maßzahlen entnommen haben – eine Zahl, die man heute noch um einiges nach oben korrigieren müsste. Berücksichtigt man, dass einige dieser Maßzahlen – z. B. der Atkinson-Index – parametrisch sind, gibt es im Prinzip unendlich viele.

es letztlich keinen Index gibt, der für alle Belange die beste Lösung darstellt (Lieberman 1980: 253, Massey/Denton 1988: 283).

Trotz aller Kontroversen und Unübersichtlichkeiten dürfte jedoch weithin Einigkeit darin bestehen, dass dem Dissimilaritätsindex (D) unter allen Maßzahlen eine herausragende Rolle zukommt. Interessanterweise wird dieser Index in der Segregationsforschung einerseits am ausgiebigsten und heftigsten kritisiert, andererseits aber auch am meisten verwendet. Letzteres ist bei näherem Hinsehen auf mindestens vier Gründe zurückzuführen. Erstens kann der Dissimilaritätsindex sehr einfach und eingängig interpretiert werden. Zweitens weist er eine Reihe von Eigenschaften auf, die für eine Maßzahl der Segregation wünschenswert sind. Drittens ist bislang keine echte Alternative in Sicht, und viertens entsteht eine Art von Selbstverstärkungsmechanismus dadurch, dass es sich schon aus Gründen der Vergleichbarkeit anbietet, den bisher verbreitetsten Index zu verwenden, zumindest solange der gerade genannte dritte Grund Geltung besitzt.

Es ist also durchaus nahe liegend, den Dissimilaritätsindex D zur Beschreibung der Gleichheit von Verteilungen zu verwenden. Im Anwendungsfall ist die Verwendung auch inhaltlich sinnvoll, denn die Maßzahl ist eine sehr direkte Umsetzung des vorgeschlagenen Assimilationsbegriffes (siehe oben, 2.). D ist folgendermaßen definiert:

$$D = \frac{1}{2} \sum_{k=1}^J \left| \frac{A_k}{A} - \frac{B_k}{B} \right|, \quad (1)$$

wobei J die Anzahl der Kategorien, A die Anzahl der Personen aus Gruppe A, B die Anzahl der Personen aus Gruppe B, A_k die Anzahl der Personen aus Gruppe A in Kategorie k und B_k die Anzahl der Personen aus Gruppe B in Kategorie k bezeichnet.

Der Wert von D liegt stets im Intervall [0, 1]. Der maximale Wert von 1 ergibt sich, wenn in keiner Kategorie sowohl Mitglieder von A als auch von B sind, wenn die Gruppen A und B also disjunkt über die Kategorien verteilt sind. Der minimale Wert von 0 ergibt sich, wenn für jede Kategorie der Anteil von A-Mitgliedern an Gesamt-A gleich dem Anteil von B-Mitgliedern an Gesamt-B ist, wenn also die Gruppen A und B anteilmäßig exakt gleich über die einzelnen Kategorien verteilt sind. Die geläufigste Interpretation ist, dass D gerade den Anteil derjenigen aus einer der beiden Gruppen angibt, die ihre Kategorie wechseln müssten, damit es zu einer gleichen Verteilung bei-

der Gruppen über die Kategorien käme (Duncan/Duncan 1955: 211; zum Nachweis: Cortese et al. 1976: 634).⁶

Neben der äußerst einfachen Interpretation besitzt der Dissimilaritätsindex noch weitere wünschenswerte Eigenschaften (James/Taeuber 1985: 11–19). Hier sind vor allem zwei davon von Bedeutung: a) D ist „Größen-invariant“, d. h. der Wert von D ändert sich nicht, wenn man jede Zelle der zugrunde liegenden Kreuztabelle mit der gleichen Konstanten multipliziert: b) D erfüllt auch das Prinzip der Kompositions-Invarianz. Hierunter verstehen James und Taeuber (1985: 15f.), dass eine proportionale Veränderung der Größe einer Gruppe, die die Verteilung dieser Gruppe auf die Kategorien unverändert lässt, keine Auswirkungen auf das Segregationsmaß besitzt. Mit anderen Worten: Multipliziert man nur eine Spalte der zugrunde liegenden Kreuztabelle mit einer Konstanten, bleibt der Wert von D ebenfalls unverändert.

Der Dissimilaritätsindex ist aufgrund seiner Definition, Interpretation und Eigenschaften somit ein geeignetes Maß zur Messung von Assimilation. Wendet man nun die Definition von D auf die Zahlen in Tabelle 1 an, so ergeben sich für die beiden Jahre folgende Werte:

$$D_{1970} = .195$$

$$D_{1995} = .206$$

Der Dissimilaritätsindex zeigt also an, dass die familiäre Ungleichheit zwischen Migranten und Deutschen von 1970 bis 1995 leicht (um .011) zugenommen hat. Wegen der Größen-Invarianz und der Kompositions-Invarianz von D ist dieser Befund unbeeinflusst von Veränderungen in der Bevölkerungs- bzw. Stichprobengröße und von Veränderungen im Migrantenanteil.

4. Problem: Der Einfluss demographischer Veränderungen auf die familiäre Ungleichheit

Das Ergebnis des letzten Abschnittes widerspricht der grundsätzlichen theoretischen Erwartung offensichtlich. Bevor die Assimilationsthese jedoch vorschnell verworfen wird, sollte noch einmal bedacht werden, dass sie auf einem ‚Ceteris-paribus-Argument‘ beruht. Somit könnte eine entscheidende Veränderung von wichtigen Randbedingungen den

⁶ Eine weitere Interpretation ist, dass D den maximalen Abstand der Lorenzkurve zur ersten Hauptdiagonalen bezeichnet (James/Taeuber 1985: 6).

Tabelle 2 Alters- und Geschlechtsstruktur von Deutschen und Migranten 1970 und 1995 (Spaltenprozent)

	Deutsche		Migranten	
	1970	1995	1970	1995
Männer unter 20 Jahren	15.0 %	10.2 %	11.4 %	16.1 %
Männer, 20 bis 29 Jahre	6.7 %	6.9 %	17.6 %	10.4 %
Männer, 30 bis 39 Jahre	7.2 %	7.9 %	21.0 %	9.3 %
Männer, 40 bis 49 Jahre	5.8 %	6.4 %	8.8 %	7.3 %
Männer, 50 bis 59 Jahre	4.4 %	7.4 %	2.8 %	6.6 %
Männer, 60 Jahre und älter	8.1 %	9.3 %	1.6 %	2.7 %
Frauen unter 20 Jahren	14.3 %	9.7 %	11.3 %	14.6 %
Frauen, 20 bis 29 Jahre	6.3 %	6.7 %	11.3 %	10.4 %
Frauen, 30 bis 39 Jahre	7.0 %	7.7 %	7.4 %	8.1 %
Frauen, 40 bis 49 Jahre	6.9 %	6.4 %	3.7 %	7.7 %
Frauen, 50 bis 59 Jahre	6.1 %	7.5 %	1.4 %	4.5 %
Frauen, 60 Jahre und älter	12.3 %	13.9 %	1.5 %	2.1 %
D: Deutsche 1970 vs. Migranten 1970		0,333		
Deutsche 1995 vs. Migranten 1995		0,221		
Deutsche 1970 vs. Deutsche 1995		0,099		
Migranten 1970 vs. Migranten 1995		0,213		

Quelle: Volkszählung 1970 (1 %-Stichprobe), Mikrozensus 1995 (ZUMA-File), eigene Berechnungen

grundsätzlichen Assimilationstrend auch nur konkurrenzieren haben. Eine nahe liegende Vermutung wäre beispielsweise, dass demographische Veränderungen mit im Spiel sind. Während sich die Zuwanderung in der Anwerbephase vorwiegend auf männliche Personen jüngeren und mittleren Alters konzentrierte, hat sich die demographische Struktur der Migrantenpopulation in Deutschland mittlerweile fundamental geändert. Gleichzeitig ist die deutsche Bevölkerung bekanntermaßen tendenziell ‚gealtert‘. Beide Entwicklungen werden deutlich, wenn man die Verteilung auf Alters- und Geschlechtsgruppen in den beiden bisher betrachteten Jahren vergleicht (Tabelle 2).

Im Jahr 1970 unterscheiden sich die Alters- und Geschlechtsverteilungen zwischen Deutschen und Migranten äußerst stark. Letztere sind vor allem in den Kategorien ‚Männer zwischen 20 und 29‘, ‚Männer zwischen 30 und 39‘ und ‚Frauen zwischen 20 und 29‘ deutlich überrepräsentiert. Wenn man die Alters-Geschlechts-Ungleichheit zwischen Deutschen und Migranten durch den Dissimilaritätsindex ausdrückt, ergibt sich 1970 ein Wert von .333. Diese Diskrepanz hat sich 1995 etwas verrin-

gert, ist aber mit einem Wert von .221 immer noch erheblich. Wie man in der Tabelle erkennen kann, haben sich beide Gruppen im Laufe der Zeit verändert. Im Vergleich von Deutschen 1970 mit Deutschen 1995 macht sich eine leichte ($D = .099$), im Vergleich von Migranten 1970 und Migranten 1995 eine deutliche ($D = .213$) Verschiebung bemerkbar. 1995 sind die Ausländer vor allem in den jüngeren Altersgruppen bei beiden Geschlechtern überrepräsentiert, während sie in den älteren Altersgruppen, vor allem unter den älteren Frauen, deutlich weniger vertreten sind.

Die Wahl des Familientyps hängt nun trivialerweise eng mit Alter und Geschlecht zusammen. Um die ungleiche Verteilung über beide Variablen in der Beurteilung der familialen Assimilation zu kontrollieren, kann man die entsprechenden Dissimilaritätsindizes für 1970 und 1995 getrennt nach Alters×Geschlechtsgruppen ermitteln. Abbildung 1 präsentiert die Werte, die sich bei einer entsprechenden Berechnung ergeben, graphisch.

Im Vergleich zu den oben ermittelten Gesamtwerten ergibt sich nun ein völlig anderes Bild und demnach auch ein anderer Schluss im Hinblick auf die

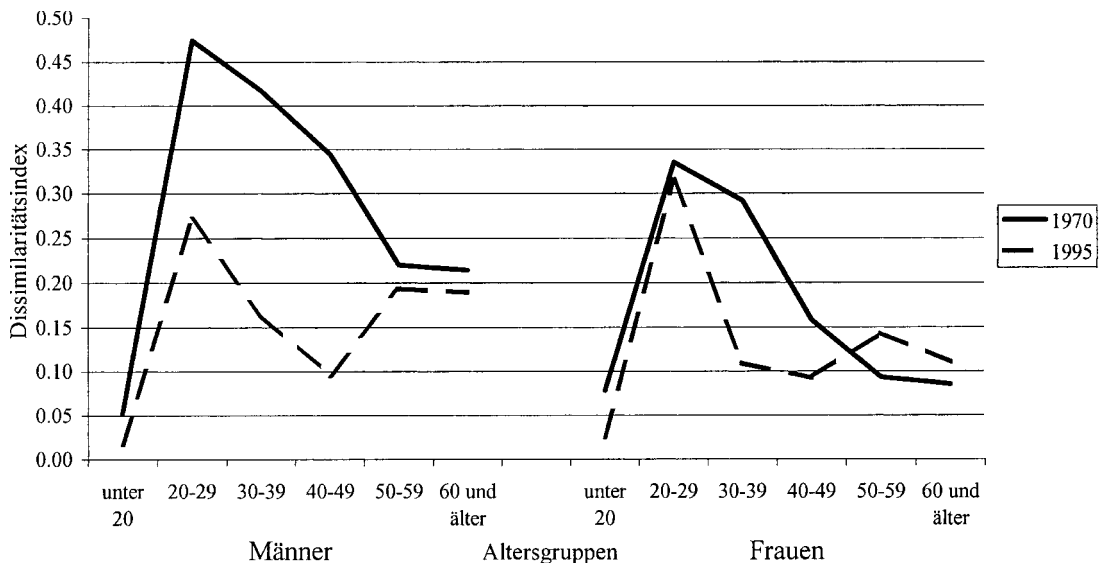


Abb. 1 Dissimilarität der Familientypen 1970 und 1995 getrennt für Alters- und Geschlechtsgruppen

familiale Assimilation. Mit Ausnahme der beiden höchsten Altersgruppen der Frauen ist D 1970 für alle Alters×Geschlechtsgruppen höher als 1995, teilweise sogar beträchtlich. Der familiäre Assimilationstrend in den betrachteten 25 Jahren ist bei den Männern mittleren Alters am ausgeprägtesten.

Abbildung 1 gibt eine korrektere Auskunft über den Trend zur familialen Assimilation, da die weitreichenden strukturellen Veränderungen in der Alters- und Geschlechtskomposition kontrolliert werden. Ein Nachteil dieser Darstellung ist jedoch, dass die enthaltene Information über die Ungleichheit sehr detailliert ist. Man könnte nun darüber nachdenken, durch angemessene Gewichtungen der Einzel-Indizes für die Alters×Geschlechtsgruppen wieder zu einem Gesamtindex, und damit zu einem einfacheren Bild der Ungleichheit, zu kommen. Dies wäre prinzipiell möglich, allerdings stieße ein solches Verfahren einer getrennten Berechnung für einzelne Subgruppen schnell an seine Grenzen, wenn man weitere unabhängige Variablen oder gar unabhängige Variablen in kontinuierlicher Form berücksichtigen wollte.

Neben einem solchen Separations- und Gewichtsungsverfahren bietet sich jedoch noch eine ‚elegantere‘ und umfassendere Lösung an. Sie beruht darauf, die Regressionsanalyse in der Analyse mitzuverwenden, und wird im folgenden Abschnitt erläutert.

5. Lösung: Die Verbindung des Dissimilaritätsindex mit dem Multinomialen Logitmodell

Eine Kontrolle von unabhängigen Variablen wie Alter und Geschlecht wird möglich, wenn man das Multinomiale Logitmodell (MNL) in die Berechnung des Dissimilaritätsindex ‚einschaltet‘. Die Idee des Vorgehens soll im Folgenden erläutert werden. Zunächst wird die grundsätzliche Verbindung zwischen beiden Analysewerkzeugen hergestellt; dann wird gezeigt, wie sich aus der Einbeziehung von Drittvariablen in das MNL ein korrigierter Wert von D ermitteln lässt.⁷

5.1 Die Berechnung von D aus den Koeffizienten eines MNL

Das Multinomiale Logitmodell ist die Erweiterung der logistischen Regression für abhängige Variablen mit mehr als zwei Kategorien. Gibt es im Allgemeinen J Kategorien, so lässt sich das MNL durch folgende Formel (2) definieren (z. B. Long 1997: 152):

⁷ Spriggs/Williams (1996) verwenden ebenfalls das MNL im Zusammenhang mit der Messung von Segregation. Allerdings schlagen sie einen neuen Index vor, während hier lediglich eine Korrektur von D vorgenommen wird.

$$\Pr(y_i = j | x_i) = \frac{\exp(x_i \beta_j)}{\sum_{k=1}^J \exp(x_i \beta_k)} \quad (2)$$

Dabei bezeichnet y_i den Wert der abhängigen Variable für Beobachtung i , x_i einen Vektor, der die Ausprägungen von m unabhängigen Variablen für Beobachtung i enthält, und β_k einen Vektor von $m+1$ Parametern ($\beta_{0k}, \beta_{1k}, \dots, \beta_{mk}$) für jedes $k = 1, \dots, J$. Die Wahrscheinlichkeit (\Pr), dass eine Beobachtung mit der Konstellation von unabhängigen Variablen x_i auf der abhängigen Variable den Wert $y_i = j$ aufweist, ist damit gleich der exponierten Linearkombination von unabhängigen Variablen und den Parametern der jeweiligen Kategorien, geteilt durch die Summe der entsprechenden exponierten Linearkombinationen über alle Kategorien. Um die Parameter eindeutig zu identifizieren, bestimmt man gewöhnlich eine Referenzkategorie und setzt deren Parameter alle gleich Null.

Eine Beurteilung der Ungleichheit zweier Gruppen kann nun dadurch vorgenommen werden, dass die Gruppenzugehörigkeit als unabhängige Variable eingeführt wird. Im Anwendungsfall kann beispielsweise mit den Daten, die Tabelle 1 zugrunde liegen, jeweils ein MNLM für 1970 und 1995 geschätzt werden. Die abhängige Variable bilden die neun Kategorien der Familientypologie (Referenzkategorie = 9), die einzige unabhängige Variable x_1 bildet das Merkmal ‚Migrant‘ ($x_1 = 0$ für ‚deutsch‘, $x_1 = 1$ für ‚nicht-deutsch‘). Die geschätzten Koeffizienten sind in Tabelle 3 zu finden.

Gäbe es in der Verteilung auf die Familientypen keinerlei Unterschiede zwischen den Deutschen und den Migranten, so wären die β_1 -Koeffizienten alle gleich Null, d. h. das Merkmal ‚Migrant‘ hätte keinen Effekt auf die Zuordnung zu den einzelnen Kategorien.⁸ In Tabelle 3 ist jedoch erkennbar, dass so-

wohl 1970 als auch 1995 zum Teil erhebliche Abweichungen von Null auftreten. Will man nun beurteilen, ob sich die Ungleichheit zwischen den betrachteten Jahren insgesamt vergrößert oder verkleinert hat, so werden die Grenzen eines einfachen Blickes auf die Koeffizienten deutlich: Einige Effekte des Merkmals ‚Migrant‘ haben sich von 1970 bis 1995 betragsmäßig vergrößert (für die Kategorien 1, 2, 3, 5), andere hingegen haben sich verkleinert (Kategorien 4, 6, 7, 8). Es tauchen also die gleichen Schwierigkeiten wie beim Blick auf die Prozentsatzdifferenzen auf (siehe oben, Abschnitt 2).

Um die *Gesamt*-Veränderung der Ungleichheit einzuschätzen, müssen die Koeffizienten der MNL-Modelle somit in irgendeiner sinnvollen Weise zu einer Maßzahl zusammengefasst werden. Es lässt sich nun zeigen, dass gerade der Dissimilaritätsindex eine solche Vorschrift liefern kann, denn D lässt sich auch aus den MNLM-Koeffizienten bestimmen: Die Berechnung von D (siehe Formel (1)) stützt sich letztlich auf die Spaltenprozentage einer Kreuztabelle, in der die Kategorien die Zeilen und die betrachteten Gruppen die Spalten bilden. Interpretiert man diese Spaltenprozentage als bedingte Wahrscheinlichkeiten, so lautet die Berechnungsvorschrift für D :

$$D = \frac{1}{2} \sum_{k=1}^J |\Pr(k|A) - \Pr(k|B)|, \quad (3)$$

wobei $\Pr(k|A)$ bzw. $\Pr(k|B)$ die Wahrscheinlichkeit bezeichnet, dass sich ein Mitglied von Gruppe A bzw. B in Kategorie k befindet. Diese bedingten Wahrscheinlichkeiten lassen sich nun – wie Formel (2) zeigt – durch die Regressionskoeffizienten eines MNLM ausdrücken.

Somit kann man eine naheliegende Verbindung zwischen den Formeln (2) und (3) – und damit zwischen D und dem MNML – herstellen. Allgemein gilt: Man schätze ein Modell, das nur eine einzige unabhängige Variable enthält, nämlich eine Dummy-Variablen für die Gruppenmitgliedschaft x_1 . Es gelte also $x_{1i} = 0$, wenn die Beobachtung i der Gruppe A angehört, und $x_{1i} = 1$, wenn die Beobachtung i der Gruppe B angehört.⁹ Ist die J -te Kategorie die Referenzkategorie, dann liefert die Modellschätzung $J-1$ Konstanten, d. h. die Koeffizienten $\beta_{01}, \beta_{02}, \dots, \beta_{0(J-1)}$, und $J-1$ Koeffizienten für die Grup-

⁸ Es ist eine sehr nützliche Eigenschaft des MNLM (wie auch der logistischen Regression), dass die exponierten Koeffizienten von Variablen als ‚odds‘ bzw. ‚odds-ratios‘ interpretiert werden können (z. B. Hosmer/Lemeshow 1989: 220–225, Long 1997: 154). Als ‚odds‘ bezeichnet man dabei die relativen Chancen zu einer gegebenen Kategorie vs. der Referenzkategorie zu gehören. Die bedingten ‚odds‘ der Referenzgruppe (in unserem Falle ‚Deutsche‘) ergeben sich im MNLM durch die exponierten Konstanten. Die bedingten ‚odds‘ der alternativen Gruppe (in unserem Falle ‚Migranten‘) ergeben sich durch die ‚odds‘ der Referenzgruppe multipliziert mit dem exponierten β_1 -Koeffizienten der Gruppenzugehörigkeitsvariable, den so genannten ‚odds-ratios‘. Die Exponierung der β_1 -Koeffizienten in Tabelle 3 kennzeichnet also jeweils die relativen Chancen der Migranten im Verhältnis zu den relativen

Chancen der Deutschen, einer bestimmten Kategorie (vs. der Referenzkategorie) anzugehören.

⁹ Damit gilt nach den Konventionen in Formel (2) und (3): $\Pr(k|A) = \Pr(y_i = k | x_1 = 0)$ und $\Pr(k|B) = \Pr(y_i = k | x_1 = 1)$

Tabelle 3 Koeffizienten der Multinomialen Logitmodelle für 1970 und 1995

K	1970		1995	
	Konstante β_{0k}	„Migranten“ β_{1k}	Konstante β_{0k}	„Migranten“ β_{1k}
1 verheiratetes Paar, ohne Kind	-0.68	0.22	-0.10	-0.85
2 verheiratetes Paar, mit Kind(ern)	-0.10	0.01	-0.03	0.06
3 geschieden oder verwitwet, ohne Kind	-1.43	-0.79	-0.91	-1.47
4 geschieden oder verwitwet, mit Kind(ern)	-2.78	-0.93	-2.50	-0.59
5 nie verheiratet, mit Kind(ern)	-5.02	0.27	-3.76	-0.45
6 verheiratet, getrennt lebend, ohne Kind	-3.56	2.53	-3.27	0.62
7 verheiratet, getrennt lebend, mit Kind(ern)	-4.92	1.44	-4.26	0.24
8 nie verheiratet, ohne Kind, nicht bei den Eltern lebend	-2.00	1.35	-0.85	-0.55
9 nie verheiratet, ohne Kind, bei den Eltern lebend	0	0	0	0
D entsprechend Formel (3)	0,195		0,206	

penmitgliedschaftsvariable $\beta_{11}, \beta_{12}, \dots, \beta_{1(j-1)}$. Es lässt sich unter diesen Voraussetzungen aus (2) und (3) nun ableiten, dass

$$D = \frac{1}{2} \sum_{k=1}^J \left| \frac{\exp(\beta_{0k})}{\sum_{i=1}^J \exp(\beta_{0i})} - \frac{\exp(\beta_{0k} + \beta_{1k})}{\sum_{i=1}^J \exp(\beta_{0i} + \beta_{1i})} \right|, \quad (4)$$

wobei für die beiden Koeffizienten der Referenzkategorie gilt: $\beta_{0j} = \beta_{1j} = 0$.

Rechnet man die Koeffizienten in Tabelle 3 entsprechend der Anweisung (4) in den Dissimilaritätsindex um, so erhält man genau die Werte, die man auch bei der Standardberechnung erhält: .195 für 1970 und .206 für 1995.

5.2 Der korrigierte Index unter Kontrolle von Drittvariablen

Bislang wurde durch den formalen Aufwand einer Verknüpfung von D und dem MNLM noch nichts gewonnen. Es ist aber nur noch ein kleiner Schritt, um die Fruchtbarkeit der Verbindung deutlich zu machen. Innerhalb des MNLM können nun nämlich noch weitere unabhängige Variablen berücksichtigt werden. Nehmen wir an, dass außer der Gruppenzugehörigkeit x_1 noch die unabhängigen Variablen x_2, \dots, x_m in das Modell eingehen und man insgesamt $m+1$ Parameter ($\beta'_{0k}, \beta'_{1k}, \dots, \beta'_{mk}$) für alle $k = 1, \dots, J$ erhält. Die Koeffizienten β'_{1k} werden dann „unter Kontrolle der unabhängigen Variablen x_2, \dots, x_m “ berechnet, d.h. sie sind nun interpretierbar als die Effekte der Gruppenzuge-

hörigkeit auf die relativen Zugangschancen zur Kategorie k (vs. der Referenzkategorie), wenn man unterstellt, dass die Ausprägungen der Variablen x_2, \dots, x_m gleich sind.¹⁰

Diese „kontrollierten Koeffizienten“ bilden nun den einfachen Ausgangspunkt einer Drittvariablenkontrolle bei der Berechnung von D. Der Vorschlag, einen *korrigierten Dissimilaritätsindex D' unter Kontrolle der unabhängigen Variablen x_2, \dots, x_m* zu berechnen, wäre somit, folgende Formel (5) zur Berechnung zu verwenden:

$$D' = \frac{1}{2} \sum_{k=1}^J \left| \frac{\exp(\beta_{0k})}{\sum_{i=1}^J \exp(\beta_{0i})} - \frac{\exp(\beta_{0k} + \beta'_{1k})}{\sum_{i=1}^J \exp(\beta_{0i} + \beta'_{1i})} \right|, \quad (5)$$

Dabei sind (für alle $k = 1, \dots, J$) β_{0k} die Konstanten eines MNLM, das nur die Gruppenzugehörigkeit x_1 als unabhängige Variable enthält,¹¹ und β'_{1k} die Koeffizienten der Gruppenzugehörigkeit x_1 in einem

¹⁰ Der exponentierte Wert von β'_{1k} ist dann immer noch als ‚odds-ratio‘ interpretierbar, nur handelt es sich jetzt nicht mehr um das generelle ‚odds-ratio‘, sondern um das ‚odds-ratio‘ unter Konstanthaltung der unabhängigen Variablen x_2, \dots, x_m (Long 1997: 154). Das heißt, das Verhältnis der relativen Chancen beider Gruppen wird nun für den Fall ausgedrückt, dass die Ausprägungen der Variablen x_2, \dots, x_m gleich sind.

¹¹ Es erscheint nicht sinnvoll, die Konstanten β'_{0k} zu verwenden, die dem Modell mit weiteren unabhängigen Variablen x_2, \dots, x_m entstammen, da man dann die Ungleichheit in Bezug auf Personen bestimmen würde, die auf allen unabhängigen Variablen x_2, \dots, x_m den Wert Null aufweisen. Demgegenüber spiegeln die ‚ursprünglichen‘ β_{0k} -Wer-

Tabelle 4 Koeffizienten der Gruppenzugehörigkeit x_1 in einem MNLM unter Kontrolle der Alters×Geschlechtsgruppen

K	1970 β'_{1k}	1995 β'_{1k}
1 verheiratetes Paar, ohne Kind	2.16	1.22
2 verheiratetes Paar, mit Kind(ern)	1.09	1.46
3 geschieden oder verwitwet, ohne Kind	1.92	1.01
4 geschieden oder verwitwet, mit Kind(ern)	0.96	0.98
5 nie verheiratet, mit Kind(ern)	1.61	0.32
6 verheiratet, getrennt lebend, ohne Kind	3.41	2.07
7 verheiratet, getrennt lebend, mit Kind(ern)	2.52	1.32
8 nie verheiratet, ohne Kind, nicht b. d. Eltern lebend	2.22	0.24
9 nie verheiratet, ohne Kind, bei den Eltern lebend	0	0

Modell, das außerdem noch die unabhängigen Variablen x_2, \dots, x_m enthält.

Die vorgeschlagene Korrektur des Dissimilaritätsindex soll nun auf den Beispielfall angewendet werden. Dazu werden wiederum MNL-Modelle für 1970 und 1995 geschätzt, die die Kategorien des Familientyps als abhängige Variable enthalten. Neben der Gruppenzugehörigkeit x_1 wird nun aber auch die Alters×Geschlechtsgruppe in Form von 11 zusätzlichen Dummy-Variablen x_2, \dots, x_{12} (Referenzgruppe: „Männer unter 20 Jahren“) berücksichtigt. Tabelle 4 zeigt die geschätzten Koeffizienten der Variable x_1 in diesen Modellen.

Wendet man die Berechnungsvorschrift (5) an, so ergeben sich für die beiden betrachteten Jahre die Werte:

$$D'_{1970} = .362$$

$$D'_{1995} = .226$$

Im Vergleich zum herkömmlichen Dissimilaritätsindex erhält man nun eine völlig andere Einschätzung im Hinblick auf die familiäre Assimilation. Die Ungleichheit zwischen Deutschen und Ausländern hat in den betrachteten Jahren erheblich abgenommen, wenn man die demographische Komposition der beiden Gruppen berücksichtigt. Die Zahlen der korrigierten Indizes spiegeln somit das Bild wider, das sich auch schon in der graphischen Betrachtung in Abbildung 1 zeigt.

6. Zugabe: Die Kontrolle des Wandels in der abhängigen Variable

Im ersten einleitenden Abschnitt wurde bereits angesprochen, dass Strukturveränderungen in unabhängigen Variablen nicht das einzige Problem bei der Verwendung von Segregationsindizes sind. Die meisten Indizes werden auch durch Strukturveränderungen der abhängigen Variable beeinflusst. Dies gilt insbesondere für den Dissimilaritätsindex. Wenn sich die Größe einer Kategorie im Verhältnis zu den anderen Kategorien ändert, ändert sich in der Regel auch der Indexwert, selbst wenn die relative Verteilung auf beide Gruppen innerhalb dieser Kategorie gleich bleiben. Technisch gesprochen ist D nicht invariant gegenüber Multiplikationen von Kategorien (Zeilen der Kreuztabelle) mit einer Konstanten.¹²

Im Prinzip können zwei verschiedene Strategien unterschieden werden, um mit dieser Schwäche von D umzugehen. Zum einen kann man für dieses Problem weniger ‚anfällige‘ bzw. gänzlich „unanfällige“ Indizes vorschlagen. In diesem Zusammenhang sind insbesondere die randverteilungsfreien („margin free“) Segregationsmaße zu nennen, beispielsweise der bekannte größen-standardisierte Dissimilaritätsindex (Gibbs 1965) oder das auf der log-linearen Analyse beruhende Verfahren von Charles und Grusky (1995). Zu beachten ist allerdings, dass mit solchen randverteilungsfreien Maßen Veränderungen in den Größenverhältnissen einfach ausgeklammert werden, was unter einer bestimmten Perspektive von Ungleichheitsstrukturen oft erwünscht ist. Ist die Berücksichtigung von Grö-

te gewissermaßen die mittlere Konstellation der Kovariaten x_2, \dots, x_m wider.

¹² Letztlich ist die Sensibilität von D für Veränderungen in der Größe einzelner Kategorien ein Hauptgrund dafür, warum so viele alternative Indizes existieren.

ßenverhältnissen jedoch – wie im Anwendungsfall hier – theoretisch sinnvoll bzw. im inhaltlichen Verständnis von Ungleichheit impliziert, so bietet sich eine zweite Strategie an. Man kann den Dissimilaritätsindex (und damit seine Vorzüge) grundsätzlich beibehalten und versuchen, die Effekte durch Größenverschiebungen der Kategorien von den Effekten durch eine Verschiebung der relativen Verteilung innerhalb der Kategorien zu trennen, denn D misst beide Effekte gewissermaßen zusammen. Mittlerweile existieren eine Reihe von Vorschlägen, wie man entsprechende Dekompositionen vornehmen kann (z. B. Blau/Hendricks 1979, Handl 1984, Karmel/McLachlan 1988, Watts 1992). Diese Verfahren beruhen in der Regel auf einer Anpassung einer Kreuztabelle an eine gegebene Randverteilung durch das Deming-Stephan-Verfahren bzw. auf dem log-linearen Modell.

Eine Trennung der entsprechenden Effekte ist jedoch auch durch die Verknüpfung von D mit dem MNLM leicht möglich. Wie bei den anderen zitierten Vorschlägen besteht die Grundidee der Dekomposition darin, zu simulieren, welche Dissimilarität vorzufinden wäre, wenn sich über die Zeit nur die Größenveränderungen in den Kategorien ergeben hätten, jedoch keine Verschiebung in den relativen Zugangsstrukturen der Gruppen für die Kategorien (Handl 1984). Mit Blick auf die Berechnung von D durch die Koeffizienten von MNL-Modellen (Formeln 4 bzw. 5) entspräche dies einer Verwendung der Konstanten des Zieljahres (im Anwendungsfall: β_{0k}^{1995}), aber der Gruppenzugehörigkeits-Koeffizienten des Ausgangsjahres (im Anwendungsfall: β_{1k}^{1970} bzw. β_{1k}^{1970}).

Für den Anwendungsfall könnte man beispielsweise folgenden Indexwert D'' berechnen

$$D'' = \frac{1}{2} \sum_{k=1}^J \left| \frac{\exp(\beta_{0k}^{1995})}{\sum_{l=1}^J \exp(\beta_{0l}^{1995})} - \frac{\exp(\beta_{0k}^{1995} + \beta_{1k}^{1970})}{\sum_{l=1}^J \exp(\beta_{0l}^{1995} + \beta_{1l}^{1970})} \right|, \quad (6)$$

wobei β_{0k}^{1995} die ‚unkorrigierten‘ Konstanten des MNLM für das Jahr 1995 (siehe Tabelle 3) und β_{1k}^{1970} die ‚korrigierten‘ Gruppenzugehörigkeits-Koeffizienten des MNLM für das Jahr 1970 (siehe Tab. 4) sind.

D'' ist interpretierbar als Dissimilarität, die sich ergeben hätte, wenn sich zwischen 1970 und 1995 zwar der Umfang einzelner Familientypen, nicht aber die um Alter×Geschlecht kontrollierten Zugangsstrukturen von Migranten relativ zu den

Deutschen geändert hätten. Wendet man die Koeffizienten aus Tabelle 3 und Tabelle 5 gemäß Formel (6) an, so ergibt sich:

$$D'' = .327$$

Verglichen mit dem D' -Wert von .362 für 1970 besagt dies, dass die (demographisch korrigierte) Ungleichheit der Familientypen von 1970 bis 1995 um .035 Indexpunkte abnimmt, wenn man nur eine strukturelle Verschiebung der Familientypen zulässt, aber eine Konstanz im Verhältnis der relativen Zugangsstrukturen von Migranten vs. Deutschen zu den einzelnen Typen unterstellt. Ein Teil der Verringerung der familialen Ungleichheit ($-.035 = .327 - .362$) scheint also auf reine Strukturverschiebungen in der abhängigen Variable (‚familialer Wandel‘) zurückzugehen. Verglichen mit dem tatsächlichen (demographisch kontrollierten) Ausmaß der Ungleichheit von $D' = .226$ im Jahr 1995 liegt der Wert jedoch noch wesentlich höher. Ein ‚Rest‘ der Veränderung ($-.101 = .226 - .327$) ist also einer darüber hinausgehenden Verhaltensangleichung zwischen den beiden Gruppen zuzuschreiben.

7. Resümee: Abschließende Einschätzung des Anwendungsfalls und des generellen methodischen Verfahrens

7.1 Gesamteinschätzung der familialen Assimilation von Migranten

Im Anwendungsbeispiel der familialen Assimilation von Migranten in Deutschland zwischen 1970 und 1995 führt die Kombination von Regressions- und Segregationsansatz zu einer angemesseneren Einschätzung der Situation als die Verwendung der beiden Einzelmethoden. Während sich mit dem isolierten Regressionsansatz nur schwer überhaupt ein Urteil über die Richtung der grundsätzlichen Veränderung treffen lässt, kommt der isolierte Segregationsansatz zunächst zum Befund einer leichten Vergrößerung der Abstände. Aus diesem ‚Gesamteffekt‘ lässt sich durch die Einschaltung von Regressionsverfahren jedoch ein ‚demographischer Effekt‘ und ein Struktureffekt ‚familialer Wandel‘ herausrechnen. Übrig bleibt dann ein ‚bereinigter Assimilationseffekt‘, der den theoretischen Erwartungen einer Angleichung im familialen Verhalten entspricht. Dies lässt sich folgendermaßen zusammenstellen:

Gesamteffekt:				
$D_{1995} - D_{1970}$	=	$.206 - .195$	= $+ .011$	
demographisch kontrollierter Effekt:				
$D'_{1995} - D'_{1970}$	=	$.226 - .362$	= $-.136$	
Zerlegung:				
$+ .011$	=	$-.136$	+ $.147$	
		demographisch bereinigter Effekt	demographischer Effekt	
	=	$(.226 - .362)$	+ $.147$	
	=	$(.226 - .327)$	+ $(.327 - .362)$	+ $.147$
	=	$-.101$	+ $-.035$	+ $.147$
		verbleibender Assim.effekt	Struktureffekt familiärer Wandel	

7.2 Abschließende Bemerkungen zum generellen methodischen Verfahren

Die Verbindung von Regressions- und Segregationsansatz stellt – wie am Anwendungsfall illustriert – ein wichtiges Instrument zur Analyse sozialer Ungleichheit zur Verfügung. Bei einer Einschätzung der Gleichheit von Verteilungen über nominalskalierte Merkmale mit relativ vielen Ausprägungen (der Spezialität der Segregationsindizes) wird eine Kontrolle von Strukturveränderungen in unabhängigen Merkmalen (der Spezialität des Regressionsansatzes) möglich. Eine gewisse ‚methodische Lücke‘ der Ungleichheitsforschung kann somit auf naheliegende Weise geschlossen werden. Darüber hinaus kann mit der vorgeschlagenen Verbindung auch einer Strukturveränderung im abhängigen Merkmal Rechnung getragen werden, auf die Leistung bisheriger alternativer Dekompositions-Ansätze muss somit keineswegs verzichtet werden.

Das hier demonstrierte Verfahren kombiniert das Multinomiale Logitmodell mit dem Dissimilaritätsindex. Grundsätzlich lassen sich auch andere Segregationsindizes mit dem Regressionsansatz kombinieren, um unabhängige Variablen zu kontrollieren. Eine hilfreiche Eigenschaft ist dabei die Kompositions-Invarianz. Darüber hinaus wäre auch eine Verbindung des MNLM mit Maßen für univariate Verteilungen denkbar, etwa dem Entropie-Index, der zur Beschreibung des Ausmaßes einer Streuung über bestimmte Kategorien verwendet wird (White

1986; 200f.). Ein Anwendungsgebiet ist z. B. die Frage nach der Pluralisierung von Lebensformen (Wagner/Franzmann 2000) und auch hier ist sicherlich eine Kontrolle der demographischen Struktur der Bevölkerung sinnvoll. Insgesamt ergeben sich somit aus der Verbindung der elaborierten Regressionsverfahren mit den ‚klassischen‘ Maßzahlen der Ungleichheitsforschung nützliche Werkzeuge und interessante Perspektiven.

Literatur

- Alba, R. / Nee V., 1997: Rethinking Assimilation Theory for a New Era of Immigration. *International Migration Review* 31: 826–874.
- Becker, G.S., 1981: *A Treatise on the Family*. Cambridge: Harvard University Press.
- Becker, G.S. / Landes E.M. / Michael R.S., 1977: An Economic Analysis of Marital Instability. *Journal of Political Economy* 85: 1141–1187.
- Blackburn, R.M. / Jarman, J. / Siltanen, J., 1993: The Analysis of Occupational Gender Segregation over Time and Place: Considerations of Measurement and Some New Evidence. *Work, Employment & Society* 7: 335–362.
- Blau, F.D. / Hendricks, W.E., 1979: Occupational Segregation by Sex: Trends and Prospects. *Journal of Human Resources* 14: 197–210.
- Brüderl, J., 2000: Regressionsverfahren in der Bevölkerungswissenschaft. S. 589–642 in: U. Mueller, B. Nauck, A. Diekmann (Hrsg.), *Handbuch der Demographie 1. Modelle und Methoden*. Berlin: Springer.
- Cortese, C.F. / Falk, R.F. / Cohen, J.K., 1976: Further

- Considerations on the Methodological Analysis of Segregation Indices. *American Sociological Review* 41: 630–637.
- Charles, M. / Grusky, D.B., 1995: Models for Describing the Underlying Structure of Sex Segregation. *American Journal of Sociology* 100: 931–971.
- Duncan, O.D. / Duncan, B., 1955: A Methodological Analysis of Segregation Indexes. *American Sociological Review* 20: 210–217.
- Fawcett, J.T., (Hrsg.), 1972: *The Satisfactions and Costs of Children: Theories, Concepts, and Methods*. Honolulu: East-West Center.
- Gibbs, J.P., 1965: Occupational Differentiation of Negroes and Whites in the United States. *Social Forces* 44: 159–165.
- Gordon, M., 1964: *Assimilation in American Life*. New York: Oxford University Press.
- Handl, J., 1984: Chancengleichheit und Segregation: Ein Vorschlag zur Messung ungleicher Chancenstrukturen und ihrer zeitlichen Entwicklung. *Zeitschrift für Soziologie* 13: 328–345.
- Hakim, C., 1981: Job Segregation: Trends in the 1970 s. *Employment Gazette* 89: 521–529.
- Hakim, C., 1993: Segregated and Integrated Occupations: A New Approach to Analysing Social Change. *European Sociological Review* 9: 289–314.
- Hakim, C., 1996: Theoretical and Measurement Issues in the Analysis of Occupational Segregation. S. 67–88 in: P. Beckmann (Hrsg.), *Gender Specific Occupational Segregation*. Beiträge zur Arbeitsmarkt- und Berufsforschung. Nürnberg: Institut für Arbeitsmarkt- und Berufsforschung der Bundesanstalt für Arbeit.
- Hosmer, D.W. / Lemeshow, S., 1989: *Applied Logistic Regression*. New York: Wiley.
- James, D.R. / Taeuber, K.E., 1985: Measures of Segregation. S. 1–32 in: N.B. Tuma (Hrsg.), *Sociological Methodology*. San Francisco: Jossey Bass.
- Kagitcibasi, C., 1982: Sex Roles, Values of Children, and Fertility. S. 151–180 in: C. Kagitcibasi (Hrsg.), *Sex Roles, Family, and Community in Turkey*. Bloomington: Indiana University.
- Kalter, F. / Granato, N., 2001: Recent Trends of Assimilation in Germany. ZUMA-Arbeitsbericht 01/02, Mannheim.
- Kalter, F. / Granato, N., 2002: Demographic Change, Educational Expansion, and Structural Assimilation of Immigrants: The Case of Germany. *European Sociological Review* 18: (forthcoming).
- Karmel, T. / MacLachlan M., 1988: Occupational Sex Segregation – Increasing or Decreasing? *The Economic Record* 64: 187–195.
- Lieberson, S., 1980: *A Piece of the Pie: Black and White Immigrants Since 1880*. Berkeley and Los Angeles: University of California Press.
- Long, J.S., 1997: *Regression Models for Categorical and Limited Dependent Variables*. Advanced Quantitative Techniques in the Social Sciences Series 7. Thousand Oaks: Sage.
- Lüttringer, P. / Riede T., 1997: *Der Mikrozensus: amtliche Daten für die Sozialforschung*. ZUMA-Nachrichten 41: 19–43.
- Massey, D.S., 1977: On the Measurement of Segregation as a Random Variable. *American Sociological Review* 42: 587–590.
- Massey, D.S. / Denton, N.A., 1987: Trends in the Residential Segregation of Blacks, Hispanics, and Asians: 1970–1980. *American Sociological Review* 52: 802–825.
- Massey, D.S. / Denton, N.A., 1988: The Dimensions of Residential Segregation. *Social Forces* 67: 281–315.
- Massey, D.S. / White, M.J. / Voon-Chin P., 1996: The Dimensions of Segregation Revisited. *Sociological Methods and Research* 25: 172–206.
- Nauck, B., 1997: Sozialer Wandel, Migration und Familienbildung bei türkischen Frauen, S. 162–199 in: B. Nauck / U. Schönplugg, (Hrsg.), *Familien in verschiedenen Kulturen*. Stuttgart: Enke.
- Siltanen, J., 1990: Social Change and the Measurement of Occupational Segregation by Sex: An Assessment of the Sex Ratio Index. *Work, Employment & Society* 4: 1–29.
- Spriggs, W.E. / Williams, R.M., 1996: A Logit Decomposition Analysis of Occupational Segregation: Results for the 1970 s and 1980 s. *The Review of Economics and Statistics* 78: 348–354.
- Wagner, M. / Franzmann, F., 2000: Die Pluralisierung der Lebensformen, *Zeitschrift für Bevölkerungswissenschaft* 25: 151–173.
- Watts, M., 1992: How Should Occupational Sex Segregation be Measured? *Work, Employment & Society* 3: 475–487.
- White, M.J., 1986: Segregation and Diversity Measures in Population Distribution. *Population Index* 52: 198–221.

Summary: This paper suggests a procedure to control for independent variables in the measurement of segregation. It links the well known Index of Dissimilarity to the Multinomial Logit Model. This combination of the regression approach and the segregation approach turns out to be a useful tool in analyzing structures of social inequality. Using data from the German population census of 1970 and the German microcensus of 1995 the trend toward family-type assimilation on the part of immigrants in Germany exemplifies the technique. In this case the regression approach would lead to an ambiguous picture of the general development, while the standard segregation approach provides a dubious answer as it does not take into account demographic change. However, combining both approaches allows us to control for structural changes in the assessment of dissimilarity. In the end, this lends support to the hypothesis that the family types of Germans and of immigrants have converged over the last decades.