

Hundertmark, Carsten

**Working Paper**

## Ökonometrische Verfahren zur Messung von Segregation: Eine theoretische und empirische Studie

SOEPPapers on Multidisciplinary Panel Data Research, No. 559

**Provided in Cooperation with:**

German Institute for Economic Research (DIW Berlin)

*Suggested Citation:* Hundertmark, Carsten (2013) : Ökonometrische Verfahren zur Messung von Segregation: Eine theoretische und empirische Studie, SOEPPapers on Multidisciplinary Panel Data Research, No. 559, Deutsches Institut für Wirtschaftsforschung (DIW), Berlin

This Version is available at:

<https://hdl.handle.net/10419/75401>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

**SOEPpapers**  
on Multidisciplinary Panel Data Research

SOEP – The German Socio-Economic Panel Study at DIW Berlin

559-2013

**Ökonometrische Verfahren zur Messung von  
Segregation – eine theoretische und  
empirische Studie**

Carsten Hundertmark

## **SOEPPapers on Multidisciplinary Panel Data Research** at DIW Berlin

This series presents research findings based either directly on data from the German Socio-Economic Panel Study (SOEP) or using SOEP data as part of an internationally comparable data set (e.g. CNEF, ECHP, LIS, LWS, CHER/PACO). SOEP is a truly multidisciplinary household panel study covering a wide range of social and behavioral sciences: economics, sociology, psychology, survey methodology, econometrics and applied statistics, educational science, political science, public health, behavioral genetics, demography, geography, and sport science.

The decision to publish a submission in SOEPPapers is made by a board of editors chosen by the DIW Berlin to represent the wide range of disciplines covered by SOEP. There is no external referee process and papers are either accepted or rejected without revision. Papers appear in this series as works in progress and may also appear elsewhere. They often represent preliminary studies and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be requested from the author directly.

Any opinions expressed in this series are those of the author(s) and not those of DIW Berlin. Research disseminated by DIW Berlin may include views on public policy issues, but the institute itself takes no institutional policy positions.

The SOEPPapers are available at  
**<http://www.diw.de/soeppapers>**

### **Editors:**

Jürgen **Schupp** (Sociology, Vice Dean DIW Graduate Center)  
Gert G. **Wagner** (Social Sciences)

Conchita **D'Ambrosio** (Public Economics)  
Denis **Gerstorff** (Psychology, DIW Research Director)  
Elke **Holst** (Gender Studies, DIW Research Director)  
Frauke **Kreuter** (Survey Methodology, DIW Research Professor)  
Martin **Kroh** (Political Science and Survey Methodology)  
Frieder R. **Lang** (Psychology, DIW Research Professor)  
Henning **Lohmann** (Sociology, DIW Research Professor)  
Jörg-Peter **Schräpler** (Survey Methodology, DIW Research Professor)  
Thomas **Siedler** (Empirical Economics)  
C. Katharina **Spieß** (Empirical Economics and Educational Science)

ISSN: 1864-6689 (online)

German Socio-Economic Panel Study (SOEP)  
DIW Berlin  
Mohrenstrasse 58  
10117 Berlin, Germany

Contact: Uta Rahmann | [soeppapers@diw.de](mailto:soeppapers@diw.de)

# Ökonometrische Verfahren zur Messung von Segregation - eine theoretische und empirische Studie

Dr. Carsten Hundertmark

## Zusammenfassung

Segregation ist ein viel diskutiertes soziales Thema. Berufliche Segregation liegt vor, wenn sich unterschiedliche Gruppen ungleich auf einzelne Berufe oder Berufsgruppen aufteilen. Üben beispielsweise Frauen in größerem Maße schlechter bezahlte Berufe aus als Männer, so kann dies eine Erklärung für die Lohnlücke zwischen weiblichen und männlichen Arbeitnehmern sein. Regionale Einkommenssegregation liegt vor, wenn sich verschiedene ethnische Gruppen ungleich auf einzelne Regionen aufteilen. Segregation in Schulen kann zu Unterschieden im Bildungsniveau zwischen Schülern aus unterschiedlichen ethnischen Gruppen beitragen. Im Folgenden werden die gängigen Methoden zur Messung von Segregation vorgestellt, verglichen sowie Stärken und Schwächen der einzelnen Verfahren diskutiert. Die methodischen Ansätze stehen hierbei im Vordergrund, aber auch die Darstellung empirischer Ergebnisse ist Gegenstand der nachfolgenden Ausführungen.

Keywords: Einkommen und Armut, Ökonometrie

# 1 Einleitung

Zur Messung von Segregation wurden in den letzten Jahren zahlreiche Indizes entwickelt. Der Dissimilaritätsindex  $ID$  geht auf Duncan und Duncan (1955) zurück ist auch heute noch der vor allem in der amerikanischen Literatur weit verbreitetste Index. Nach Falk (2002) wird der  $ID$  interpretiert als Prozentsatz der Frauen oder Männer, die ihren Beruf wechseln müssten, damit eine Gleichverteilung von Männern und Frauen entsprechend ihrer Verteilung in der Gesamtbeschäftigung vorliegt. Während der oben beschriebene Dissimilaritätsindex  $ID$  die Segregation entsprechend der Berufsgröße gewichtet, werden bei der Berechnung des standardisierten Dissimilaritätsindex  $SID$  nach Gibbs (1965) und Gross (1968) alle Berufsgruppen gleich gewichtet. Der Informationstheorieindex  $H$  (Theil (1972), Theil und Finezza (1971), Zoloth (1976)) misst, wie ungleich sich Männer und Frauen auf Berufe verteilen, wobei die ungleiche Verteilung von Arbeitskräften als Entropie definiert wird. Der Sex Ratio  $SR$  wird von Hakim (1981) angewandt, um Segregation am Britischen Arbeitsmarkt zu analysieren. Hakim (1981) beschreibt dieses Maß als den Unterschied zwischen dem Ausmaß an Überrepräsentation von weiblichen Arbeitskräften in typischen Frauenberufen und Unterrepräsentation von weiblichen Arbeitskräften in typischen Männerberufen. Blackburn et al. (1995) entwickeln eine standardisierte Form dieses Indizes. James und Taeuber (1985) definieren den Variations-Ratio-Index  $R$  als Neigung ethnischer Gruppen, sich ungleich auf lokale Einheiten wie Schulen oder Regionen aufzuteilen. Ein weiteres verbreitetes statistisches Maß zur Darstellung von Einkommensungleichverteilungen ist der Gini-Koeffizient  $G$  nach James und Taeuber (1985). Im Unterschied zum  $ID$  reagiert er auch auf Veränderungen in der Berufsstruktur, die sich zwischen stark und weniger stark geschlechtsspezifisch segregierten Berufen ereignen. Der  $IP$ -Index von Karmel und Maclachlan (1988) stellt eine Weiterentwicklung des zuvor beschriebenen Dissimilaritätsindex  $ID$  dar unter dem Gesichtspunkt der besseren Interpretierbarkeit. Der Index misst den Anteil der Beschäftigten, die ihren Beruf wechseln müssten, um eine Gleichverteilung der Geschlechter über die Berufe zu erzielen unter Beibehaltung der Berufsstruktur der Beschäftigten. Der Marginal-Matching-Index  $MM$  nach Siltanen, Jarman und Blackburn (1995) misst, wie stark Frauenberufe von weiblichen Arbeitskräften und wie stark Männerberufe von männlichen Arbeitskräften ausgeübt werden. Je stärker diese Beziehung, desto höher ist der Grad der Segregation. Hutchens (2001) definiert wünschenswerte Eigenschaften, die ein Index erfüllen sollte und charakterisiert eine Familie von Indizes, die mehrere dieser Eigenschaften erfüllen. Der Square-Root-Index  $S$  nach Hutchens (2004) wird mit der Absicht eingeführt, einen Index zu entwickeln, der ein Maximum wünschenswerter Eigenschaften erfüllt. Mit Ausnahme des Informationstheorieindex  $H$  bleibt die Anwendung sämtlicher zuvor beschriebenen Methoden auf binäre Variablen beschränkt. Es kann aber durchaus von Interesse sein, wie gleich bzw. ungleich sich mehr als zwei Gruppen auf unterschiedliche Berufe oder Regionen verteilen.

Eine weitere Möglichkeit zur Messung von Segregation besteht in der Anwendung des Mutual-Information-Indizes  $M$ . Er bezieht sich auf den Informationstheorieindex  $H$  und

wurde als erstes von Theil (1971) vorgeschlagen und von Fuchs (1975) sowie Mora und Ruiz-Castillo (2003, 2004) angewandt, um geschlechtsspezifische Segregation am Arbeitsmarkt zu messen. Mora und Ruiz-Castillo (2007) sowie Frankel und Volij (2008) entwickeln einen additiv zerlegbaren Index  $M$ , mit dem gemessen werden kann, wie gleich oder ungleich sich ethnische Gruppen auf Schulen über unterschiedliche Regionen verteilen.

Kim und Jargowsky (2005) entwickeln eine allgemeine Form des Gini-Koeffizienten, der als Segregationsmaß sowohl für stetige als auch dichotome Variablen angewandt werden kann. Chakravarty, d'Ambrosio und Silber (2009) beabsichtigen, einen Koeffizienten herzuleiten unter der Annahme, dass die Anzahl der zu vergleichenden Gruppen frei wählbar ist und die Anzahl der Personen in diesen Gruppen auch null sein kann. Der entsprechende Index entspricht ebenfalls einer allgemeinen mehrdimensionalen Form des Gini-Koeffizienten.

Reardon (2008) argumentiert, dass die herkömmlichen Verfahren zur Messung von Segregation auf nominalen Kategorien beruhen. Diese Methoden sind jedoch unangebracht, wenn entweder die Gruppen oder aber die organisatorischen Einheiten (Regionen, Berufe) ordinal skaliert sind und entwickelt Methoden, die in diesem Fall anwendbar sind.

Ziel dieses Beitrags ist, die gängigen Methoden zur Messung von Segregation vorzustellen, zu vergleichen sowie Stärken und Schwächen der einzelnen Verfahren zu diskutieren. Die methodischen Ansätze stehen hierbei im Vordergrund, aber auch die Darstellung empirischer Ergebnisse ist Gegenstand der nachfolgenden Ausführungen. Dieser Artikel ist folgendermaßen gegliedert: In den Abschnitten 2-5 werden die gängigen Indizes vorgestellt und in Abschnitt 6 erfolgt ein abschließendes Resümee.

## 2 Traditionelle Messverfahren

### 2.1 Methodologie

Im Folgenden entspricht  $m_k \in \{M_k, F_k\}$  der Anzahl der Männer  $M_k$  bzw. der Anzahl der Frauen  $F_k$  in Beruf  $k$ ,  $m \in \{M, F\}$  der Gesamtzahl beschäftigter Männer  $M$  bzw. Frauen  $F$  auf dem Arbeitsmarkt,  $\pi_{km} = \frac{m_k}{M_k + F_k}$  dem Anteil von Gruppe  $m$  in Beruf  $k$ ,  $\pi_m = \frac{m}{M + F}$  dem Anteil von Gruppe  $m$  auf dem Arbeitsmarkt insgesamt,  $t_k = M_k + F_k$  der Anzahl der Beschäftigten in Beruf  $k$  und  $T = M + F$  der Anzahl der Beschäftigten insgesamt.

**Der Dissimilaritätsindex  $ID$**  geht auf Duncan und Duncan (1955) zurück und wurde zur Bestimmung der Wohnungssegregation zwischen der schwarzen und weißen Bevölkerung entwickelt. Er entspricht

$$ID = \frac{1}{2} \sum_{k=1}^K \left| \frac{M_k}{M} - \frac{F_k}{F} \right| \quad (1)$$

bzw. nach Taeuber und Taeuber (1965)

$$ID = \sum_{k=1}^K \frac{t_k |\pi_{km} - \pi_m|}{2T\pi_m(1 - \pi_m)}. \quad (2)$$

Der  $ID$  ist symmetrisch hinsichtlich der beiden Gruppen und insofern ist es egal, ob in Gleichung (2.2) Gruppe  $M$  oder  $F$  bei der Berechnung herangezogen wird. Er kann interpretiert werden als Prozentsatz von Männern und Frauen, die ihren Beruf wechseln müssten damit sich Männer und Frauen gleich auf die Berufe aufteilen.

**Der standardisierte Dissimilaritätsindex  $ID_{st}$**  nach Gibbs (1965) und Gross (1968) entspricht

$$ID = \frac{1}{2} \sum_{k=1}^K \left| \frac{F_k/t_k}{\sum_{k=1}^K F_k/t_k} - \frac{M_k/t_k}{\sum_{k=1}^K M_k/t_k} \right|. \quad (3)$$

Er beschreibt ebenfalls den Anteil der Frauen oder Männer, die ihren Beruf wechseln müssten, so dass die Verteilung von Männern und Frauen innerhalb der Berufe ihrer Verteilung im Arbeitsmarkt bzw. in der Belegschaft des Betriebes entspräche, ohne allerdings dabei die Größe des Berufes zu berücksichtigen. Da der Einfluss der Berufsstruktur konstant gehalten wird, sind Vergleiche über die Zeit oder zwischen Betrieben möglich, auch wenn sich die Beschäftigtenzahlen in den Berufsgruppen unterschiedlich entwickeln.

**Der Sex Ratio  $SR$**  wird von Hakim (1981) angewandt, um Segregation am Britischen Arbeitsmarkt zu analysieren. Er wird definiert als

$$H = \frac{F_f/T_f}{F/T} - \frac{F_m/T_m}{F/T}, \quad (4)$$

wobei die Indizes  $f, m$  Frauen- und Männerberufe kennzeichnen. Beispielsweise stellt  $F_m$  die Anzahl von Frauen in männerdominierten Berufen dar und  $T_m$  die Anzahl sämtlicher Erwerbstätigen in männerdominierten Berufen. Ein Beruf wird als frauendominiert definiert, wenn der Anteil der Frauen im Beruf größer ist als der Anteil der Frauen auf dem gesamten Arbeitsmarkt, also  $F_k/t_k > F/T$ . Die Definition männerdominierter Berufe erfolgt analog. Hakim (1981) beschreibt dieses Maß als den Unterschied zwischen dem Ausmaß an Überpräsentation von weiblichen Arbeitskräften in typischen Frauenberufen und Unterpräsentation von weiblichen Arbeitskräften in typischen Männerberufen.

**Der  $IP$ -Index** von Karmel und Maclachlan (1988) stellt eine Weiterentwicklung des zuvor beschriebenen Dissimilaritätsindex  $ID$  dar unter dem Gesichtspunkt der besseren Interpretierbarkeit. Er wird definiert als

$$IP = \frac{1}{T} \sum_{k=1}^K |m_k - \pi_m t_k|. \quad (5)$$

Nach Falk (2002) misst der *IP*-Index den Anteil der Beschäftigten im Gegensatz zu Frauen oder Männern wie beim *ID*, die ihren Beruf wechseln müssten, um eine Gleichverteilung der Geschlechter über die Berufe zu erzielen unter Beibehaltung der Berufsstruktur der Beschäftigten. Die Indizes *IP* und *ID* stehen in folgender Beziehung zueinander:

$$IP = 2\pi_m(1 - \pi_m)D. \quad (6)$$

**Der Marginal-Matching-Index *MM*** nach Siltanen, Jarman und Blackburn (1995) misst, vergleichbar mit *SR* wie stark Frauenberufe von weiblichen Arbeitskräften und wie stark Männerberufe von männlichen Arbeitskräften ausgeübt werden. Es wird zwischen Frauen- und Männerberufen unterschieden und folgende Segregationsmatrix erzeugt:

$$\begin{pmatrix} F_f & M_f \\ F_m & M_m \end{pmatrix}, \quad (7)$$

Es gilt nun, eine Scheidelinie so zu ziehen, dass die Segregationsmatrix symmetrisch ist, d.h., dass  $M_f = F_m$  und somit

$$\sum_{k=1}^{n^F} (M_k + F_k) = F = \sum_{k=1}^n F_k \quad (8)$$

mit  $n^F$  als Anzahl frauendominierter Berufe. Für den üblichen Fall, dass die Gleichung nicht exakt aufgeht, schlagen Blackburn, Jarman und Siltanen (1993) eine Interpolation vor. Die Scheidelinie variiert je nach der sich verändernden Geschlechterzusammensetzung der Beschäftigung im Zeitverlauf. Der Marginal-Matching-Index *MM* wird definiert als

$$MM = \frac{1}{FM} (F_f M_m - F_m M_f) = \frac{F_f}{F} - \frac{M_m}{M}. \quad (9)$$

**Der Gini-Index *G*** nach James und Taeuber (1985) entspricht

$$G = \sum_{k=1}^K \sum_{s=1}^K \frac{t_k t_s |\pi_{km} - \pi_{sm}|}{2T^2 \pi_m (1 - \pi_m)}, \quad (10)$$

und kann interpretiert werden als Summe der gewichteten durchschnittlichen absoluten Differenzen der Anteile einer bestimmten Personengruppe  $m$  zwischen allen möglichen Paaren von Berufsgruppen geteilt durch den größtmöglichen Wert dieser Summe, der sich bei vollständiger Segregation ergeben würde. Im Unterschied zum *ID* reagiert er auch auf Veränderungen in der Berufsstruktur, die sich zwischen stark und weniger stark geschlechtsspezifisch segregierten Berufen ereignen.

**Der Variations-Ratio-Index *R*** nach James und Taeuber (1985) wird definiert als

$$R = \sum_{k=1}^K \frac{t_k(\pi_{km} - \pi_m)^2}{T\pi_m(1 - \pi_m)}, \quad (11)$$

und kann interpretiert werden als Varianz der Anteile von Arbeitnehmern einer bestimmten Personengruppe  $m$  in den Berufen geteilt durch die Varianz der dichotomen Variablen Geschlecht bei sämtlichen Arbeitskräften. Er wird als Neigung von Gruppen definiert, sich ungleich auf Berufe aufzuteilen.

**Der Square-Root-Index**  $S$  nach Hutchens (2004) wird mit der Absicht eingeführt, einen Index zu entwickeln, der ein Maximum wünschenswerter Eigenschaften erfüllt. Er entspricht

$$S = \sum_{k=1}^K \left[ \sqrt{\frac{F_k}{F} \cdot \frac{M_k^{\text{keine Segregation}}}{M}} - \sqrt{\frac{F_k}{F} \cdot \frac{M_k}{M}} \right] \quad (12)$$

und kann interpretiert werden als Summe der Unterschiede zwischen den beobachteten Anteilen von Männern und Frauen über die Berufsgruppen hinweg und den Anteilen, die sich bei gleichmäßiger Aufteilung der beiden Gruppen über die Berufe ergeben würden. Für jede Berufsgruppe wird dieser Unterschied definiert als Differenz des geometrischen Mittels der Anteile von Männern und Frauen auf die Berufe in Abwesenheit von Segregation und dem geometrischen Mittel der tatsächlich beobachteten Anteile der beiden Geschlechter. Da im Fall nicht vorhandener Segregation  $\frac{F_k}{F} = \frac{M_k}{M}$  gilt stets

$$S = \sum_{k=1}^K \left[ \frac{F_k}{F} - \sqrt{\frac{F_k}{F} \cdot \frac{M_k}{M}} \right]. \quad (13)$$

Mit Ausnahme des  $SR$  nehmen sämtliche vorgestellten Indizes den Wert 0 bei nicht vorhandener Segregation und den Wert 1 bei vollständiger Segregation an. Die vorgestellten traditionellen Messverfahren sind nicht für den Mehrgruppenfall geeignet. Sie können auf beliebige Personengruppen und andere organisatorische Einheiten wie z.B. Schulen oder Regionen übertragen werden.

## 2.2 Empirische Resultate

In der folgenden Studie werden die zuvor beschriebenen Indizes angewandt, um berufliche Segregation zu messen. Für die Segregationsmessung wird das vom Deutschen Institut für Wirtschaftsforschung (DIW) erstellte Sozio-ökonomische Panel (SOEP) verwendet (Wagner, Frick und Schupp (2007)). Es handelt sich hierbei um eine repräsentative Befragung privater Haushalte in Deutschland, die im jährlichen Rhythmus bei denselben Personen und Haushalten seit 1984 durchgeführt wird. Bereits im Juni 1990, also noch vor der

Währungs-, Wirtschafts- und Sozialunion, wurde die Studie auf das Gebiet der ehemaligen DDR ausgeweitet.

Bei der Auswahl einer geeigneten Stichprobe werden die selben Kriterien wie bei den Lohn-diskriminierungsmessungen herangezogen, d.h., dass sozialversicherungspflichtige Arbeiter, Angestellte und Beamte in Vollzeit- oder Teilzeittätigkeit von 18-65 Jahren aus den zugehörigen SOEP-Wellen herausgefiltert werden. Nicht berücksichtigt sind Selbständige, Praktikanten und Auszubildende, unbezahlt mithelfende Familienangehörige und Personen, die ausschließlich in so genannten Mini-Jobs oder 1-Euro-Jobs tätig sind. Die vorgestellten Studien beziehen sich auf die Jahre 1999 und 2006 für Gesamtdeutschland. Arbeitnehmer mit fehlenden oder unplausiblen Antworten werden wie in vergleichbaren, auf dem SOEP basierenden Analysen (vgl. Melly (2006)) nicht berücksichtigt. Hierbei wird unterstellt, dass keine systematische Antwortverweigerung oder -verfälschung vorliegt. Fehlende Werte sind über sämtliche Beobachtungen zufällig verteilt und somit ist kein Selektionsbias zu erwarten.

Die in dieser Studie gewählten Berufsgruppen unterscheiden nach beruflichem Status, wobei das zugrunde liegende Skalenniveau stark mit der Treiman-Prestige-Skala<sup>1</sup> korreliert. Es werden folgende Berufsgruppen gewählt:

- Gruppe 1: Tätigkeit mit geringem Status und wenig Eigenständigkeit.
- Gruppe 2: Beschäftigung mit einem geringen Maß an Spezialisierung.
- Gruppe 3: Tätigkeit, die eine gewisse Schulbildung erfordert und ein begrenztes Maß an Verantwortung einschließt.
- Gruppe 4: Leitende aber nicht prestigeträchtige Tätigkeit, die Fachhochschul oder Universitätsabschluss erfordert.
- Gruppe 5: Prestigeträchtige Tätigkeit mit umfassenden Führungsaufgaben.

Die Stichprobe umfasst für das Jahr 1999 bei Männern 1979 Beobachtungen, die der Frauen 1136 Beobachtungen und für das Jahr 2006 bei Männern 3270 Beobachtungen und bei Frauen 3166 Beobachtungen. Tabelle (1) zeigt wie sich Frauen und Männer auf die Berufsgruppen aufteilen,

Tabelle (2) die Ergebnisse der Segregationsmessungen.

Wird der Gini-Koeffizient  $G$  als Maßzahl herangezogen, so fällt die Segregation mit 0.291 (2006) und 0.298 (1999) am Größten, beim Square-Root-Index  $S$  mit 0.039 (2006) und 0.041 (1999) am Geringsten aus. Die Werte des Sex Ratios können nicht mit den anderen Indizes verglichen werden, da es sich hierbei um ein nicht normiertes Maß handelt.

---

<sup>1</sup>Der Prestigerang von Berufen wird in bildungssoziologischen Untersuchungen als relativ aussagekräftige Kategorie benutzt. Donald J. Treiman entwickelte den Standard Index of Occupational Prestige Scala (SIOPS) (auch Treiman-Index genannt) auf Grundlage des Klassifikationssystems ISCO von 1968.

Jahr	Gruppe	Männer	Frauen	$\Sigma$
2006	1	522	393	915
	2	1257	835	2092
	3	845	1637	2482
	4	314	181	495
	5	332	120	452
	$\Sigma$		3270	3166
1999	1	335	149	484
	2	645	296	941
	3	417	526	943
	4	511	148	659
	5	71	17	88
	$\Sigma$		1979	1136

Tabelle 1: Aufteilung von Frauen und Männern auf die Berufsgruppen

Jahr	$ID$	$ID_{st}$	$SR$	$IP$	$G$	$R$	$S$
2006	0.259	0.198	0.555	0.129	0.291	0.076	0.039
1999	0.252	0.219	0.759	0.117	0.298	0.076	0.041

Tabelle 2: Ergebnisse der Segregationsmessungen

Die Indexwerte bleiben im Zeitverlauf relativ konstant. Während Segregationsmessungen auf Basis des  $ID$  oder  $IP$  zu dem Ergebnis kommen, dass die Segregation im Zeitverlauf geringfügig zugenommen hat, bleiben die anderen Maßzahlen konstant oder gehen geringfügig zurück.

### 3 Auf Entropie basierende Verfahren zur Messung von Segregation

#### 3.1 Der Informationstheorieindex als Maß für Segregation

##### 3.1.1 Methodologie

Beim Informationstheorieindex  $H$  nach Theil (1972) wird die Entropie  $E$  als Maß für die Ungleichverteilung von Männern und Frauen auf die Berufe herangezogen. Es gilt für die Entropie insgesamt

$$E = \sum_{k=1}^K \pi_k \log_2 \frac{1}{\pi_k} \quad (14)$$

bzw. für die Entropie  $E_m$  in den einzelnen Personengruppen

$$E_m = \sum_{k=1}^K \pi_{mk} \log_2 \frac{1}{\pi_{mk}} \quad (15)$$

mit  $\pi_k = \frac{t_k}{T}$  dem Anteil von in Berufsgruppe  $k$  beschäftigter Arbeitnehmer auf dem Arbeitsmarkt insgesamt und  $\pi_{mk} = m_k/m$  dem Anteil von in Beruf  $k$  beschäftigter Erwerbstätiger in Gruppe  $m$ .<sup>2</sup> Die Entropie nimmt den Wert 0 bei nur einer Berufsgruppe an und erreicht ihr Maximum, wenn die Anteile der Individuen auf die Berufe bei den ethnischen Gruppen identisch sind. Der Informationstheorieindex  $H$  wird definiert als

$$H = \sum_{i=1}^M \frac{m_i(E - E_m)}{TE}, \quad (16)$$

mit  $M \geq 2$  Personengruppen. Er kann auf beliebige Personengruppen und andere organisatorische Einheiten wie z.B. Schulen oder Regionen übertragen werden.

## 3.2 Der Mutual-Information-Index als Maß für Segregation

### 3.2.1 Methodologie

Frankel und Volij (2008) entwickeln einen additiv zerlegbaren Segregationsindex<sup>3</sup>, mit dem gemessen werden kann, wie gleich oder ungleich sich ethnische Gruppen auf Schulen über unterschiedliche Regionen verteilen. Im Folgenden soll gelten:  $G$  sei eine diskrete Zufallsvariable mit  $M$  möglichen Realisationen und  $p_m$  die zugehörige Wahrscheinlichkeit. Sei beispielsweise  $G$  die ethnische Gruppe eines zufällig ausgewählten Schülers, dann entspricht  $p_m$  dem Anteil der Schüler der  $m$ -ten Gruppe in einer bestimmten Region. Als Maß für die Ungleichverteilung der ethnischen Gruppen wird die bereits in Gleichung (15) beschriebene Entropie  $E = \sum_{m=1}^M \pi_m \log_2 \frac{1}{\pi_m}$  herangezogen. Im Folgenden sei die ethnische Gruppe, der ein Schüler angehört, nicht bekannt, dafür aber die Schule  $n$ , die er besucht. Wenn Segregation hinsichtlich der ethnischen Verteilung zwischen den Schulen  $N$  in der Region vorliegt, können daraus zusätzliche Informationen über die Gruppenzugehörigkeit gezogen werden. Als Maß für den erwarteten Informationsgewinn zwischen  $G$  und  $N$  wird die gewichtete Summe der Entropien innerhalb der Schulen definiert, wobei die Gewichtung durch den Anteil der Schüler in den Schulen erfolgt. Der Mutual-Information-Index  $M$  entspricht

<sup>2</sup>Man definiert  $0 \cdot \log_2(0) = \lim_{\pi \rightarrow 0} \pi \log_2 \pi = 0$

<sup>3</sup>Mit additiver Zerlegbarkeit sind starke Schul- und Gruppenzerlegbarkeit (vgl. Gleichungen (21) und (22)) gemeint

$$M = \left( \begin{array}{c} \text{Entropie der} \\ \text{ethnischen Verteilung} \\ \text{in der Region} \end{array} \right) - \sum_{\substack{\text{Schulen } n \\ \text{in der Region}}} \left( \begin{array}{c} \text{Anteil der} \\ \text{Schüler in} \\ \text{Schule } n \end{array} \right) \left( \begin{array}{c} \text{Entropie der} \\ \text{ethnischen Verteilung} \\ \text{in Schule } n \end{array} \right). \quad (17)$$

dh. der Minderung der Ungewissheit über die ethnische Herkunft eines Schülers aus der Kenntnis heraus, welche Schule dieser besucht. Da der Mutual-Information-Index  $M$  symmetrisch bezüglich der beiden Variablen  $G$  und  $N$  ist, gilt

$$M = \left( \begin{array}{c} \text{Entropie der} \\ \text{Schulverteilung} \\ \text{in der Region} \end{array} \right) - \sum_{\substack{\text{ethnische Gruppen } g \\ \text{in der Region}}} \left( \begin{array}{c} \text{Anteil der} \\ \text{Schüler in} \\ \text{Gruppe } g \end{array} \right) \left( \begin{array}{c} \text{Entropie der} \\ \text{Schulverteilung} \\ \text{in Gruppe } g \end{array} \right). \quad (18)$$

Der Informationstheorieindex  $H$  nach Gleichung (16) und der Mutual-Information-Index  $M$  nach Gleichung (17) stehen in folgender Beziehung zueinander:

$$H = M / \left( \begin{array}{c} \text{Entropie} \\ \text{innerhalb} \\ \text{der Region} \end{array} \right). \quad (19)$$

Während  $H$  nur Werte zwischen 0 und 1 annehmen kann, ist der Wertebereich für  $M$  nach oben nicht beschränkt. Im Folgenden soll gelten:

**Definition 1:** Eine *Schulregion*  $X$  setzt sich zusammen aus:

- Einer nichtleeren und endlichen Menge von ethnischen Gruppen  $\mathbf{G}(X)$ .
- Einer nichtleeren und endlichen Menge von Schulen  $\mathbf{N}(X)$ .
- Für jede ethnische Gruppe  $g \in \mathbf{G}(X)$  und jede Schule  $n \in \mathbf{N}(X)$  eine nichtnegative Anzahl  $T_g^n$  von Mitgliedern in Gruppe  $g$ , die Schule  $n$  besuchen.

Eine Schulregion kann auch im Listenformat  $\langle (T_g^n)_{g \in \mathbf{G}} \rangle_{n \in \mathbf{N}}$  spezifiziert werden. Beispielsweise bezeichnet  $\langle (10, 20), (30, 10) \rangle$  eine Region mit zwei ethnischen Gruppen und zwei Schulen. Die erste Schule (10, 20) umfasst 10 Inländer und 20 Ausländer, die zweite Schule (30, 10) 30 Inländer und 10 Ausländer.

Für zwei beliebige Schulregionen  $X$  und  $Y$  bezeichnet  $X \uplus Y$  den Zusammenschluss der Schulen in  $X$  und der Schulen in  $Y$  in einer Region. Ist  $\alpha$  ein nichtnegativer Skalar, so meint  $\alpha X$ , dass die Anzahl der Schüler in jeder Gruppe und Schule in Region  $X$  mit  $\alpha$  multipliziert wird.  $c(X)$  entspricht dem Zusammenschluss aller Schulen in  $X$  zu einer Schule. Ist beispielsweise  $X = \langle (10, 20), (30, 10) \rangle$  und  $Y = \langle (40, 50) \rangle$ , so ergibt  $X \uplus Y = \langle (10, 20), (30, 10), (40, 50) \rangle$ ,  $2X = \langle (20, 40), (60, 20) \rangle$  und  $c(X) = \langle (40, 30) \rangle$ .

Es wird folgende Ausdrucksweise gewählt:

- $T_g = \sum_{n \in \mathbf{N}} T_g^n$ : Anzahl der Schüler in Gruppe  $g$  in einer Region
- $T^n = \sum_{g \in \mathbf{G}} T_g^n$ : Anzahl der Schüler, die Schule  $n$  besuchen
- $T = \sum_{g \in \mathbf{G}} T_g$ : Anzahl der Schüler insgesamt in einer Region

Im Folgenden sei  $C$  die Menge aller Schulregionen. Eine Segregationsrelation  $\succeq$  sei eine zweistellige, transitive und totale Relation auf  $C$ . Die Beziehung  $X \succeq Y$  meint, dass Region  $X$  mindestens genauso segregiert ist, wie Region  $Y$ . Die Relationen  $\succ$  und  $\sim$  werden von  $\succeq$  auf die übliche Weise abgeleitet<sup>4</sup>. Der Segregationsindex ist eine Funktion  $S : C \rightarrow R$ , d.h., dass jeder Region aus  $C$  eine reelle Zahl zugeordnet wird. Der Index  $S$  repräsentiert die Segregationsrelation  $\succeq$  für alle Paare von Regionen  $X, Y \in C$ :

$$X \succeq Y \iff S(X) \geq S(Y). \quad (20)$$

Wird die Transformation  $\sigma : C \rightarrow C$  auf eine Schulregion  $X$  angewandt, so bleibt die Segregation in der Region  $X$  unverändert, d.h., dass  $\sigma(X) \sim X$ . Gilt diese Beziehung für sämtliche Regionen, so ist die Segregation invariant gegenüber der Transformation  $\sigma$ .

Nach Frankel und Volij (2008) erfüllt der Mutual-Information-Index  $M$  folgende wünschenswerte Eigenschaften:

- Symmetrie: Die Segregation in einer Region ist invariant gegenüber Umbenennungen oder Neusortierungen von Gruppen und Schulen.
- Schulspaltung: Es sei  $\mathbf{N}$  die Menge aller Schulen in der Region  $X \in C$  und  $X'$  das Ergebnis, wenn  $n \in \mathbf{N}$  Schulen in zwei Schulen  $n_1$  und  $n_2$  aufgespalten werden. Dann gilt  $X' \succeq X$ . Haben die beiden Schulen dieselbe ethnische Verteilung, dann gilt  $X' \sim X$ .
- Typ-1-Unabhängigkeit: Es seien  $X, Y \in C$  zwei Regionen mit identischer Anzahl an Schülern und denselben ethnischen Verteilungen, wobei  $Y$  dabei als Neuverteilung der Schüler auf die Schulen angesehen werden kann.<sup>5</sup> Durch das Hinzufügen eines Clusters  $Z$  bestehend aus nur einer Schule soll gelten  $X \succeq Y$  genau dann, wenn  $X \uplus Z \succeq Y \uplus Z$ , d.h., dass durch das Hinzufügen eines Clusters  $Z$  bestehend aus nur einer Schule nicht beeinflusst wird, welche Region stärker segregiert. Der  $ID$  nach Duncan und Duncan (1955) verletzt dieses Prinzip. Beispielsweise sei  $X = \langle (50, 100), (50, 0) \rangle$ ,  $Y = \langle (100, 40), (0, 60) \rangle$  und  $Z = \langle (100, 0) \rangle$ . Es gilt:  $ID(X) = 0.5 < ID(Y) = 0.6$ ,

<sup>4</sup> $x \sim y$ , wenn  $X \succeq Y$  und  $Y \succeq X$ ,  $X \succ Y$ , wenn  $X \succeq Y$  aber nicht  $Y \succeq X$

<sup>5</sup>Typ-1-Unabhängigkeit setzt nicht voraus, dass die Anzahl der Schulen in  $X$  und  $Y$  identisch ist.

aber  $ID(X \uplus Z) = 0.75 > ID(X \uplus Y) = 0.6$ . Wird als Segregationsmaß der Mutual-Information-Index herangezogen, so ist für das gewählte Beispiel  $M(X) = 0.311 < M(Y) = 0.396$  und  $M(X \uplus Z) = 0.459 < M(X \uplus Y) = 0.516$ , womit Typ-1-Unabhängigkeit erfüllt ist.

- Typ-2-Unabhängigkeit: Es seien  $X, Y, Z \in C$  drei Regionen mit  $T(X) = T(Y)$ , d.h., dass die Anzahl der Schüler in  $X$  und  $Y$  identisch ist. Weiterhin sei das Cluster  $c(Z)$  der Zusammenschluss aller Schulen in  $Z$  zu einer Schule. Es soll gelten  $X \uplus Z \succeq Y \uplus Z$  genau dann, wenn  $X \uplus c(Z) \succeq Y \uplus c(Z)$ . Die meisten Indizes verletzen dieses Prinzip, da sie nicht zwischen Segregation zwischen Clustern und innerhalb eines Clusters unterscheiden. Beispielsweise sei  $X = \langle (50, 0), (0, 100) \rangle$ ,  $Y = \langle (100, 0), (0, 50) \rangle$  und  $Z = \langle (50, 0), (0, 100) \rangle$ . Die ethnische Verteilung in den Schulregionen  $X$  und  $Z$  ist identisch mit jeweils 50 In- und 100 Ausländern und weicht von  $Y$  ab. Demzufolge ist die Segregation zwischen den Clustern  $X$  und  $Z$  geringer als zwischen  $Y$  und  $Z$ . Allerdings berücksichtigen die meisten Indizes nur Segregation innerhalb eines Clusters. Wird als Segregationsmaß der  $ID$  nach Duncan und Duncan (1955) herangezogen, werden die Cluster  $X \uplus Z$  und  $Y \uplus Z$  als maximal und identisch segregiert angesehen mit  $ID(X \uplus Z) = ID(Y \uplus Z) = 1$ . Für das gewählte Beispiel ergibt sich  $ID(X \uplus c(Z)) = 0.5 < ID(Y \uplus c(Z)) = 2/3$ , womit Typ-2-Unabhängigkeit verletzt ist. Bei Verwendung des Mutual-Information-Indexes gilt:  $M(X \uplus Z) = 0.981 < M(Y \uplus Z) = 1$  und  $M(X \uplus c(Z)) = 0.459 < M(Y \uplus c(Z)) = 0.541$ , womit Typ-2-Unabhängigkeit erfüllt ist.
- Gruppenspaltung: Es sei  $\mathbf{G}$  die Menge der ethnischen Gruppen in einer Region  $X \in C$  und  $X'$  das Ergebnis, wenn  $g \in \mathbf{G}$  Gruppen in zwei Gruppen  $g_1$  und  $g_2$  aufgespalten werden in der Art, dass  $\frac{T_{g_1}^n}{T_{g_1}} = \frac{T_{g_2}^n}{T_{g_2}}$  für alle  $n \in \mathbf{N}$ . Dann ist  $X \sim X'$ .
- Größeninvarianz: Wenn die Anzahl der Schüler in jeder ethnischen Gruppe und jeder Schule  $n$  mit derselben positiven Konstanten  $\alpha$  multipliziert wird, dann gilt für jede Region  $X \in C$  und jeden Skalar  $\alpha$ , dass  $X \sim \alpha X$ .
- Skaleninvarianz: Sei  $X'$  das Ergebnis, wenn die Anzahl der Schüler einer bestimmten ethnischen Gruppe  $g \in \mathbf{G}(X)$  in jeder Schule  $n$  mit derselben positiven Konstanten  $\alpha$  multipliziert wird. Dann gilt  $X' \sim \alpha X$ .
- Starke Schulzerlegbarkeit: Bei jeder beliebigen Aufteilung  $X = X^1 \uplus \dots \uplus X^K$  der Schulen einer Region in  $K$  Cluster gilt

$$S(X) = S(c(X^1) \uplus \dots \uplus c(X^K)) + \sum_{k=1}^K P^k S(X^k) \quad (21)$$

mit  $P^k$  gleich dem Anteil der Schüler in Cluster  $k$ , wobei  $S(c(X^1) \uplus \dots \uplus c(X^K))$  der Segregation zwischen den  $K$  Clustern und  $S(X^k)$  der Segregation innerhalb des

Clusters  $k$  entspricht. Dieses Prinzip wird von den meisten Indizes verletzt.<sup>6</sup>

- Starke Gruppenzerlegbarkeit: Bei jedem beliebigen Zusammenschluss von ethnischen Gruppen einer Region in  $K$  Supergruppen gilt

$$S = S_K + \sum_{k=1}^K P^k S_k \quad (22)$$

mit  $P_k$  gleich dem Anteil der Schüler in Supergruppe  $k$ .  $S_K$  entspricht der Segregation der Region, wenn jede Supergruppe wie eine einzelne Gruppe behandelt wird und  $S_k$  der Segregation der Region, wenn alle Schüler, die nicht der Supergruppe angehören getrennt behandelt werden. Dieses Prinzip soll an folgendem Beispiel erläutert werden: In einer Region wird zwischen vier ethnischen Gruppen A, B, C und D unterschieden, wobei A, B und C zu einer Supergruppe zusammengefasst werden. Unter zweimaliger Verwendung von Gleichung (22) kann die Segregation in drei Terme zerlegt werden:

$$\begin{aligned} \begin{pmatrix} \text{Segregation} \\ \text{zwischen} \\ \text{A,B,C \& D} \end{pmatrix} &= \begin{pmatrix} \text{Segregation} \\ \text{zwischen} \\ \{A,B,C\} \& D \end{pmatrix} + \begin{pmatrix} \text{Anteil der} \\ \text{Schüler in} \\ \text{A,B, oder C} \end{pmatrix} \begin{pmatrix} \text{Segregation} \\ \text{zwischen} \\ \text{A \& \{B,C\}} \end{pmatrix} \\ &+ \begin{pmatrix} \text{Anteil der} \\ \text{Schüler in} \\ \{B,C\} \end{pmatrix} \begin{pmatrix} \text{Segregation} \\ \text{zwischen} \\ \text{B \& C} \end{pmatrix} \end{aligned} \quad (23)$$

Starke Gruppenzerlegbarkeit wird ebenfalls von den meisten Indizes nicht erfüllt.

- Stetigkeit: Für drei beliebige Regionen  $X, Y, Z \in C$  stellen  $\{c \in [0, 1] : cX \uplus (1 - c)Y \succeq Z\}$  und  $\{c \in [0, 1] : Z \succeq cX \uplus (1 - c)Y\}$  abgeschlossene Mengen dar. Dieses Axiom ist technischer Natur und wird benötigt, um zu beweisen, dass die Segrationsrelation durch einen Index repräsentiert wird.
- Nichttrivialität: Es existieren zwei Regionen  $X, Y \in C$  mit zwei nichtleeren ethnischen Gruppen, die nicht identisch segregiert sind, also  $X \succ Y$ . Dieses Axiom wird benötigt, um die triviale Segrationsrelation auszuschließen.

### 3.3 Empirische Resultate

In der folgenden Studie wird die vom Landesbetrieb für Statistik und Kommunikationstechnologie Niedersachsen bereitgestellte und frei verfügbare LSKN-Online Datenbank herangezogen. Die betrachteten Schulegionen sind die Samtgemeinden Nenndorf und Rodenberg im Landkreis Schaumburg in Niedersachsen. Bei den Schulen werden Grundschulen ausgewählt, da sich diese im Gegensatz zu anderen Schulformen in beiden Gemeinden

---

<sup>6</sup>Siehe Empirische Resultate

befinden. Bei den ethnischen Gruppen wird zwischen In- und Ausländern unterschieden. Die Spezifikation für das Jahr 2010 lautet im Listenformat:

- $N = \langle (337, 26), (240, 2) \rangle$  für die Samtgemeinde Nenndorf
- $R = \langle (364, 12), (278, 13) \rangle$  für die Samtgemeinde Rodenberg
- $N \uplus R = \langle (337, 26), (240, 2), (364, 12), (278, 13) \rangle$  für den Zusammenschluss beider Gemeinden

Segregation wird zunächst für beide Gemeinden separat ermittelt und anschließend für den Zusammenschluss beider Gemeinden in einer Region. Tabelle (3) zeigt die Ergebnisse der Segregationsmessungen,

Region	$H$	$M$
$N$	0.072	0.020
$R$	0.003	0.001
$N \uplus R$	0.040	0.010

Tabelle 3: Ergebnisse der Segregationsmessungen

wobei unter Verwendung des Mutual-Information-Indizes  $S(N \uplus R)$  auch gemäß Gleichung (21) ermittelt werden kann. Bei nur zwei Schulregionen  $N$  und  $R$  gilt:

$$S(N \uplus R) = S(c(N) \uplus c(R)) + P^R S(R) + P^N S(N)$$

mit  $c(N) \uplus c(R) = \langle (577, 28) \rangle \uplus \langle (642, 25) \rangle = \langle (577, 28), (642, 25) \rangle$ .

Die Schulsegregation ist bei beiden Indizes in der Gesamtgemeinde Rodenberg (0.003 bzw. 0.001) geringer als in der Gesamtgemeinde Nenndorf (0.072 bzw. 0.020) und liegt für den Zusammenschluss beider Gemeinden (0.040 bzw. 0.010) dazwischen.

## 4 Der Gini-Koeffizient als Maß für Segregation bei stetigen Variablen

### 4.1 Methodologie

Der Gini-Koeffizient  $G$  ist ein weit verbreitetes statistisches Maß zur Darstellung von Einkommensungleichverteilungen. Er wird ebenfalls als Segregationsmaß verwendet, obwohl seine Anwendung in diesem Zusammenhang im Allgemeinen auf binäre Variablen beschränkt bleibt. Kim und Jargowsky (2005) entwickeln eine modifizierte Form dieses

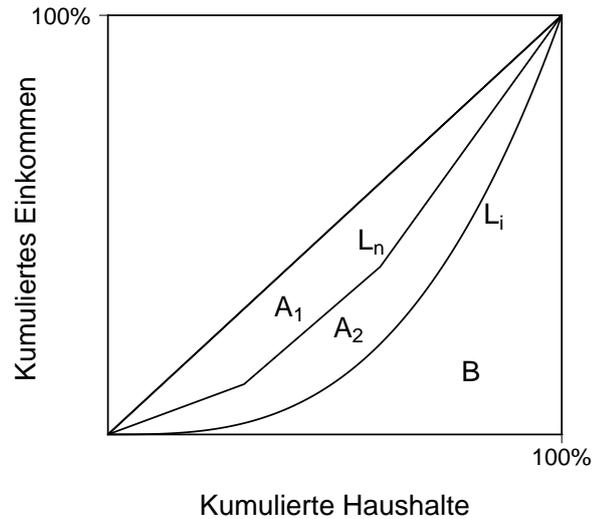


Abbildung 1: Lorenzkurven für individuelle Haushalte und Regionen

Indizes, der auch als Segregationsmaß für stetige Variablen angewandt werden kann.

Wie in Abbildung (1) dargestellt, spiegelt die Lorenzkurve den kumulierten Prozentsatz des gesamten Haushaltseinkommens, geordnet nach der Größe des Einkommens, als Funktion des kumulierten Anteils an den gesamten Haushalten wieder. Der Gini-Koeffizient der Einkommensungleichverteilung wird definiert als Fläche zwischen Lorenzkurve und Winkelhalbierender geteilt durch die gesamte Fläche unterhalb der Winkelhalbierenden. In Abbildung (1) stellt  $L_i$  die Lorenzkurve für die individuellen Haushaltseinkommen und  $L_n$  die Lorenzkurve für die mittleren Einkommen in unterschiedlichen Regionen dar. Der Gini-Koeffizient für die individuellen Haushalte entspricht

$$G^i = (A_1 + A_2)/(A_1 + A_2 + B), \quad (24)$$

und für die Regionen

$$G^n = A_1/(A_1 + A_2 + B). \quad (25)$$

Nach Kim und Jargowsky (2005) entspricht Segregation dem Ausmaß der Ungleichheit zwischen Gruppen bei *gegebener* Verteilung der individuellen Einheiten. Sie geben zu bedenken, dass der Index  $G^n$  als Segregationsmaß ungeeignet ist. Zum Einen misst er lediglich die Ungleichheit zwischen Gruppen und berücksichtigt dabei nicht, wie sich die Verteilung der Individuen ändert, wenn individuelle Einheiten zu Gruppen zusammengefasst werden. Zum Anderen hängt die obere Grenze des Gini-Koeffizienten  $G^n$  ebenfalls von der Verteilung der individuellen Einheiten ab. Bei maximaler Segregation - Haushalte verteilen sich auf Regionen in der Art, dass in einer Region nur Haushalte mit demselben Einkommen vorkommen - konvergiert die Lorenzkurve für Haushalte  $L_n$  gegen die

Lorenzkurve für Individuen  $L_i$  und die obere Grenze von  $G^n$  wird durch  $G^i$  festgelegt. Wird in allen Regionen dasselbe durchschnittliche Einkommen erzielt, nimmt  $G^n$  den Wert null an. In Anlehnung an Jahn, Schmid und Schrag (1947) schlagen Kim und Jargowsky (2005) vor, bei der Messung von Einkommenssegregation zwischen Regionen die Einkommensungleichverteilung zwischen individuellen Haushalten mit zu berücksichtigen und den Koeffizienten  $G^n$  neu zu skalieren. Sie definieren

$$G^S = \frac{G^n}{G^i} = \frac{A_1/(A_1 + A_2 + B)}{(A_1 + A_2)/(A_1 + A_2 + B)} = \frac{A_1}{A_1 + A_2}, \quad (26)$$

wobei  $G^S$  ebenfalls Werte zwischen null bei nicht vorhandener Segregation und eins bei maximaler Segregation annehmen kann.

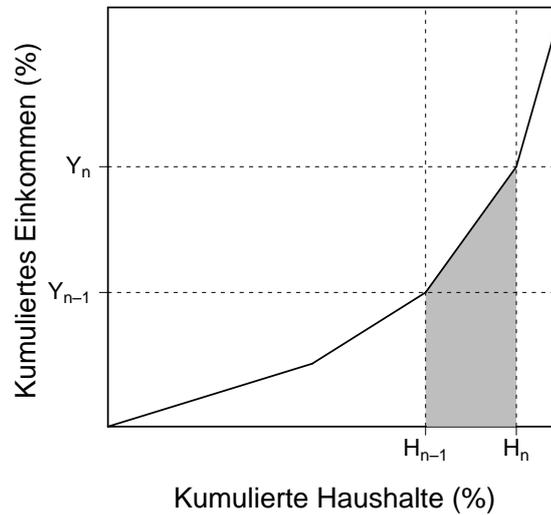


Abbildung 2: Kumuliertes Haushaltseinkommen

Aus Abbildungen (1) und (2) geht hervor, dass

$$\begin{aligned} G^i &= (A_1 + A_2)/(A_1 + A_2 + B) \\ &= (0.5 - B)/0.5 \\ &= 1 - 2B \end{aligned} \quad (27)$$

$$2B = \sum_{i=1}^N (Y_{i-1} + Y_i)(H_i - H_{i-1}). \quad (28)$$

$Y_i$  stellt den kumulierten Prozentsatz der Haushaltseinkommen und  $H_i$  den kumulierten Anteil an den gesamten Haushalten dar, aufsteigend geordnet von 1 bis  $N$  nach der Höhe des Einkommens, wobei  $N$  der Anzahl der Haushalte insgesamt entspricht.  $G^n$  wird auf analoge Weise ermittelt, wobei die Haushalte nun aufsteigend nach der Höhe des mittleren Einkommens in den Regionen geordnet sind und  $Y_{ni}$  den kumulierten Prozentsatz der durchschnittlichen Einkommen darstellt. Es gilt  $H_i - H_{i-1} = 1/N$  und somit ist

$$\begin{aligned}
G^S = \frac{G^n}{G^i} &= \frac{1 - \sum_{i=1}^N (Y_{ni-1} + Y_{ni})(H_i - H_{i-1})}{1 - \sum_{i=1}^N (Y_{i-1} + Y_i)(H_i - H_{i-1})} \\
&= \frac{N - \sum_{i=1}^N (Y_{ni-1} + Y_{ni})}{N - \sum_{i=1}^N (Y_{i-1} + Y_i)}.
\end{aligned} \tag{29}$$

Nach Anand (1983) kann Gleichung (29) wie folgt umgeformt werden:

$$G^S = \frac{G^n}{G^i} = \frac{\frac{1}{2N^2m} \sum_{i=1}^N \sum_{j=1}^N |y_{ni} - y_{nj}|}{\frac{1}{2N^2m} \sum_{i=1}^N \sum_{j=1}^N |y_i - y_j|}, \tag{30}$$

wobei  $y_i$  dem Einkommen des  $i$ ten Haushalts,  $y_{ni}$  dem durchschnittliche Haushaltseinkommen der Region, in dem sich der Haushalt befindet und  $m$  dem mittleren Einkommen sämtlicher Haushalte entspricht. Die Gleichungen (29) und (30) können nicht nur für stetige sondern auch für binäre Variablen angewandt werden, was an folgendem Beispiel verdeutlicht werden soll: Eine Grundgesamtheit  $T$  setzt sich aus  $F$  Frauen und  $M$  Männern zusammen ( $T = F + M$ ), wobei jeder Frau der Wert eins und jedem Mann der Wert null zugeordnet wird. Kim und Jargowsky (2005) definieren  $P = M/T$  und fassen die Abzisse der Lorenzkurve als Prozentsatz der Gesamtbevölkerung und die Ordinate als Prozentsatz der Männer auf. In Gleichung (30) wird  $m$  durch  $P$  ersetzt,  $y_i$  ist gleich eins bei Frauen und null bei Männern,  $y_{ni}$  entspricht  $P_{ni}$ , dem Anteil der Frauen in der Region, in der sich der  $i$ -te Haushalt befindet. Der Nenner  $G^i$  in Gleichung (30) wird  $1 - P$  und somit gilt

$$G^S = \frac{\sum_{i=1}^N \sum_{j=1}^N |P_{ni} - P_{nj}|}{2N^2P(1 - P)}. \tag{31}$$

Gleichung (31) wird in der Literatur häufig bei binären Variablen verwendet und entspricht Gleichung (10), Gleichungen (29) und (30) stellen eine Verallgemeinerung dar.

## 4.2 Empirische Resultate

Bei der folgenden Untersuchung wird erneut auf das Sozio-ökonomische Panel (SOEP) zurückgegriffen. Es wird zwischen Haushalten in unterschiedlichen Regionen in Gesamtdeutschland unterschieden, wobei sich die Segregationsmessungen auf die Jahre 1999 und 2006 beziehen. Innerhalb dieses Zeitraums sind unterschiedliche politische Maßnahmen durchgeführt worden, die das Ausmaß der Segregation beeinflussen können. Arbeitsmarktpolitisch wurden einzelne Gesetze zur Reform des Arbeitsmarktes mit den Kurbezeichnungen Hartz I, Hartz II, Hartz III und Hartz IV umgesetzt. Steuerpolitisch wurden u.a. Eingangs- und Spitzensteuersatz kontinuierlich herabgesetzt. Weiterhin kann das Maß der Einkommensungleichverteilung von Veränderungen der Haushaltszusammensetzungen und einer im betrachteten Zeitraum gesunkenen Arbeitslosenquote abhängen.

In dieser Studie wird zwischen folgenden Regionen unterschieden:

- Nord: Schleswig-Holstein, Hamburg, Niedersachsen und Bremen.
- Süd: Hessen, Baden-Württemberg und Bayern.
- West: Nordrhein-Westfalen, Rheinland-Pfalz und Saarland .
- Ost: Berlin, Brandenburg, Mecklenburg-Vorpommern, Sachsen, Sachsen-Anhalt und Thüringen.

Das Haushaltseinkommen entspricht dem jährlichen Nettohaushaltseinkommen nach Steuern und staatlichen Transferzahlungen sämtlicher Mitglieder eines Haushaltes, wobei das Einkommen für das Jahr 1999 entsprechend dem vom Statistischen Bundesamt erhobenen Verbraucherpreisindex<sup>7</sup> auf das Jahr 2006 umgerechnet wurde.

Jahr		Nord	Süd	West	Ost
2006	Anzahl	1697	3876	3067	3063
	Einkommen	35052	36595	35561	27087
1999	Anzahl	1852	4368	3637	3401
	Einkommen	31224	32200	30811	26386

Tabelle 4: Aufteilung von Haushalten auf die Regionen

Tabelle (4) zeigt, wie sich die Haushalte in der Stichprobe auf die unterschiedlichen Regionen in Gesamtdeutschland aufteilen mit den zugehörigen mittleren Haushaltseinkommen. In sämtlichen Regionen ist das Realeinkommen im Zeitverlauf angestiegen, wobei der Zuwachs in den neuen Bundesländern am Kleinsten ausfällt. Zu beiden Zeitpunkten ist das Einkommen im Süden am Höchsten und im Osten am Geringsten.

<sup>7</sup>DESTATIS 2009, Verbraucherpreisindex für Deutschland, Lange Reihe ab 1948-Juni 2009

Jahr	$G^i$	$G^n$	$G^S$
2006	0.373	0.056	0.151
1999	0.324	0.039	0.119

Tabelle 5: Ergebnisse der Segregationsmessungen

Tabelle (5) zeigt die Ergebnisse der Segregationsmessungen. Der Gini-Koeffizient  $G^i$  für die Einkommen der individuellen Haushalte steigt im betrachteten Zeitraum von 0.324 auf 0.373, der Gini-Koeffizient  $G^n$  für die mittleren Einkommen der Haushalte in den Regionen von 0.039 auf 0.056 und der Segregationsindex  $G_S$  von 0.119 auf 0.151.

## 5 Methoden zur Messung ordinaler Regression

### 5.1 Methodologie

Traditionelle Methoden zur Messung von Segregation beschreiben, wie sich nominal skalierte Gruppen (z.B. Männer und Frauen) auf ebenfalls nominal skalierte organisatorische Einheiten (z.B. Regionen, Schulen) verteilen. Reardon (2008) entwickelt einen allgemeinen Ansatz, mit dem Segregation gemessen werden kann, wenn entweder die Gruppen oder die organisatorischen Einheiten ordinal skaliert sind.

Zur Motivation soll folgendes Beispiel dienen: Auf einem Arbeitsmarkt befinden sich 100 männliche und 100 weibliche Arbeitskräfte, die sich unterschiedlich auf vier Berufsgruppen verteilen. Die Berufsgruppen unterscheiden nach beruflichem Status, ansteigend von Kategorie eins bis vier. In Szenario A befinden sich Männer überproportional in Berufen mit hohem Status, Frauen dagegen überproportional in Berufen mit niedrigem Status. In Szenario B tauschen bei beiden Geschlechtern die Verteilungen auf die Berufsgruppen 1 und 2 genauso wie die Verteilungen auf die Gruppen 3 und 4 (siehe Tabelle 6). Ein traditionelles Segregationsmaß wie beispielsweise der Dissimilaritätsindex nach Duncan und Duncan (1955) würde die Berufsgruppen kategorial behandeln und beide Szenarien gleich einschätzen. In jedem Szenario gibt es eine Berufsgruppe mit einem 40/10 Männer/Frauen Verhältnis, eine mit einem 30/20 Männer/Frauen Verhältnis, eine mit einem 20/30 Männer/Frauen Verhältnis und mit einem 10/40 Männer/Frauen Verhältnis. Ziel des Ansatzes nach Reardon (2008) ist deshalb, ein geeignetes Maß zu finden, wie gleich oder ungleich sich ungeordnete Gruppen von Arbeitnehmern auf geordnete Berufsgruppen aufteilen.

Im Folgenden sei  $k$  eine ordinal skalierte Variable mit den Ausprägungen  $1,2,\dots,K$  für unterschiedliche nach beruflichem Status geordnete Berufsgruppen und  $m$  bezeichne un-

beruflicher Status					
	1	2	3	4	insgesamt
Szenario A					
Männer	10	20	30	40	100
Frauen	40	30	20	10	100
insgesamt	50	50	50	50	200
Szenario B					
Männer	20	10	40	30	100
Frauen	30	40	10	20	100
insgesamt	50	50	50	50	200

Tabelle 6: Unterschiedliche Szenarien der Berufsaufteilungen

geordnete Kategorien  $1, 2, \dots, M$  von Arbeitnehmern. Es wird angenommen, dass auf dem Arbeitsmarkt  $T$  Arbeitskräfte beschäftigt sind, wobei jedes Individuum mit  $m$  und  $t$  klassifiziert wird. Die Ausdrücke  $t_{m\cdot}, t_{\cdot k}, t_{mk}$  geben die Anzahl der Arbeiter in Kategorie  $m$ , Kategorie  $k$  und beiden Kategorien an. Innerhalb jeder ungeordneten Kategorie  $m$  sei  $c_{mk}$  das kumulierte Verhältnis

$$c_{mk} = \sum_{j=1}^k \frac{t_{mj}}{t_m}. \quad (32)$$

Die Verteilung der Individuen in Kategorie  $m$  auf die Berufsgruppen kann nun als  $K$ -Tupel  $\mathbf{c}_m = (c_{m1}, c_{m2}, \dots, c_{mK})$  dargestellt werden. Da  $c_{mK} = 1$  wird nur das  $[K-1]$ -Tupel  $\mathbf{c}_m = (c_{m1}, c_{m2}, \dots, c_{m[K-1]})$  benötigt, um die Verteilung in  $m$  vollständig zu beschreiben. Auf dieselbe Weise wird die Verteilung sämtlicher Arbeiter auf die Berufsgruppen als  $[K-1]$ -Tupel  $\mathbf{c} = (c_1, c_2, \dots, c_{[K-1]})$  dargestellt. Zusätzlich wird folgende Ausdrucksweise gewählt: Die Verteilung in  $m$  dominiert die Verteilung in  $n$  über die geordneten Berufsgruppen von  $h$  bis  $j$  ( $1 \leq h \leq j \leq K$ ), wenn  $c_{mk} < c_{nk}$  für alle  $k \in (h, \dots, j)$ .

Folgende Eigenschaften sind bei der Messung ordinaler Segregation wünschenswert:

- **Skaleninterpretierbarkeit:** Ein Segregationsindex wird genau dann maximiert, wenn in jeder ungeordneten Kategorie  $m$  die Arbeitskräfte nur einer Berufsgruppe zugehören ( $c_{mk} \in \{0, 1\}$  für alle  $m$  und  $k$ ). Ein Segregationsindex wird genau dann minimiert, wenn in jeder ungeordneten Kategorie  $m$  die Verteilung der einzelnen Arbeitskräfte auf die Berufsgruppen identisch ist mit der Verteilung sämtlicher Erwerbstätigen auf die Berufsgruppen ( $c_{mk} = c_k$  für alle  $m$  und  $k$ ). Die Intuition ist, dass maximale Segregation dann auftritt, wenn keine Variation in den geordneten Kategorien innerhalb jeder ungeordneten Gruppe auftritt und minimale Se-

gregation dann, wenn die Verteilung von  $k$  innerhalb  $m$  der Verteilung sämtlicher Erwerbstätigen gleicht.

- Ungeordnete Kategorienäquivalenz: Wenn sich in zwei ungeordneten Kategorien  $m$  und  $n$  die Arbeitskräfte im selben Verhältnis auf die Berufsgruppen  $k$  aufteilen und  $m$  und  $n$  zu einer Kategorie zusammengefasst werden, bleibt die Segregation unverändert. Absicht hierbei ist, dass zwei identisch verteilte ungeordnete Kategorien nicht voneinander segregieren und in sofern auch der Zusammenschluss dieser Kategorien das Ausmaß der Segregation nicht beeinflussen darf.
- Größeninvarianz: Wenn die Anzahl der Erwerbstätigen in jeder ungeordneten Kategorie  $m$  und jeder geordneten Berufsgruppe  $k$  mit derselben positiven Konstanten multipliziert wird ( $c \cdot t_{mk}$  für alle  $m$  und  $k$ ), bleibt die Segregation unverändert. Die Intuition hierbei ist, dass nur die Verteilung zwischen Kategorien und nicht die Anzahl der Arbeitskräfte insgesamt die Segregation beeinflusst.
- Nullgruppeneffekte: Wenn eine geordnete Kategorie  $j$  ohne Arbeitskräfte ( $t_{mj} = t_{.j} = 0$  für alle  $m$ ) den anderen geordneten Berufsgruppen  $k = 1, 2, \dots, K$  hinzugefügt wird, sollen folgende Bedingungen erfüllt sein: 1. Ist  $j = 0$  (d.h., Berufsgruppe  $j$  hat geringeren Status als sämtliche  $K$  Berufsgruppen) oder ist  $j = K + 1$  (d.h., Berufsgruppe  $j$  hat höheren Status als sämtliche  $K$  Berufsgruppen) bleibt die Segregation unverändert. 2. Ist  $1 < j < K$  und  $c_{mj} \in \{0, 1\}$  für alle  $m$  (d.h., dass für sämtliche Arbeitskräfte in jeder ungeordneten Kategorie  $m$   $k < j$  oder  $k > j$  gilt) und existieren ungeordnete Kategorien  $m$  und  $n$  mit  $c_{mj} = 0$  und  $c_{nj} = 1$  nimmt die Segregation zu, es sei denn sie hat schon ihren maximalen Wert erreicht. Die Absicht ist hierbei, dass sich Segregation durch das Hinzugügen von Nullkategorien nur dann ändern soll, wenn sich auch der Abstand zwischen den Berufsgruppen verändert.
- Wechsel: Wenn die Verteilung in  $m$  die Verteilung in  $n$  über die geordneten Berufsgruppen von  $j$  bis  $k$  dominiert und wenn ein Individuum aus Kategorie  $j$  von  $n$  nach  $m$  wechselt während ein Individuum aus Kategorie  $k$  von  $m$  nach  $n$  wechselt, nimmt die Segregation ab. Auf dieselbe Weise wird das Ausmaß der Segregation geringer, wenn ein Individuum aus Kategorie  $m$  von  $k$  nach  $j$  wechselt während ein Individuum aus Kategorie  $n$  von  $j$  nach  $k$  wechselt. Die Intuition ist hierbei, dass bei einem Wechsel von Individuen aus unterschiedlichen geordneten Kategorien in der Art, dass sich die Verteilungen in zwei ungeordneten Kategorien angleichen, die Segregation geringer wird.
- geordnete Wechsel: Wenn die Verteilung in  $m$  die Verteilung in  $n$  über die geordneten Berufsgruppen von  $j$  bis  $l$  dominiert (mit  $j < k < l$ ) und wenn ein Individuum aus Kategorie  $j$  von  $n$  nach  $m$  wechselt und gleichzeitig ein Individuum aus Kategorie  $l$  von  $m$  nach  $n$  wechselt, dann ist das Ausmaß der Segregationsreduktion größer als wenn ein Individuum aus Kategorie  $j$  von  $n$  nach  $m$  wechselt während ein Individuum aus Kategorie  $k$  von  $m$  nach  $n$  wechselt. Gleichermäßen ist der Rückgang

der Segregation größer, wenn ein Individuum aus Kategorie  $m$  von  $l$  nach  $j$  wechselt während ein Individuum aus Kategorie  $n$  von  $j$  nach  $l$  wechselt als wenn ein Individuum aus Kategorie  $m$  von  $k$  nach  $j$  wechselt während ein Individuum aus Kategorie  $n$  von  $j$  nach  $k$  wechselt. Die Absicht hierbei ist, dass Wechsel zwischen Individuen aus weiter entfernten Kategorien die Segregation stärker beeinflussen als Wechsel zwischen Individuen aus näher gelegenen Kategorien.

- Additive ungeordnete Gruppenzerlegung: Wenn  $M$  ungeordnete Kategorien in  $N$  Cluster zusammengefasst werden, wird die Segregation wie in Reardon und Firebaugh (2002) beschrieben in unabhängige Komponenten innerhalb und zwischen den Clustern zerlegt.
- Additive geordnete Gruppenzerlegung: Wenn  $K$  geordnete Kategorien in  $J$  Superkategorien zusammengefasst werden durch den Zusammenschluss benachbarter Gruppen, soll die Segregation in eine Summe unabhängiger Komponenten innerhalb und zwischen den Superkategorien zerlegt werden.

Reardon und Firebaugh (2002) folgend wird ein ordinales Segregationsmaß  $\Lambda^v$  wie folgt definiert:

$$\Lambda^v = \sum_{m=1}^M \frac{t_m}{T_v} (v - v_m), \quad (33)$$

wobei  $v$  die Streuung der ordinalen Variablen  $k$  und  $v_m$  die Streuung der ordinalen Variablen  $k$  innerhalb Kategorie  $m$  beschreibt<sup>8</sup>. Um ordinale Segregation zu bestimmen, wird für die Streuung einer ordinal skalierten Variablen mit der Ausprägung  $k = 1, 2, \dots, K$  ein Maximum  $\mathbf{c}_0$  definiert, wenn  $k = 1$  für die eine Hälfte der Arbeitskräfte und  $k = K$  für die andere Hälfte, also  $\mathbf{c} = \mathbf{c}_0 = (c_1, c_2, \dots, c_{[K-1]}) = (1/2, 1/2, \dots, 1/2)$ . Ein Minimum wird definiert, wenn  $k = c$  mit  $c \in (1, 2, \dots, K)$  für alle Beobachtungen. In diesem Fall existieren  $K$  mögliche Verteilungen von  $k$ , wie z.B.  $\mathbf{c} = (0, 0, \dots, 0, 1, 1, \dots, 1)$  ( $c_j = 0$  für  $j < k$  und  $c_j = 1$  für alle  $j \geq k$ ). Messung der ordinalen Streuung läuft dann darauf hinaus, wie weit die Verteilung der ordinalen Variablen vom Maximum oder Minimum entfernt ist, wobei das Maximum zu 1 und das Minimum zu 0 normiert werden kann. Reardon (2008) schlägt vor,  $\mathbf{c}$  als Verteilungsfunktion einer ordinalen Variablen zu betrachten, wobei die Streuung als Funktion des Abstandes zwischen  $\mathbf{c}$  und der Geraden  $c_K = 1/2$  gemessen werden kann. Eine allgemeine Formulierung für das Streuungsmaß lautet:

$$v = \frac{1}{K-1} \sum_{j=1}^{K-1} f(c_j), \quad (34)$$

wobei  $f(c)$  als stetige Funktion auf dem Intervall  $[0,1]$  definiert wird in der Form, dass  $f(c)$  monoton wachsend für  $c \in [0, \frac{1}{2}]$  und fallend für  $c \in (\frac{1}{2}, 1]$  ist mit einem Maximalwert

---

<sup>8</sup>Wird  $k$  als intervallskalierte Variable behandelt und ist  $v$  die Varianz dieser Variablen, so ist  $\Lambda^v$  äquivalent zu  $R^2$ , dem Bestimmtheitsmaß einer Regression von  $k$  auf eine Menge von Dummy-Variablen für die Kategorien  $m = 1, 2, \dots, M$

$f(1/2) = 1$  auf dem Intervall  $[0,1]$  und einem Minimalwert  $f(0) = f(1) = 0$  auf dem Intervall  $[0,1]$ . Reardon (2008) nennt vier mögliche Funktionen  $f$ :<sup>9</sup>

$$\begin{aligned} f_1(c) &= -[c\log_2 c + (1-c)\log_2(1-c)] \\ f_2(c) &= 4c(1-c) \\ f_3(c) &= 2\sqrt{c(1-c)} \\ f_4(c) &= 1 - |2c - 1| \end{aligned} \tag{35}$$

Werden diese Funktionen in Gleichung (34) eingesetzt, ergeben sich vier mögliche Maße für die Streuung einer ordinalen Variablen, die entsprechend mit  $v_1, v_2, v_3$  und  $v_4$  bezeichnet werden. Sie nehmen alle den Wert 1 bei maximaler Streuung und den Wert 0 bei minimaler Streuung an. Gemäß Gleichung (33) erhält man vier Indizes zur Messung von Segregation:

$$\begin{aligned} \Lambda^1 &= \sum_{m=1}^M \frac{t_m}{Tv_1} (v_1 - v_{1m}) \\ \Lambda^2 &= \sum_{m=1}^M \frac{t_m}{Tv_2} (v_2 - v_{2m}) \\ \Lambda^3 &= \sum_{m=1}^M \frac{t_m}{Tv_3} (v_3 - v_{3m}) \\ \Lambda^4 &= \sum_{m=1}^M \frac{t_m}{Tv_4} (v_4 - v_{4m}) \end{aligned} \tag{36}$$

Der Index  $\Lambda^1$  ist eine ordinale Verallgemeinerung des konventionellen Zweigruppen-Informationstheorieindizes  $H$  (Theil (1972), Theil und Finezza (1971), Zoloth (1976)) und soll im Folgenden  $H^0$  bezeichnet werden. Bei zwei Berufsgruppen ( $K = 2$ ) entsprechen  $H^0$  und  $H$ . Auf ähnliche Art entspricht  $\Lambda^2$  einer ordinalen Verallgemeinerung des Variationsratioindizes (James und Taeuber (1985), Reardon und Firebaugh (2002)) und soll im Folgenden  $R^0$  bezeichnet werden. Für  $K = 2$  entsprechen  $R^0$  und  $R$ . Des Weiteren kann  $\Lambda^3$  als ordinale Verallgemeinerung des Hutchens-Square-Root-Indizes  $S$  (Hutchens (2004)) angesehen werden und wird im Folgenden mit  $S^0$  bezeichnet. Auch hier entsprechen für  $K = 2$  die Indizes  $S^0$  und  $S$ . Im Gegensatz zu den vorher beschriebenen Indizes ist  $\Lambda^4$  nicht vergleichbar mit anderen Segregationsmaßen. Die vier Indizes können als durchschnittliche Distanz zwischen der Geamtstreuung einer ordinalen Variablen und ordinaler Streuung innerhalb kategorialer Gruppen betrachtet werden. Sie unterscheiden sich nur in der funktionalen Form, wie die Streuung gemessen wird. Reardon (2008) zeigt, dass die meisten zuvor beschriebenen Eigenschaften, die bei der Messung von Segregation

---

<sup>9</sup>Reardon (2008) definiert  $0 \cdot \log_2(0) = \lim_{x \rightarrow 0} x \log_2 x = 0$

wünschenswert sind, von den Indizes erfüllt werden. Die Eigenschaften Skaleninterpretierbarkeit und Wechsel werden nicht von  $\Lambda^4$  und additive geordnete Gruppenzerlegung wird von keinem der aufgeführten Segregationsmaße erfüllt.

## 5.2 Empirische Resultate

In der folgenden Studie wird das Verfahren nach Reardon (2006) zunächst angewandt, um berufliche Segregation zwischen Männern und Frauen und anschließend zwischen Ausländern, Ausländerinnen, Inländern und Inländerinnen zu ermitteln. Die Stichproben entsprechen den schon in Kapitel 2 verwendeten und umfassen für das Jahr 2006 3012 Inländer, 258 Ausländer, 2974 Inländerinnen und 192 Ausländerinnen, für das Jahr 1999 1748 Inländer, 231 Ausländer, 1065 Inländerinnen und 71 Ausländerinnen. Die Berufsgruppen unterscheiden erneut nach beruflichem Status, ansteigend von Kategorie eins bis fünf.

Tabelle (7) zeigt, wie sich die vier ungeordneten Kategorien von Arbeitnehmern auf die Berufsgruppen aufteilen,

Jahr	Gruppe	Inländer	Ausländer	Inländerinnen	Ausländerinnen	$\Sigma$
2006	1	404	118	306	87	915
	2	1160	97	778	57	2092
	3	811	34	1591	46	2482
	4	309	5	180	1	495
	5	0328	4	119	1	452
	$\Sigma$	3012	258	2974	192	6436
1999	1	219	116	112	37	484
	2	569	76	279	17	941
	3	399	18	511	15	943
	4	492	19	146	2	659
	5	69	2	17	0	88
	$\Sigma$	1748	231	1065	71	3115

Tabelle 7: Berufsaufteilung

Tabelle (8) die Ergebnisse der Segregationsmessungen.

Die Segregation zwischen Männern und Frauen fällt bei der Methode nach Reardon (2006) geringer aus als bei den in Kapitel 2 verwendeten, die Berufsgruppen kategorial behandelnden Indizes, wobei die Ungleichverteilung auf die Berufe zwischen 1999 und 2006 geringfügig angestiegen ist. Wird zusätzlich zwischen In- und Ausländern unterschieden erhöht sich das Ausmaß der Segregation im betrachteten Zeitraum.

Jahr	Gruppen	$\Lambda^1$	$\Lambda^2$	$\Lambda^3$	$\Lambda^4$
2006	Männer, Frauen	0.018	0.018	0.016	0.054
	Inländer/,-innen	0.046	0.048	0.038	0.072
	Ausländer/,-innen				
1999	Männer, Frauen	0.013	0.014	0.010	0.000
	Inländer/,-innen	0.056	0.064	0.044	0.071
	Ausländer/,-innen				

Tabelle 8: Ergebnisse der Segregationsmessungen

## 6 Resüme

In dieser Studie wurden zunächst traditionelle Methoden angewandt, um zu messen, wie sich berufliche Segregation zwischen männlichen und weiblichen Arbeitnehmern im Zeitraum von 1999 - 2006 entwickelt hat. Mit Ausnahme des  $SR$  nehmen sämtliche vorgestellten Indizes den Wert 0 bei nicht vorhandener Segregation und den Wert 1 bei vollständiger Segregation an. Die vorgestellten traditionellen Messverfahren sind nur für den Zweigruppen- und nicht für den Mehrgruppenfall geeignet. Wird der Ginikoeffizient  $G$  als Maßzahl herangezogen, so fällt die Segregation zu beiden Zeitpunkten mit 0.291 (2006) und 0.298 (1999) am Größten, beim Square-Root-Index  $S$  mit 0.039 (2006) und 0.041 (1999) am Geringsten aus. Die Werte des Sex Ratios können nicht mit den anderen Indizes verglichen werden, da es sich hierbei um ein nicht normiertes Maß handelt. Die Indexwerte bleiben im Zeitverlauf relativ konstant. Während  $ID$  und  $IP$  zu dem Ergebnis kommen, dass die Segregation im Zeitverlauf geringfügig zugenommen hat, bleiben die anderen Maßahlen konstant oder gehen geringfügig zurück.

Eine weitere Möglichkeit zur Messung von Segregation besteht in der Anwendung des Mutual-Information-Indizes  $M$ , der sich auf den Informationstheorieindex  $H$  bezieht. Mit Hilfe des von Frankel und Volij (2008) entwickelten Indizes  $M$  kann gemessen werden, wie gleich oder ungleich sich ethnische Gruppen auf Schulen über unterschiedliche Regionen verteilen. Der Vorteil des Mutual-Information-Indizes gegenüber anderen Messverfahren besteht darin, dass er eine Vielzahl wünschenswerter Eigenschaften erfüllt. Die in dieser Studie betrachteten Schulegionen sind die Samtgemeinden Nenndorf und Rodenberg im Landkreis Schaumburg in Niedersachsen. Bei den Schulen wurden Grundschulen ausgewählt, da sich diese im Gegensatz zu anderen Schulformen in beiden Gemeinden befinden, bei den ethnischen Gruppen wird zwischen In- und Ausländern unterschieden. Die Segregation fällt in beiden Gemeinden gering aus. Der Mutual-Information-Index nimmt für die Samtgemeinde Nenndorf den Wert 0.020 und für die Samtgemeinde Rodenberg den Wert 0.001 an. Die Messung der Segregation für den Zusammenschluss beider Gemeinden in einer Region kann nach Gleichung (17) oder (21) erfolgen und ist gleich 0.010. Wird

der Informationstheorieindex zur Messung der Segregation herangezogen ist das Ergebnis vergleichbar.

Kim und Jargowsky (2005) entwickeln eine modifizierte Form des Gini-Koeffizienten  $G$ , der auch als Segregationsmaß für stetige Variablen angewandt werden kann. In dieser Studie wurde gemessen, wie gleich oder ungleich sich das jährliche Nettohaushaltseinkommen auf die Regionen Nord, Süd, West und Ost verteilt. Der Gini-Koeffizient  $G_i$  für die Einkommen der individuellen Haushalte steigt im betrachteten Zeitraum von 0.324 auf 0.373, der Gini-Koeffizient  $G_n$  für die mittleren Einkommen der Haushalte in den Regionen von 0.039 auf 0.056 und der Segregationsindex  $G_S$  von 0.119 auf 0.151.

Traditionelle Methoden zur Messung von Segregation beschreiben, wie sich nominal skalierte Gruppen (z.B. Männer und Frauen) auf ebenfalls nominal skalierte organisatorische Einheiten (z.B. Regionen, Schulen) verteilen. Reardon (2008) entwickelt einen allgemeinen Ansatz, mit dem Segregation gemessen werden kann, wenn entweder die Gruppen oder die organisatorischen Einheiten ordinal skaliert sind. In dieser Studie wurde zunächst berufliche Segregation zwischen Männern und Frauen und anschließend zwischen Ausländern, Ausländerinnen, Inländern und Inländerinnen ermittelt. Die Berufsgruppen unterscheiden sich nach beruflichem Status, ansteigend von Kategorie eins bis fünf. Die Segregation zwischen Männern und Frauen fällt bei der Methode nach Reardon (2006) geringer aus als bei den die Berufsgruppen kategorial behandelnden Indizes, wobei die Ungleichverteilung auf die Berufe zwischen 1999 und 2006 geringfügig angestiegen ist. Wird zusätzlich zwischen In- und Ausländern unterschieden, so erhöht sich das Ausmaß der Segregation im betrachteten Zeitraum.

## Literatur

- ANAND, S. (1983) *Inequality and Poverty in Malaysia: Measurement and Decomposition*, Oxford: Oxford University Press.
- BLACKBURN R.-M., SILTANEN J., JARMAN J. (1995). The Measurement of Occupational Gender Segregation: Current Problems and a New Approach *Journal of the Royal Statistical Society*, Vol. 158, Part 2, S. 319-331.
- CHAKRAVARTY, S-R., D'AMBROSIO, C., SILBER J. (2009). Generalized gini occupational segregation indices, in Yves Flückiger, Sean F. Reardon, Jacques Silber (ed.) *Occupational and Residential Segregation, Research on Economic Inequality*, Vol. 17, Emerald Group Publishing Limited, S.71-95.
- DUNCAN, O.D., DUNCAN, B. (1955). A Methodological Analysis of Segregation Indices, *American Sociological Review*, Vol. 20, No. 2, S. 210-217.
- FALK, S. (2002). Geschlechtsspezifische berufliche Segregation in Ostdeutschland zwischen Persistenz, Verdrängung und Angleichung: ein Vergleich mit Westdeutschland

- für die Jahre 1991-20, *Mitteilungen aus der Arbeitsmarkt- und Berufsforschung*, Vol. 35, Issue 1, S. 37-59.
- FRANKEL, D.-M., VOLIJ, O. (2008). Measuring school segregation, *Discussion Paper*, No. 07-03, Monaster Center for Economic Research, Ben Gurion University of the Negev.
- FUCHS, V. (1975). A Note on Sex Segregation in Professional Occupations, *Explorations in Economic Research*, Vol. 2, S. 105-111.
- GIBBS, J. (1965). Occupational Differentiation of Negroes and Whites *Social Forces*, Vol. 44, No 2, S. 59-165.
- GROSS, E. (1968). Plus ça change...? The Sexual Structure of Occupations Over Time, *Social Problems*, Vol. 16, No 2, S. 198-208.
- HAKIM, C. (1981). Job Segregation: Trends in the 1970s, *Employment Gazette*, December, S. 521- 529.
- HUTCHENS, R.-M. (2001). Numerical measures of segregation: Desirable properties and their implications. *Mathematical Social Sciences*, Vol 42, No 1, S. 13-29.
- HUTCHENS, R.-M. (2004). One measure of segregation *International Economic Review*, Vol. 45 , No. 2), S. 555-578.
- JAHN, J., SCHMID, C.-F., SCHRAG, C. (1947). The Measurement of ecological segregation, *American Sociological Review*, Vol. 3, S. 293-303.
- JAMES, D. R., TAEUBER, K. E. (1985). Measures of segregation, *Sociological Methodology*, Vol. 14, S. 1-32.
- KARMEL, T. AND MACLACHLAN, M. (1988). Occupational Sex Segregation - Increasing or Decreasing, *Economic Record*, Vol. 64, S. 187-95.
- KIM, J., JARGOWSKY, P.-A. (2005). *The Gini Coefficient and Segregation on a Continuous Variable* National Poverty Center Working Paper Series, No. 05-02, Gerald R. Ford School of Public Policy, University of Michigan.
- MELLY, B. (2005). Public-private sector wage differentials in Germany: Evidence from quantile regression, *Empirical Economics*, Vol. 30, No 2, S. 505-520.
- MORA, R., RUIZ-CASTILLO, J. (2003). Additively Decomposable Segregation Indexes. The Case of Gender Segregation by Occupations and Human Capital Levels in Spain, *Journal of Economic Inequality*, Springer, vol. 1, No. 2, S. 147-179.
- MORA, R., RUIZ-CASTILLO, J. (2004). Gender segregation by occupations in the public and the private sector. The case of Spain, *Investigaciones Economicas, Fundación SEPI*, vol. 28, No 3, S. 399-428.

- MORA, R., RUIZ-CASTILLO, J. (2007). The invariance properties of the Mutual Information index of multigroup segregation, *Economics Working Papers*, we077544, Universidad Carlos III, Departamento de Economía.
- REARDON, S.F. (2008). Measures of Ordinal Segregation, *Research on Economic Inequality*, Working Paper, 2008-11.
- REARDON, S.-F., FIREBAUGH, G. (2002). Measures of multi-group segregation. *Sociological Methodology*, Vol 32, S. 33-67.
- SILTANEN, J., JARMAN, J. AND BLACKBURN, R.-M. (1995). Gender Inequality in the Labour Market: Occupational Concentration and Segregation, *Geneva: International Labour Office*.
- TAEUBER, A-F., TAEUBER, K.-E. (1965). Negroes in Cities: Residential Segregation and Neighborhood Change *Chicago: Aldine Publishing Company*.
- THEIL, H. (1972). Statistical decomposition analysis, *Amsterdam: North Holland Publishing Company*, Vol. 14.
- THEIL, H., FINEZZA, A. J. (1971). A note on the measurement of racial integration of schools by means of informational concepts, *Journal of Mathematical Sociology*, Vol. 14, S. 187-194.
- TREIMAN, D. J. (1977). Occupational prestige in comparative perspective, Academic Press, New York.
- WAGNER, G. G., FRICK, J. R., SCHUPP, S. (2007). The German Socio-Economic Panel Study (SOEP) - Scope, Evolution and Enhancements, *Schmollers Jahrbuch* 127, no. 1, 139-169.
- ZOLOTH, B. S. (1976). Alternative measures of school segregation, *Land Economics*, Vol. 52 No.3, S. 278-298.