

**Daten über die personelle Einkommensverteilung -
Was man beachten sollte**

- 20. März 2009 -

Datenbank Makroindikatoren des Graduiertenkolleg SOCLIFE
Arbeitspapier Nr. 3

Dipl. Volksw. Hawal Shamon
Wissenschaftlicher Mitarbeiter
Lehrstuhl für empirische Sozial- und Wirtschaftsforschung
WISO-Fakultät
Universität zu Köln
Herbert-Lewin-Str. 2
D-50931 Köln
shamon@wiso.uni-koeln.de

Prof. Dr. Hans-Jürgen Andreß
Lehrstuhl für empirische Sozial- und Wirtschaftsforschung
WISO-Fakultät
Universität zu Köln
Herbert-Lewin-Str. 2
D-50931 Köln
hja@wiso.uni-koeln.de

„Empirical work is a matter of choosing the definition of data which fits the theory one wants to test or apply.“ (Atkinson/Bourguignon 2000: 40)

1) Einleitung

Zahlreiche Studien der sozialwissenschaftlichen Disziplinen verwenden in ihren Forschungsarbeiten Daten über die personelle Einkommensverteilung.¹ Ungleichheitsdaten gehören zu der Gruppe der aggregierten Makroindikatoren, denen gemeinsam ist, dass sie auf der Grundlage von Daten auf der Mikroebene berechnet werden (cf. Scheuch 1966). Ungleichheitsdaten werden auf der Basis von Einkommensdaten von Individuen oder Personengruppen berechnet. Für gewöhnlich wird ihnen ein einjähriger Betrachtungszeitraum zu Grunde gelegt. Allerdings können sich Einkommensdaten in anderen konzeptionellen Aspekten erheblich unterscheiden. Bei Einkommensdaten gilt es zu beachten, wer als Einkommensempfängereinheit definiert ist, wie das Einkommen konzeptualisiert ist und ob die erhobenen Einkommen an die unterschiedlichen Haushaltsstrukturen angepasst werden. Jeder konzeptionelle Aspekt der Einkommensdaten wirkt sich in unterschiedlich starkem Maße auf die Ungleichheitsdaten aus, so dass Ungleichheitswerte eines Landes zu einem bestimmten Jahr in Abhängigkeit der konzeptionellen Ausgestaltung der Einkommensdaten variieren.

Aus diesem Kontext ergeben sich für den Umgang mit Verteilungsdaten zwei Implikationen, die von Wissenschaftlern idealerweise befolgt werden sollten. Zum einen sollten stets Ungleichheitsdaten verwendet werden, deren zugrundeliegende Einkommensdaten in ihrer Konzeption im Einklang mit der zu bearbeitenden Forschungsfrage stehen. Wissenschaftliche Abhandlungen, die sich mit der Leistungsgerechtigkeit beschäftigen, erfordern eine andere konzeptionelle Ausgestaltung der den Verteilungsdaten zu Grunde liegenden Einkommensdaten als Forschungsarbeiten, die sich mit der wirtschaftlichen Wohlfahrt oder aber der sozialen Schichtung auseinandersetzen. Zum anderen sollten für wissenschaftliche Analysen ausschließlich Ungleichheitsdaten verwendet werden, die auf identisch konzipierten Einkommensdaten beruhen. Die Verwendung von Verteilungsdaten, die auf unterschiedlich konzipierten Einkommensdaten beruhen, birgt die Gefahr, Fehlschlüsse in der empirischen Analyse zu erliegen, die einzig auf konzeptbedingte Variationen in den Daten zurückzuführen sind. In der Praxis lassen sich die beiden Handlungsempfehlungen allerdings nicht ohne Abstriche machen zu müssen umsetzen. Die

¹ Im Folgenden werden für die Daten der personellen Einkommensverteilung die Begriffe Verteilungsdaten und Ungleichheitsdaten synonym verwendet.

weltweit verfügbaren Ungleichheitsdaten sind in Bezug auf die ihnen zu Grunde liegenden Einkommensdaten bis dato sehr heterogen. Befolgt man die Handlungsempfehlungen, kann sich die Zahl der verfügbaren Ungleichheitswerte je nach Konzeption so stark verringern, dass die Quer- bzw. Längsschnittprepräsentativität einer international vergleichenden Untersuchung in Frage gestellt werden kann.

Diese Arbeit ist Teil eines Datenbankprojektes des *Graduiertenkolleg Sozialordnung und Lebenschancen im internationalen Vergleich* der Wirtschafts- und Sozialwissenschaftlichen Fakultät der *Universität zu Köln*, in dem diverse quantitative Makroindikatoren für ausgewählte institutionelle Rahmenbedingungen auf nationaler Ebene über einen längeren Zeitraum erfasst werden sollen. Die Einkommensverteilung ist ein Makroindikator der Datenbank, welcher für unterschiedliche Forschungszwecke Verwendung finden soll. In Anbetracht der konzeptionellen Vielfalt der Einkommensdaten richtet sich das Papier an jene Personen, welche auf bestehende Ungleichheitsdaten zurückgreifen möchten. Es soll für sie die konzeptionelle Komplexität der den verfügbaren Verteilungsdaten zu Grunde liegenden Einkommensdaten reduzieren und damit eine Hilfestellung bei der Auswahl von Verteilungsdaten bieten. Um diesem Anspruch gerecht zu werden, wird zunächst die Berechnung eines der am meisten gebrauchten statistischen Maßzahlen für Ungleichheitsdaten – des Gini-Koeffizienten – erläutert (Abschnitt 2). Im darauffolgenden Abschnitt 3 wird der Leser mit den verschiedenen Einkommenskonzepten, den verschiedenen Untersuchungseinheiten und dem Zweck einer Bedarfsgewichtung vertraut gemacht. Darüber hinaus wird in Abschnitt 3 auf die Repräsentativität der Verteilungsdaten eingegangen. Nachdem im 4. Abschnitt die derzeit verfügbaren Quellen für Verteilungsdaten vorgestellt werden, die eine international vergleichende Untersuchung ermöglichen, wird im 5. Abschnitt der Aufbau und im Abschnitt 6 der Umgang mit der Datenbank des Graduiertenkollegs SOCLIFE geschildert. Von einer Diskussion der Problematiken die bei einer Einkommenserhebung bestehen, sowie den sich hieraus ergebenden Messfehlern, wird in diesem Papier Abstand genommen. Hierfür wird an dieser Stelle auf Gottschalk und Huynh (2006) verwiesen. Abschnitt 7 liefert eine Zusammenfassung.

2) Maßzahlen zur Messung der Einkommensungleichheit

In der Sozialwissenschaft existiert eine Fülle an statistischen Maßzahlen, die für die Berechnung von Einkommensungleichheiten verwendet werden können. Neben den sogenannten Quantilverhältnissen, welche sich lediglich aus dem Verhältnis zweier

Einkommensquantile ergeben, gibt es sogenannte integrale Maße.² Diese integralen Maßzahlen haben gegenüber den Verhältniszahlen den Vorteil, dass sie in ihrer Berechnung alle Einkommensquantile bzw. alle Einkommensklassen berücksichtigen. Für den Wissenschaftler entfällt bei der Verwendung eines integralen Maßes damit der Rechtfertigungszwang für die Auswahl bestimmter Einkommensquantile. Darüber hinaus lässt sich vom methodischen Standpunkt her argumentieren, dass ein integrales Maß für die Messung der gesamtgesellschaftlichen Einkommensungleichheit valider erscheint („face validity“) als das bloße Verhältnis zweier Quantile zueinander. Zu den integralen Maßzahlen zählen beispielsweise: das Atkinson-Maß, die Hoover-Ungleichverteilung, der Theil-Index und der Gini-Koeffizient. Jedes Maß bietet Vor- und Nachteile, mit denen sich ein Wissenschaftler zunächst vertraut machen sollte, bevor er sich für die Verwendung einer bestimmten Maßzahl entschließt.

Im Kontext der international vergleichenden Forschung sind einem Wissenschaftler, welcher auf existierende Einkommensungleichheitsdaten zurückgreifen möchte, bei der Auswahl der passenden Maßzahl allerdings Grenzen gesetzt. In diesem Fall bestimmt die Datenverfügbarkeit die Auswahlmöglichkeiten eines Wissenschaftlers. Die international am häufigsten verwendete Maßzahl für Einkommensungleichheit ist der Gini-Koeffizient. Daher wird im Folgenden zunächst auf den Wertebereich sowie auf die Interpretation des Gini-Koeffizienten eingegangen. Anschließend werden die Berechnung des Gini-Koeffizienten und eine mögliche grafische Veranschaulichung dargestellt.

2.1 Gini-Koeffizient: Zielsetzung

Der normierte Gini-Koeffizient kann beliebige Werte zwischen **0** und **1**, bzw. **0** und **100** Prozent annehmen. Während der Wert **0** eine vollkommene Gleichverteilung in der Gesellschaft bedeutet, bedeutet der Wert **100** eine vollkommene Ungleichverteilung des in einer Volkswirtschaft gemeinsam erwirtschafteten Einkommens aller natürlichen Personen.

² Ein Quantil ist ein Maß zur Beschreibung von Verteilungen. Mit Quantilen kann man aufsteigend geordnete Erhebungseinheiten in prozentuale Abschnitte unterteilen und den jeweiligen Merkmalswert an den Abschnitten ermitteln (vgl. Mosler/Schmid 2004: 31ff.). Je nachdem, wie viele Teilabschnitte für die Aufteilung einer Verteilung gewählt werden, ergeben sich Dezile (10%), Quintile (20%) oder Quartile (25%). Bei Quantilen der Einkommensverteilung interessiert man sich für den prozentualen Anteil, den ein bestimmter Prozentanteil der Erhebungseinheiten am insgesamt erwirtschafteten Einkommen der Volkswirtschaft hat. So hatten in Deutschland beispielsweise im Jahr 2003 die 10% einkommensschwächsten Haushalte (1. Dezil) lediglich einen Anteil von 2,7% am Gesamteinkommen Deutschlands (vgl. Deckl 2006: 21). Setzt man nun den Anteil der 10% einkommensstärksten Haushalte am gesamtwirtschaftlichen Einkommen Deutschlands in 2003 (25%) in das Verhältnis zu dem Anteil der 10% einkommensschwächsten Haushalte, ergibt sich die Verhältniszahl 9,26 (=25,7%/2,7 %). Auf EU-Ebene wird gemäß des im Jahr 2001 in Laeken (Belgien) gefassten Beschlusses des Europäischen Rates die Einkommensverteilung als das P80/P20 Quintilverhältnis gemessen. Die Einkommensverteilung wird hier als einer von mehreren Indikatoren für den sozialen Zusammenhalt einer Gesellschaft verwendet (vgl. Krause/Ritz: 2006).

Zur Veranschaulichung dient folgendes Beispiel: Man stelle sich einen Kuchen vor, welcher auf einer Festgesellschaft auf Personen aufgeteilt wurde. Der Kuchen symbolisiert hierbei das aggregierte Einkommen, welches von allen Personen einer Volkswirtschaft zusammen erwirtschaftet wurde. Die fünf Personen der Festgesellschaft repräsentieren die Bevölkerungsgröße der Volkswirtschaft. Haben alle fünf Personen gleich große Stücke des Kuchens erhalten - also jeweils ein Fünftel des Kuchens, herrscht eine vollkommene Gleichverteilung in der betrachteten Festgesellschaft und der Gini-Koeffizient hat den Wert 0. Im Gegensatz hierzu würde der normierte Gini-Koeffizient den Wert 1 haben, wenn eine der fünf Personen den gesamten Kuchen erhalten hätte und die restlichen vier Personen leer ausgehen würden.

2.2 Gini-Koeffizient: Berechnung

Die Berechnung des Gini-Koeffizienten wird aus didaktischen Gründen an einem Fallbeispiel mit diskreten Einkommenswerten demonstriert.³ Im Anschluss an die Berechnung des Gini-Koeffizienten wird auf die grafische Darstellung der Ungleichverteilung eingegangen, um einen Schwachpunkt des Gini-Koeffizienten zu verdeutlichen. Für die Berechnung des Gini-Koeffizienten stelle man sich vor, dass man das monatliche Einkommen von fünf Personen erhebt ($n = 5$). Person 1 verdient 5 Geldeinheiten, Person 2 verdient 7 Geldeinheiten, Person 3 verdient 1 Geldeinheit, Person 4 verdient 10 Geldeinheiten und Person 5 verdient 2 Geldeinheiten. Der Betrag, den die einzelne Person i ($i = 1, 2, \dots, n$) verdient, stellt hierbei den Merkmalswert x_i dar. Das Gesamteinkommen der Gruppe (Merkmalssumme) ergibt sich aus der Addition der Merkmalswerte $x_n = x_1 + x_2 + \dots + x_5$ und beträgt 25 Geldeinheiten.

In einem ersten sowohl für die Berechnung als auch für die grafische Darstellung elementaren Schritt, werden die fünf Personen nach der Größe ihrer Merkmalswerte x_i aufsteigend geordnet, wobei eine zusätzliche Nullzeile hinzugefügt wird. Aus $x_1 = 5$, $x_2 = 7$, $x_3 = 1$, $x_4 = 10$ und $x_5 = 2$ erhält man dann die geordnete Folge $x_{<j>}$ ($j = 0, 1, \dots, n$) mit $x_{<0>} = 0$, $x_{<1>} = 1$, $x_{<2>} = 2$, $x_{<3>} = 5$, $x_{<4>} = 7$ und $x_{<5>} = 10$. In einem zweiten Schritt werden dann die Summen $p_j = x_{<0>} + \dots + x_{<j>}$ der jeweils ersten j Werte der geordneten Folge errechnet. In einem dritten Schritt gewichtet man die geordneten Merkmalswerte mit ihren Rangplätzen und erhält somit $q_j = j * x_{<j>}$, um anschließend die Summe $q_n = \sum_{j=1}^n q_j$ zu ermitteln.

³ Für gewöhnlich weisen Einkommenserhebungen stetig klassierte oder kontinuierliche Einkommensinformationen auf, so dass für die Berechnung des Gini-Koeffizienten eine den kontinuierlichen Einkommensinformationen angepasste Formel verwendet wird.

Tabelle 1: Tabelle für die Berechnung des Gini-Koeffizienten und die Zeichnung der Lorenzkurve

Vorliegende Informationen		1. Schritt	2. Schritt	3. Schritt	4. Schritt	5. Schritt	
i	x _i	j	x _{<j>}	p _j (= x _{<1>+...+ x_{<j>})}	q _j (= j*x _{<j>})	u _j (= j/n)	v _j (= p _j /p _n)
		0	0	0	0	0	0
1	5	1	1	1	1	0,2	0,04
2	7	2	2	3	4	0,4	0,12
3	1	3	5	8	15	0,6	0,32
4	10	4	7	15	28	0,8	0,6
5	2	5	10	25	50	1	1
Σ	25		25		98		

Der Gini-Koeffizient G errechnet sich nun unter Verwendung folgender Formel:

$$G = \frac{1}{n} * \left(\frac{2 * q_n}{p_n} - 1 \right) - 1$$

$$\Leftrightarrow G = \frac{1}{5} * \left(\frac{2 * 98}{25} - 1 \right) - 1$$

$$\Leftrightarrow G = 0,368 \quad , \text{ mit } G \left[0; 1 - \frac{1}{n} \right].$$

Da der Gini-Koeffizient G nur Werte zwischen 0 und $G_{max} = 1 - \frac{1}{n}$ annimmt, wird anstelle von G meist der normierte Gini-Koeffizient G^* verwendet, der stets zwischen 0 und 1 liegt

$$G^* = \frac{G}{G_{max}} = \left(1 + \frac{1}{n-1} \right) * G$$

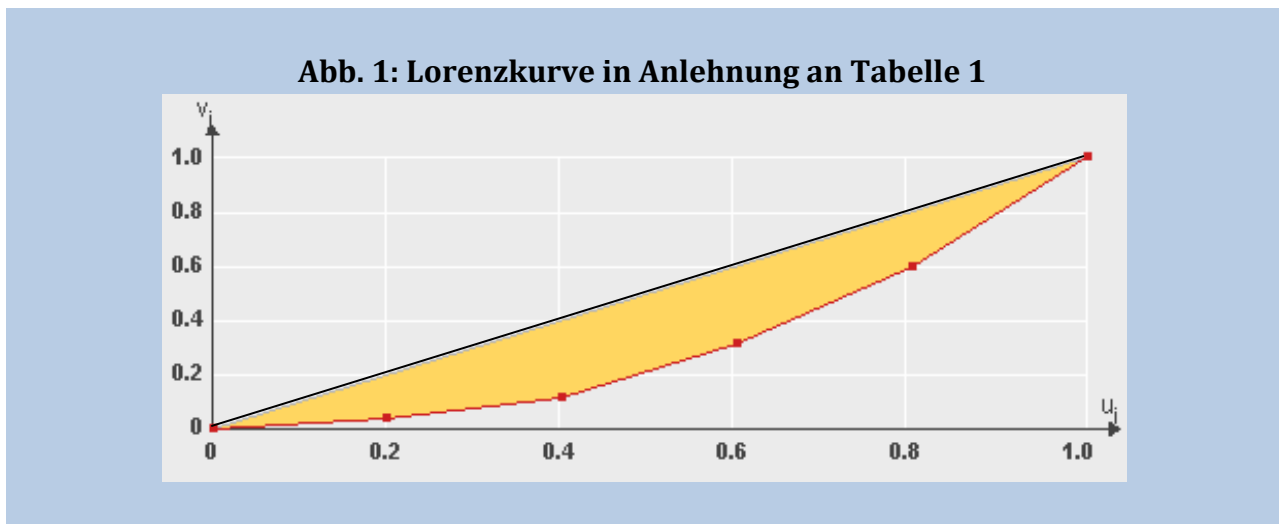
$$\Leftrightarrow G^* = \left(1 + \frac{1}{4} \right) * 0,368$$

$$\Leftrightarrow G^* = 0,46 \quad , \text{ mit } G^* \left[0; 1 \right].$$

2.3 Gini-Koeffizient: Graphische Darstellung

Grafisch lässt sich der Gini-Koeffizient mittels der Lorenzkurve veranschaulichen. Ein kurzer Exkurs ist an dieser Stelle sinnvoll, da er ein Manko des Gini-Koeffizienten zum Vorschein bringen wird, welches in wissenschaftlichen Abhandlungen bei der Verwendung des Gini-

Koeffizienten häufig vernachlässigt wird. Die Lorenzkurve ist ein grafisches Instrument zur Bestimmung der relativen Konzentration (vgl. Atkinson/Bourguignon 2000: 30). Sie ergibt sich aus der linearen Verbindung der Koordinaten (u_j, v_j) (s. Tabelle 1, Schritt 4 und 5). Die x – Koordinate $u_j = \frac{j}{n}$ steht hierbei für den Anteil der Personen der geordneten Folge j an der Gesamtzahl aller zugrunde gelegten n Personen. Mit $j = 0, 1, \dots, n$ ergeben sich im diskreten Fall damit $n + 1$ Koordinaten. Die y – Koordinate v_j wird als kumulierte relative Merkmalssumme bezeichnet. Hierunter versteht man mit $v_j = \frac{p_j}{p_n}$ den Anteil, den die jeweilige Summe p_j der jeweils ersten j Werte der geordneten Folge an der Summe p_n hat. Nach erfolgter Berechnung der Koordinaten in den Schritten 4 und 5 der Tabelle 1 ergibt sich die folgende grafische Darstellung der Lorenzkurve (vgl. Abbildung 1).⁴



Die Strecke zwischen dem Nullpunkt und der Koordinate $(1, 1)$ stellt die Hauptdiagonale dar, während die Lorenzkurve die in Tabelle 1 errechneten Koordinaten (u_j, v_j) linear verbindet. Bei vollkommener Gleichverteilung verläuft die Lorenzkurve, wie in Abbildung 2a dargestellt, entlang der Hauptdiagonalen. Würde nur eine Person die gesamten 25 Geldeinheiten verdienen und die restlichen vier Personen nichts, verläuft die Lorenzkurve wie in Abbildung 2b dargestellt, bis zu $u_j = n - \frac{1}{n}$ entlang der x – Achse und von dort bis zur Koordinate $(1, 1)$. Ergo verläuft die Lorenzkurve umso konvexer, je ungleicher die Verteilung ist. Die grau schraffierte Fläche A zwischen der Lorenzkurve und der

⁴ Die grafische Darstellung der Lorenzkurve basiert auf einem Internet Applet der Projektgruppe „Neue Statistik“, Hagen, welches über die ULR <http://www.fernuni-hagen.de/newstatistics/applets/Lorenzkurve/Lorenzkurve.htm> abgerufen werden kann.

Hauptdiagonalen kann als ein grafisches Maß der Ungleichheit verstanden werden (vgl. Bomsdorf 2002: 73). In Abbildung 1 entspricht ihr Anteil an der gesamten Fläche B unterhalb der Hauptdiagonalen (das Dreieck mit den Koordinaten $(0, 0)$, $(1, 0)$, $(1, 1)$) dem oben errechneten Gini-Koeffizienten G . Des Weiteren gibt Abbildung 1 zu erkennen, dass weit weniger als ein Fünftel (16%) des gesamten Einkommens auf die ersten beiden Personen fällt, während der Anteil der höchstverdienenden Person am Gesamteinkommen 40% beträgt.

Abb. 2a: Vollkommene Gleichverteilung

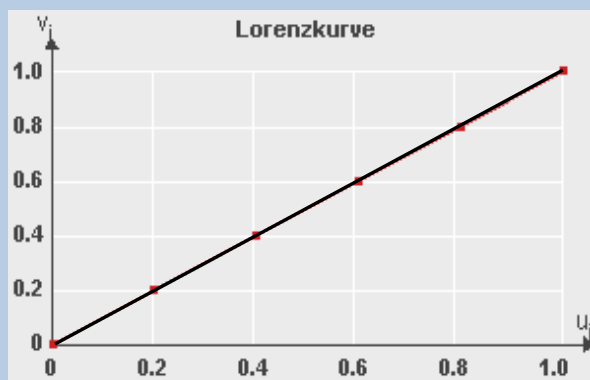
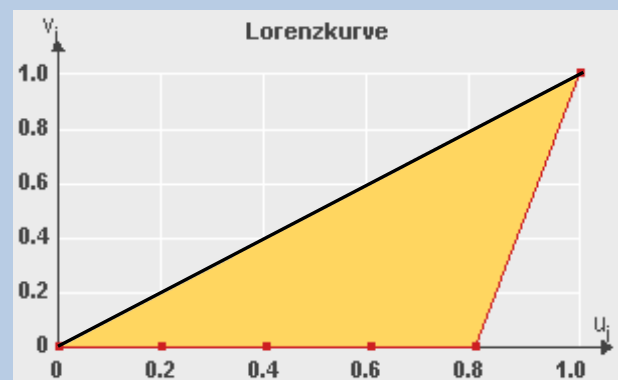


Abb. 2b: Vollkommene Ungleichverteilung



Ein Vergleich der Abbildungen 3a und 3b zeigt nun das Problem des Gini-Koeffizienten. Derselbe Gini-Koeffizient kann Ausdruck zweier unterschiedlicher Verteilungen sein. Für beide Verteilungen beträgt der Gini-Koeffizient $G = 0,263$ und der normierte Gini-Koeffizient $G^* = 0,329$ (s. Anhang Punkt 1 und 2).⁵ Aus diesem Grunde empfiehlt die statistische Literatur neben dem Gini-Koeffizienten immer auch die Lorenzkurve mit heranzuziehen, um eine eindeutige Aussage hinsichtlich der Ungleichheit treffen zu können (vgl. Bomsdorf 2001: 73). In vielen wissenschaftlichen Abhandlungen, welche Einkommensungleichheitsdaten verwenden, bleibt diese Empfehlung jedoch unberücksichtigt.

⁵ Für ein besseres Größenverständnis des Gini-Koeffizienten sei an dieser Stelle erwähnt, dass sich für die Verteilung der deutschen Haushaltsnettoeinkommen im Jahr 2003 ein normierter Gini-Koeffizient mit dem Wert 0,344 ergeben hat (vgl. Deckl 2006: 21).

Abb. 3a: Lorenzkurve bei Konzentration des Einkommens bei der fünften Person

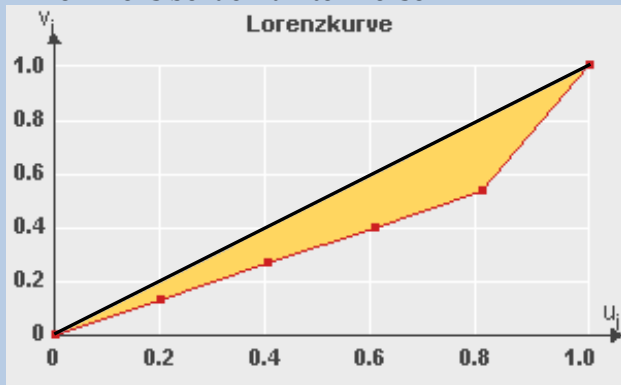
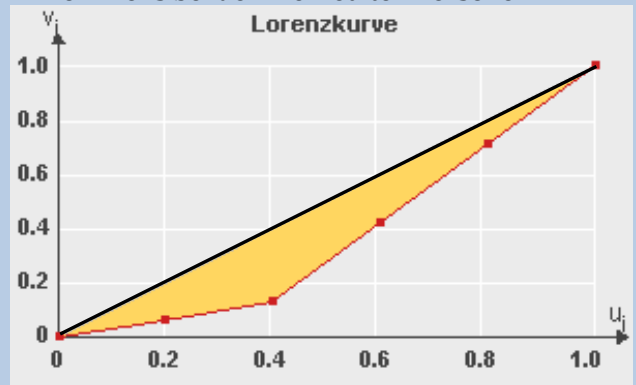


Abb. 3b: Lorenzkurve bei Konzentration des Einkommens bei den vier letzten Personen



3) Probleme der Vergleichbarkeit der Verteilungsdaten

Ein weiteres Problem im Umgang mit international verfügbaren Gini-Koeffizienten besteht in der internationalen und intertemporalen Vergleichbarkeit der Verteilungsdaten (vgl. World Institute for Development Economics Research 2008). Das Problem entspringt aus der Tatsache, dass Verteilungsdaten auf der Basis von Einkommensdaten berechnet werden, die sich in konzeptionellen und repräsentativen Aspekten erheblich unterscheiden können. Hinter dem repräsentativen Aspekt steht die Frage, ob eine Einkommenserhebung die Bevölkerung eines Landes repräsentiert. Eine Einkommenserhebung ist bevölkerungsweit nicht repräsentativ, wenn gewisse Bevölkerungsgruppen, wie z.B. Rentenempfänger, Erwerbstätige in Selbstständigkeit oder die ländliche Bevölkerung von der Untersuchungspopulation ausgeschlossen sind. Hinsichtlich der konzeptionellen Aspekte ist neben der Frage nach der Einkommensdefinition die Frage nach der Empfängereinheit – sind es einzelne Personen, Familien oder Haushalte – der Einkommen notwendig. Darüber hinaus werden Einkommen, die unter anderem in Familien oder Haushalten gemeinsam konsumiert werden, gelegentlich mittels einer sogenannten Bedarfsgewichtung um die unterschiedliche Haushaltstruktur bereinigt. Dies ist insbesondere dann der Fall, wenn eine empirische Untersuchung auf die Beantwortung einer wohlfahrtsökonomischen Fragestellung abzielt.

Die Berücksichtigung der auf der Mikrodatenebene angesiedelten konzeptionellen und repräsentativen Aspekte ist beim Umgang mit Ungleichheitsdaten unerlässlich, da sie die Ungleichheitsmaße auf der Makroebene beeinflussen. Atkinson und Brandolini (2001) zeigen, dass sowohl das Niveau als auch der Trend der Verteilung durch die Datenauswahl beeinflusst werden. Eine Missachtung der Heterogenität der Einkommensdaten birgt das Risiko von Fehlschlüssen in der empirischen Analyse, die sich allein durch konzeptbedingte

Variationen in den Daten ergeben können (vgl. Gobbin/Rayp 2004). Somit ergeben sich für den Forscher aus den konzeptionellen und repräsentativen Unterschieden der den Ungleichheitsdaten zugrunde liegenden Einkommensdaten zwei Implikationen im Umgang mit Verteilungsdaten: Zum einen sollten stets Verteilungsdaten verwendet werden, deren zugrunde liegenden Einkommensdaten im Einklang mit der zu bearbeitenden Forschungsfrage stehen (Validitäts-Argument). Zum Anderen sollte darauf verzichtet werden, auf Verteilungsdaten zurück zugreifen, die sich in den konzeptionellen und repräsentativen Aspekten der ihnen zugrunde liegenden Einkommensdaten unterscheiden (Homogenitäts-Argument). Im Folgenden werden zuerst die konzeptionellen Unterschiede erläutert (Abschnitt 3.1). Anschließend wird auf den Repräsentationsgrad eingegangen (Abschnitt 3.2).

3.1 Konzeptionelle Aspekte bei Einkommensdaten

Anders als bei Kennzahlen der volkswirtschaftlichen Gesamtrechnung oder etwa der Zahlungsbilanz hielt sich die länderübergreifende Verständigung über die Einführung konzeptioneller Standards bei der Erhebung von Einkommensdaten lange Zeit in Grenzen. Mit der *Canberra Group on Household Income Statistics* hat im Jahr 2001 erstmals eine Expertenrunde, welche unter anderem die Verbesserung der internationalen Vergleichbarkeit von Einkommensungleichheitsdaten zum Ziel hatte, gewisse Empfehlungen für die Konzeption von Haushaltseinkommensstatistiken ausgesprochen.⁶ Im Folgenden werden die wesentlichen konzeptionellen Aspekte, in denen sich Einkommensdaten unterscheiden können, in komprimierter Weise dargestellt (ausführlich s. Canberra Group 2001).

3.1.1 Einkommensdefinition

Im Allgemeinen lassen sich die verfügbaren Gini-Koeffizienten danach unterscheiden, ob sie auf der Grundlage des Gesamt-, des Netto- oder des verfügbaren Einkommens berechnet wurden.⁷ Der Bilanzierungszeitraum beträgt hierbei für gewöhnlich ein Jahr. Da jedoch die Konzeptualisierung der Einkommen lange Zeit ohne jegliche Verständigung von zahlreichen Wissenschaftlern und Forschungsinstituten in mehr oder weniger eigener Regie vorgenommen wurde, ist nicht auszuschließen, dass gleichnamige Einkommenskonzepte verschiedene Einkommensquellen in Betracht ziehen. So kann es beispielsweise sein, dass ein Wissenschaftler A bei der Konzeptualisierung des Gesamteinkommens unter anderem alle Einnahmen aus Vermögen mit berücksichtigt, während ein Wissenschaftler B das

⁶ Diese Empfehlungen richten sich sowohl an jene Forscher, welche Einkommensdaten erheben, als auch an jene, welche auf verfügbare Verteilungsdaten zurückgreifen. Sie weisen allerdings keinen Verbindlichkeitscharakter auf.

⁷ Neben einkommensbasierten Gini-Koeffizienten lassen sich auch konsumausgabenbasierte Gini-Koeffizienten finden. Diese sind allerdings zahlenmäßig selten und mit Problemen behaftet, die an dieser Stelle nicht erörtert werden können. Die konsumausgabenbasierten Gini-Koeffizienten werden im Folgenden nicht berücksichtigt.

Gesamteinkommen ausschließlich über Einnahmen aus Erwerbstätigkeit definiert. Unterschiede in den berücksichtigten Einkommenskomponenten sowie in den berücksichtigten direkten und indirekten Steuern können das Maß an Ungleichheit verzerren (vgl. Smeeding et al. 1993, Gardiner et al. 1995, Atkinson et al. 1995). Für die Beurteilung der Validität sowie der Vergleichbarkeit der Gini-Koeffizienten ist es daher unabdinglich, die den Gini-Koeffizienten zugrunde gelegten Einkommenskonzepte stets auf ihre Einkommenskomponenten hin zu überprüfen. Hierfür ist es hilfreich, auf das von der Canberra Group (2001) entwickelte Rahmenwerk für Einkommenskonzepte zurückzugreifen.

Die Canberra Group (2001) hat bei der Definition der aufeinander aufbauenden Einkommenskonzepte all jene Einkommenskomponenten berücksichtigt, welche ausschließlich die gegenwärtige ökonomische Situation einer Person beeinflussen. Demnach bleiben Einkommenskomponenten, welche die zukünftige ökonomische Situation einer Person beeinflussen, außer Betracht.⁸ Die Einkommenskomponenten lassen sich in Komponenten monetärer und nicht-monetärer Art unterscheiden und sind in Tabelle 2 zusammengefasst.

⁸ Beispiele für Einkommenskomponenten, welche die zukünftige ökonomische Situation einer Person beeinflussen können, sind Arbeitgeberzahlungen für die Altersversorgung und andere Formen der Sozialversicherung sowie Zinseinkommen aus Rentenkapital und Rentenvermögensanlagen.

Tabelle 2: Einkommenskonzepte

1 Employee income

Cash or near cash

Cash wages and salaries

Tips and bonuses

Profit sharing including stock options

Severance and termination pay

Allowances payable for working in remote locations etc, where part of conditions of employment

Cash value of 'fringe benefits'

Employers' social insurance contributions

Goods and services provided to employee as part of employment package

2 Income from self-employment

Cash or near cash

Profit/loss from unincorporated enterprise

Royalties

In-kind, imputed

Goods and services produced for barter, less cost of inputs

Goods produced for home consumption, less cost of inputs

Income less expenses from owner-occupied dwellings

3 Rentals

Income less expenses from rentals, except rent of land

4 Property income

Interest received less interest paid

Dividends

5 Current transfers received

Social insurance benefits from employers' schemes

Social insurance benefits in cash from government schemes

Universal social assistance benefits in cash from government

Means-tested social assistance benefits in cash from government

Regular inter-household cash transfers received

Regular support received from non-profit making institutions such as charities

6 Total income (*sum of 1 to 5*)

7 Current transfers paid

Employers' social insurance contributions

Employees' social insurance contributions

Taxes on income

Regular taxes on wealth

Regular inter-household cash transfers

Regular cash transfers to charities

8 Disposable income (*6 less 7*)

9 Social transfers in kind received

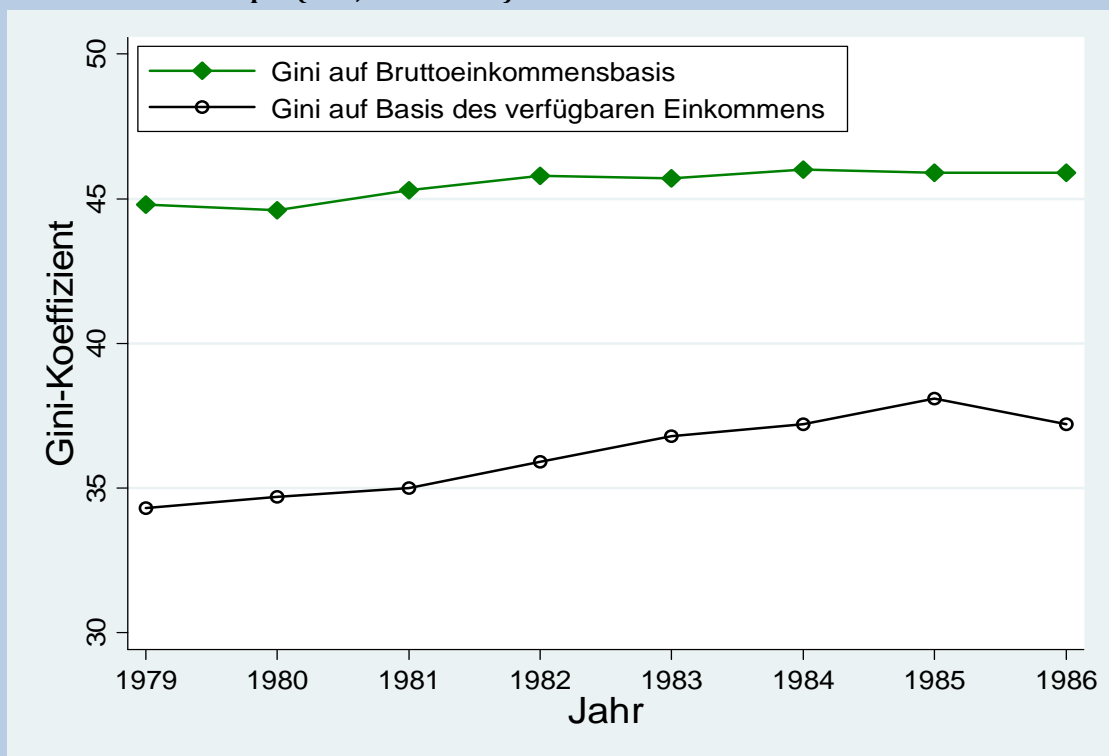
10 Adjusted disposable income (*8 plus 9*)

Quelle: Canberra Group 2001

Das **Gesamteinkommen** einer Empfängereinheit (Ziffer 6) ergibt sich aus der Addition der Einkommen aus abhängiger (Ziffer 1) und selbstständiger Beschäftigung (Ziffer 2) zuzüglich etwaiger Einkommen aus Mieteinkünften (Ziffer 3), Anlagevermögen (Ziffer 4) und empfangenen Transferzahlungen (Ziffer 5). Neben dem Begriff „Gesamteinkommen“ wird in der Literatur auch der Begriff „Bruttoeinkommen“ für Einkommen vor Steuerabzügen verwendet. Zieht man vom Gesamteinkommen der betreffenden Empfängereinheit die geleisteten Steuer- und Sozialversicherungsabgaben sowie die privaten Transferzahlungen (Ziffer 7) ab, erhält man das **verfügbare Einkommen** (Ziffer 8). Unter Berücksichtigung der Transfers, die einer Person als Sachleistungen zugehen (Ziffer 9), erhält man das **adjustierte verfügbare Einkommen** (Ziffer 10). Hierfür ist es allerdings erforderlich, den Wert der Sachleistungen zu bestimmen, was sich meist als ein schwieriges Unterfangen erweist. Das **Nettoeinkommen** einer Empfängereinheit lässt sich in Anlehnung an Tabelle 2 als das Einkommen definieren, welches nach Zahlung der Einkommens- und Vermögenssteuer sowie der vom Arbeitgeber und Arbeitnehmer zu zahlenden Pflichtbeiträge zur Sozialversicherung verbleibt (Summe der Ziffern aus 1, 2, 3 und 4 abzüglich Steuern und Sozialversicherungsbeiträge) (vgl. Becker/Hauser 2004: 58).

Abbildung 4 zeigt die Entwicklung der Einkommensungleichheit in den Vereinigten Staaten von Amerika auf der Grundlage von zwei unterschiedlichen Einkommenskonzepten (Bruttoeinkommen und verfügbares Einkommen) für den Zeitraum von 1979 bis 1986. Es sind Niveau- und Trendunterschiede in den Gini-Koeffizienten zu erkennen. Gemessen an den Gini-Koeffizienten auf der Basis des verfügbaren Einkommens ist die Verteilung der Einkommen in allen betrachteten Jahren deutlich ungleicher als die Ungleichheit vor Steuern und Transfers (Niveauunterschied). Dies kann zum einen mit dem Vorliegen eines progressiven Steuersystems erklärt werden, welches einkommensstärkeren Personen einen höheren Grenzsteuersatz zuschreibt als einkommensschwachen Personen. Zum anderen kann es mit staatlichen Transferleistungen erklärt werden, die den einkommensschwächeren Quantilen zugeführt werden. Des Weiteren zeigt sich, dass die Ungleichverteilung der Einkommen nach Steuern und Transfers im selben Zeitraum stärker zunimmt als die Ungleichverteilung der Bruttoeinkommen. Während der Gini-Koeffizient auf der Grundlage der verfügbaren Einkommen von 1979 bis 1986 um rund 3 Indexpunkte zunimmt (von 34,3 auf 37,2), nimmt die Ungleichheit der Bruttoeinkommen gemessen am Gini-Koeffizienten im selben Zeitraum nur um rund 1 Indexpunkt zu (von 44,8 auf 45,9). Vor dem Hintergrund der unterschiedlichen Trendverläufe ist davon auszugehen, dass es die steuerpolitischen Reformen Ronald Reagans zu Beginn der achtziger Jahre waren, die in den Vereinigten Staaten eine ungleichere Verteilungssituation bedingt haben.

Abbildung 4: Einkommensungleichheit (Gini) auf der Basis unterschiedlicher Einkommenskonzepte (USA, 1979-1986)



Quelle: World Income Inequality Database Version 2.c, eigene Berechnungen

Abbildung 4 verdeutlicht damit, dass bei der Auswahl der Verteilungsdaten stets auf den Einklang der zugrunde liegenden Einkommensdaten mit der zu bearbeitenden Forschungsfrage geachtet werden sollte (Validitäts-Argument). Interessiert man sich beispielsweise für den Einfluss der wirtschaftlichen Liberalisierung auf die Entwicklung der Löhne und Gehälter, ist es angebracht, Gini-Koeffizienten zu verwenden, welche auf der Grundlage des Bruttoeinkommens berechnet wurden. Diese Gini-Koeffizienten spiegeln die Verteilungssituation vor staatlicher Umverteilung wider. Interessiert man sich hingegen im international vergleichenden Kontext für die Verteilung der finanziellen Möglichkeiten der Gesellschaftsmitglieder in den betrachteten Ländern, so sollten Gini-Koeffizienten, welche auf dem verfügbaren Einkommen basieren, verwendet werden. Diese berücksichtigen die internationalen Unterschiede nach staatlicher und privater Umverteilung und erlauben damit einen direkten Vergleich der den Gesellschaftsmitgliedern zur Verfügung stehenden finanziellen Ressourcen.

Darüber hinaus lässt sich mit Hilfe von Abbildung 4 die Folge einer Missachtung des Homogenitäts-Arguments veranschaulichen. Die Verwendung von Verteilungsdaten, die sich in den konzeptionellen und repräsentativen Aspekten der ihnen zugrunde liegenden Einkommensdaten unterscheiden, lässt sich unter Verwendung der Daten aus Abbildung 4 beispielhaft wie folgt darstellen: Angenommen ein Forscher habe für die Jahre 1979, 1981, 1983 und 1985 die Gini-Koeffizienten auf der Basis des verfügbaren Einkommens herangezogen und für die restlichen Jahre (1980, 1982, 1984 und 1986) die Gini-

Koeffizienten auf Bruttoeinkommensbasis. Gemäß diesen ausgewählten Daten steigt die Einkommensverteilung in den USA von rund 34 (in 1979) auf rund 44 (in 1980) an, um dann auf einen Wert von rund 35 (in 1981) abzufallen, wobei sich die alternierende Folge bis in das Jahr 1986 fortsetzt. Es ist offensichtlich, dass diese Folge von Ungleichheitsdaten nicht der wahren Ungleichheitsentwicklung der USA entspricht. Vielmehr ist die Variation dieser Verteilungsdaten mit den Unterschieden in den konzeptionellen Aspekten der zu Grunde liegenden Einkommensdaten zu erklären. Würde ein Wissenschaftler die ausgewählten Verteilungsdaten für eine empirische Analyse verwenden, läuft er Gefahr, Fehlschlüsse zu ziehen.

3.1.2 Erhebungseinheit / Untersuchungseinheit

Neben dem Einkommenskonzept ist von Interesse, wessen Einkommen für die Berechnung der Gini-Koeffizienten herangezogen wurde und ob das Individuum, die Familie oder der Haushalt Gegenstand der Verteilungsanalyse ist. Hierfür ist es erforderlich, sowohl die für die Einkommenserhebung relevante Erhebungseinheit als auch die für die Ungleichheitsmessung relevante Untersuchungseinheit zu betrachten. Unter einer Erhebungseinheit versteht man jene Einheit, die einer Erhebung zu Grunde liegt (vgl. Friedrichs 1990: 126). In einer Einkommenserhebung kann nach dem Einkommen einer Person, einer Familie oder eines Haushaltes gefragt werden. Im ersten Fall nennt der Befragte sein persönliches Einkommen gemäß dem in der Einkommenserhebung zugrundegelegten Einkommenskonzept. Im zweiten und dritten Fall gibt der Befragte darüber hinaus das relevante Einkommen jener Personen an, deren Einkommen gemäß der jeweiligen Familien- bzw. Haushaltsdefinition für die Einkommenserhebung ebenfalls von Interesse sind.

Hiervon ist die Untersuchungseinheit strikt zu trennen, unter welcher man jene Einheit versteht, auf die sich eine Auswertung bezieht (vgl. Friedrichs 1990: 126). Innerhalb einer Verteilungsanalyse kann sich ein Forscher für die Ungleichheit von Personen, Familien oder Haushalten interessieren. Es ist dem Forscher nur dann möglich, eine von der Erhebungseinheit abweichende Untersuchungseinheit zu analysieren, wenn innerhalb einer Einkommenserhebung die für eine Umrechnung notwendigen Informationen mit abgefragt wurden. Greift beispielsweise ein an der persönlichen Einkommensverteilung interessierter Wissenschaftler auf repräsentative Einkommensdaten zurück, welche das Haushaltseinkommen widerspiegeln, so ist ihm die Ermittlung des persönlichen Einkommens der Einkommensbezieher nur möglich, wenn in der Einkommenserhebung explizit nach der Anzahl der Einkommensbezieher sowie der Höhe ihres jeweiligen Einkommens gefragt wurde. Obwohl die Wahl der Untersuchungseinheit Ausmaß und Entwicklung der Ungleichheit beeinflusst, wird diesem Problem zu wenig Beachtung geschenkt (vgl. Atkinson/Bourguignon 2000: 34). Es sollte daher darauf geachtet werden,

dass den verwendeten Gini-Koeffizienten stets jene Untersuchungseinheit unterliegt, welche im Einklang mit der Forschungsfrage steht.

Analysiert man den Arbeitsmarkt und ist hierbei die Verteilung der Markteinkommen von Interesse, so ist die Verwendung jener Gini-Koeffizienten angebracht, welche die Verteilung der Einkommen über die tatsächlichen Einkommensbezieher widerspiegeln. Demnach sollten in diesem Forschungskontext Gini-Koeffizienten verwendet werden, denen (Erwerbs)Personen als Erhebungs- und Untersuchungseinheiten zugrunde liegen (vgl. Atkinson/Bourguignon 2000: 35).

Für Untersuchungszwecke, die sich mit dem wirtschaftlichen Wohlergehen, der Kaufkraftverteilung oder der sozialen Schichtung auseinandersetzen, sollte von der Verwendung dieser Gini-Koeffizienten abgesehen werden (vgl. Canberra Group 2001: 32, vgl. Blümle 1975: 25). Realiter leben nicht alle Menschen in Single-Haushalten und in den meisten Fällen wird das Einkommen in Mehrpersonen-Haushalten mit jenen Menschen geteilt, mit denen man zusammen lebt. Wird in einer Einkommenserhebung beispielsweise das persönliche Erwerbseinkommen einer nicht arbeitenden Managergattin abgefragt, so wird sie angeben müssen, dass sie kein Einkommen bezieht. Dies spiegelt aber nicht ihr wahres wirtschaftliches Wohlergehen, ihre Kaufkraft oder sozialen Status wider, da sie einen rechtlichen Anspruch auf das Einkommen ihres Ehemannes hat.

Ist man am wirtschaftlichen Wohlergehen, an der Kaufkraftverteilung oder an der sozialen Schichtung interessiert, sollte man auf Gini-Koeffizienten zurückgreifen, welche auf der Grundlage des Einkommens der Familie oder des Haushaltes berechnet wurden (vgl. Hauser 2002: 175). Für die Erhebungseinheit Familie existieren unterschiedliche Definitionen. Man unterscheidet zwischen der „Kernfamilie“ und der „ökonomischen Familie“. Während bei der Kernfamilie nach dem gemeinsamen Einkommen jener Personen gefragt wird, die in einem Eltern-Kind-Verhältnis zueinander stehen, ist die ökonomische Familie weiter gefasst.⁹ Beide Familiendefinitionen schließen jedoch Personen der Einkommenserhebung aus, die das finanzielle Wohlergehen der Familienmitglieder beeinflussen können (vgl. Canberra Group 2001: 33). Im Falle der Kernfamilie ist es beispielsweise ein im Haushalt der Kernfamilie lebender Neffe, welcher mit seinem Einkommen die finanzielle Situation der Kernfamilie verbessert oder vom Einkommen der Kernfamilie mit versorgt wird. Im Falle der ökonomischen Familie kann es der noch nicht angeheiratete Lebenspartner eines der Familienmitglieder sein, der mit seinem Einkommen die finanziellen Möglichkeiten der ökonomischen Familie verbessert oder von deren Einkommen mit versorgt wird. Damit besteht auch hier die Möglichkeit, dass die finanzielle Situation der Familien fehlerhaft

⁹ Gemäß der Canberra Group wird die Kernfamilie in der Regel wie folgt definiert: Eltern und unverheiratete Kinder, die gemeinsam ein Gebäude teilen. In einigen Fällen liegt eine Altersbeschränkung für die Kinder vor. Die ökonomische Familie wird wie folgt definiert: Personen, die ein Gebäude teilen und blutsverwandt, angeheiratet, verschwägert oder adoptiert sind (vgl. Canberra Group 2001: 36f.).

erfasst wird. Die Canberra Group empfiehlt daher die Verwendung von Verteilungsdaten, welche auf der Berechnung des Haushaltseinkommens basieren. Unter einem Haushalt versteht die Canberra Group sowohl alleinstehende Personen als auch eine Gruppe von Personen, die gemeinsam eine Wohneinheit teilen (vgl. Canberra Group 2001: 34 und 38).¹⁰ Der Vorteil der Erhebungseinheit Haushalt besteht darin, dass sie jegliche Personenkonstellationen, die die gemeinsame finanzielle Stellung eines Haushalts zu beeinflussen vermögen, berücksichtigt. Allerdings kann die implizite Annahme, dass Personen einer Wohneinheit ungeachtet ihres verwandtschaftlichen Verhältnisses ihr Einkommen grundsätzlich teilen, nicht zutreffen.

Da sich Gruppen von Personen, egal ob sie als Familie oder Haushalt definiert werden, in Größe und Struktur unterscheiden können, sind Familien bzw. Haushaltseinkommen im Kontext einer Wohlfahrtsanalyse nicht ohne weitere Korrektur vergleichbar. Gini-Koeffizienten sollten daher auf der Grundlage von bedarfsgewichteten Familien- bzw. Haushaltseinkommen berechnet werden, die Größe und Struktur der Familien bzw. Haushalte berücksichtigen (siehe den folgenden Abschnitt 3.1.3).

Schließlich besteht die Möglichkeit, dass die einem Gini-Koeffizienten zugrunde liegende Untersuchungseinheit von der Erhebungseinheit abweicht. Wissenschaft und Politik sind üblicherweise am wirtschaftlichen Wohlergehen von Individuen und nicht am Wohlergehen von Haushalten interessiert (vgl. Atkinson et al. 1995: 21). In diesen Fällen ist es notwendig, auf Gini-Koeffizienten zurückzugreifen, die auf der Grundlage des bedarfsgewichteten Familien- oder Haushaltseinkommens der Individuen berechnet wurden (vgl. Blümle 1975:25 Hauser 2002: 175f. und Canberra Group 2001: 40). Hierbei wird jedem Mitglied einer Familie oder eines Haushalt das ermittelte bedarfsgewichtete Familien- oder Haushaltseinkommen zugewiesen. Die Berechnung basiert damit auf der Betrachtung von Individuen und nicht auf der Betrachtung von Familien bzw. Haushalten.

3.1.3 Bedarfsgewichtung und Berechnung von Äquivalenzeinkommen

Im Hinblick auf den internationalen Vergleich von Verteilungen der Wohlfahrt über Individuen, Familien oder Haushalte beeinträchtigt die Verwendung von Gini-Koeffizienten, welche nicht auf dem bedarfsgewichteten Pro-Kopf-Haushaltseinkommen basieren, die Vergleichbarkeit der Daten (vgl. Canberra Group 2001: 40). Die Bedarfsgewichtung des Haushaltseinkommens beruht auf vier Annahmen: 1. Das Einkommen aller Haushaltsmitglieder fließt in eine gemeinsame Haushaltskasse. 2. Das in der Haushaltskasse verfügbare Einkommen wird derart konsumiert, dass alle Haushaltsmitglieder das gleiche

¹⁰ Das statistische Bundesamt fasst den Haushaltsbegriff deutlich enger, indem sie den Haushalt als eine Gruppe von verwandten oder persönlich verbundenen (auch fremden) Personen bezeichnet, die sowohl Einkommens- als auch verbrauchsmäßig zusammengehören, i.d.R. zusammen wohnen und über ein oder mehrere Einkommen oder über Einkommensteile gemeinsam verfügen und voll oder überwiegend im Rahmen einer gemeinsamen Hauswirtschaft versorgt werden, wobei auch Einzelpersonen mit eigenem Einkommen, die für sich alleine wirtschaften als Haushalte gelten (vgl. Statistisches Jahrbuch 2007: 544).

Wohlfahrtsniveau erreichen. 3. Beim gemeinsamen Wirtschaften entstehen Einsparungen in Abhängigkeit der Haushaltsgröße (Skaleneffekte). 4. Kinder haben geringere Bedürfnisse als Erwachsene (vgl. Hauser 2002: 178). Vor dem Hintergrund dieser Annahmen werden die finanziellen Möglichkeiten des einzelnen Haushaltsmitglieds bei gegebenem Haushaltseinkommen maßgeblich durch die Haushaltsstruktur bestimmt, d.h. durch die Haushaltsgröße und die demographische Zusammensetzung des Haushalts. Die Bedarfsgewichtung erfolgt über die Verwendung einer sogenannten Äquivalenzskala. Mit dieser berechnet man aus dem Haushaltseinkommen das bedarfsgewichtete Haushaltseinkommen, für das auch der Begriff Äquivalenzeinkommen verwendet wird. In der Wissenschaft existieren verschiedene Äquivalenzskalen, welche die demographische Struktur der Haushalte in unterschiedlichem Maße in Betracht ziehen. Buhmann et al. (1988) identifizieren in zehn OECD-Ländern über 30 verschiedene Äquivalenzskalen. Einen Extrempol bildet hierbei jene Äquivalenzskala, nach der lediglich das Haushaltseinkommen berechnet wird, d.h. die also sowohl Haushaltsgröße als auch Skaleneffekte ignoriert. Den anderen Gegenpol bilden jene Äquivalenzskalen, welche allen vier Annahmen gerecht werden (vgl. Buhmann et al. 1988: 116).¹¹ Eine solche Skala ist z.B. die modifizierte Skala der Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (OECD), welche auch innerhalb der Europäischen Union für Vergleiche von Wohlfahrtspositionen verwendet wird (vgl. Dennis/Giuo 2004: 6). Sie weist dem Haupteinkommensbezieher im Haushalt einen Gewichtungsfaktor von 1,0 zu. Sonstige Personen ab 14 Jahren gehen mit dem Gewichtungsfaktor 0,5 und Personen unter 14 Jahren mit dem Faktor 0,3 in die Äquivalenzberechnung mit ein. Das Äquivalenzeinkommen wird ermittelt, indem das Haushaltseinkommen durch die Summe der Gewichte geteilt wird. Es wird, unabhängig davon, ob es sich um Erwachsene oder Kinder handelt, jedem Haushaltsmitglied zugewiesen. So beträgt das Äquivalenzeinkommen einer vierköpfigen Familie, welche aus einem Haupteinkommensbezieher (Gewichtungsfaktor: 1,0), einer Ehefrau (0,5) und Zwillingen im Alter von 12 Jahren (jeweils 0,3) besteht und über ein monatliches Haushaltseinkommen von 2100 Geldeinheiten verfügt, 1000 Geldeinheiten. Für ein kinderloses Ehepaar würde das bedarfsgewichtete Pro-Kopf-Einkommen unter sonst gleichen Bedingungen 1400 Geldeinheiten betragen.

Basiert der Gini-Koeffizient auf der Untersuchungseinheit Familie bzw. Haushalt, bedeutet dies, dass bei der Berechnung jeder Haushalt nur ein Mal mit dem entsprechenden bedarfsgewichteten Haushaltseinkommen berücksichtigt wird. Seine Verwendung erlaubt ausschließlich Aussagen über die Verteilung von Wohlfahrtspositionen für Familien bzw. Haushalte in einer Gesellschaft. Hingegen können Aussagen über die Verteilung der individuellen Wohlfahrtsposition getroffen werden, wenn der Gini-Koeffizient auf der

¹¹ Zwischen diesen beiden Extremen liegen jene Äquivalenzskalen, welche die Haushaltsgröße berücksichtigen, indem das Haushaltseinkommen durch die Anzahl der Haushaltsmitglieder geteilt wird, Skaleneffekte aber ebenfalls ignorieren.

Untersuchungseinheit Person beruht. Dies bedeutet, dass bei der Berechnung jedes einzelne Haushaltsmitglied mit dem entsprechenden bedarfsgewichteten Haushaltseinkommen berücksichtigt wird. Im Fall der vierköpfigen Familie würden die vier Familienmitglieder im Datensatz als unabhängige Personen jeweils mit einem Äquivalenzeinkommen von 1000 Geldeinheiten in die Berechnung eingehen.

Im nationalen Forschungskontext scheint die Verwendung von Gini-Koeffizienten, welche auf Basis unterschiedlicher Äquivalenzskalen berechnet wurden, zu geringfügigen prozentualen Abweichungen der Gini-Koeffizienten zu führen. Anhand einer Sensitivitätsanalyse zeigt Hauser (2002) unter Verwendung der *älteren OECD-Skala*, der *modifizierten OECD-Skala* sowie der *neuen US-Skala* am Beispiel Westdeutschlands, dass die Gini-Koeffizienten der Verteilung der bedarfsgewichteten Nettoeinkommen der Haushalte für die Jahre 1973 bis 1998 maximal um 3,4% voneinander abweichen.¹² Für die international vergleichende Forschung hingegen bedarf es einer differenzierteren Aussage. Buhmann et al. (1988) untersuchen die Sensitivität der Gini-Koeffizienten in Abhängigkeit unterschiedlicher Äquivalenzskalen für zehn Länder.¹³ Sie zeigen, dass die Wahl der Äquivalenzskala die auf der Größe der jeweiligen Gini-Koeffizienten basierende Rangreihe der Länder nicht beeinträchtigt, wohl aber die Werte der jeweiligen Gini-Koeffizienten in systematischer Weise beeinflusst (vgl. Buhmann et al. 1988: 128). Der Wert des Gini-Koeffizienten fällt umso größer aus, je stärker eine Skala die Skaleneffekte gewichtet. Beide Studien zeigen damit, dass die Wahl der Äquivalenzskala den Wert des Gini-Koeffizienten – wenn auch in geringer Weise – beeinflusst. Um Variationen in den Gini-Koeffizienten zu vermeiden, die einzig auf die unterschiedlichen Äquivalenzskalen zurückzuführen sind, sollte man daher stets auf Gini-Koeffizienten zurückgreifen, welche auf einer einheitlichen Äquivalenzskala basieren. Wie oben bereits dargestellt, werden allerdings viele unterschiedliche Äquivalenzskalen zur Berechnung der Äquivalenzeinkommen verwendet. Folglich kann diese Strategie dazu führen, dass die Anzahl an verfügbaren Gini-Koeffizienten für eine längsschnitts- bzw. querschnittsrepräsentative Analyse zu gering ist. Infolge dieses Zielkonfliktes scheint es angebracht, eine alternative Strategie zur Eindämmung der konzeptbedingten Datenvariation zu verfolgen. Diese kann darin bestehen, dass nur jene Gini-Koeffizienten verwendet werden, deren zugrunde liegende Äquivalenzskalen die Skaleneffekte in einem ähnlichen Maß gewichtet haben. Je unterschiedlicher das Maß an Gewichtung der Skaleneffekte innerhalb der Äquivalenzskalen ausfällt, desto größer ist die konzeptbedingte Variation. In jedem Fall aber gilt es, die den Gini-Koeffizienten zugrunde liegenden Äquivalenzskalen mit zu berichten und etwaige Auswirkungen auf die Ergebnisse in der kritischen Auseinandersetzung mit der eigenen Forschungsarbeit anzusprechen.

¹² Es werden sechs Zeitpunkte mit einem fünfjährigem Abstand betrachtet (1973, 1978, 1983, 1988, 1993, 1998). Im Durchschnitt beträgt die Abweichung der Gini-Koeffizienten lediglich 2,2%.

¹³ Untersucht werden die Länder Australien, Deutschland, Israel, Kanada, Niederlande, Norwegen, Schweden, Schweiz, Vereinigtes Königreich sowie die Vereinigten Staaten von Amerika.

Tabelle 3a: Überblick über die verschiedenen konzeptionellen Aspekte

Konzeptionelle Aspekte	Einkommensdefinition / Zugrundeliegendes Einkommenskonzept	<ul style="list-style-type: none"> • Bruttoeinkommen / Gesamteinkommen • Nettoeinkommen • Verfügbares Einkommen
	Erhebungseinheit/ Untersuchungseinheit	<ul style="list-style-type: none"> • Person • Familie • Haushalt
	Bedarfsgewichtung/ Berechnung von Äquivalenzeinkommen	<ul style="list-style-type: none"> • Keine Bedarfsgewichtung • Bedarfsgewichtung unter der Berücksichtigung der Haushaltsgröße • Bedarfsgewichtung unter der Berücksichtigung der Haushaltsgröße und der demographischen Struktur

3.2 Repräsentationsgrad

Neben den konzeptionellen Aspekten, in denen sich Einkommensungleichheitsdaten unterscheiden können, unterscheiden sich die verfügbaren Daten auch hinsichtlich ihrer Bevölkerungsrepräsentativität. So können bei der Festlegung der Stichproben für die Einkommenserhebungen gewisse Altersgruppen, geographische Regionen oder Bevölkerungsgruppen ausgeblendet worden sein. Grundsätzlich sollten Daten, sofern es das Forschungsvorhaben nicht anders erfordert bzw. festlegt, bevölkerungsweit repräsentativ sein. Liegen keinerlei repräsentative Daten vor, besteht die zweitbeste Lösung darin, jene Daten zu verwenden, welche in den gleichen Aspekten ein gleiches Maß an Repräsentativität aufweisen. Die schlechteste Lösung liegt in der Verwendung von Daten, die ein unterschiedliches Maß an Repräsentativität aufweisen. Es gilt also das Argument, konzeptbedingte Variationen in den Daten zu vermeiden. Werden beispielsweise die Einkommen eines Landes in einem Jahr ausschließlich in Städten erhoben, so kann der ermittelte Gini-Koeffizient dieses Jahres nicht als repräsentativ erachtet werden, wenn das Ausmaß der Einkommensungleichheit zwischen den städtischen und ländlichen Gebieten hoch erscheint. Dies ist häufig in Entwicklungsländern der Fall. Der Gini-Koeffizient würde das wahre Maß der Ungleichheit in dem betrachteten Jahr höchstwahrscheinlich unterschätzen, da die ökonomisch weitaus schlechter gestellte Landbevölkerung nicht berücksichtigt wurde. Die Unterschiede zu den repräsentativen Werten der anderen Jahre wären dann nicht mit dem Stichprobenfehler zu erklären, sondern wären systematischer Natur. Verwendet man diesen Gini-Koeffizienten zusammen mit den Gini-Koeffizienten der anderen Jahre für dieses Land, läuft man Gefahr, Fehlschlüsse in der empirischen Analyse zu erleiden. Es empfiehlt sich daher, jene Gini-Koeffizienten, welche sich in ihrem Maß der

Repräsentativität gegenüber den anderen Gini-Koeffizienten maßgeblich unterscheiden, von der Analyse auszuschließen.

Tabelle 3b: Überblick über die möglichen Ausprägungen bei repräsentativen Aspekten

Repräsentative Aspekte	Altersgruppen	<ul style="list-style-type: none"> • Gesamtbevölkerung • Bevölkerung im erwerbsfähigem Alter • Bevölkerung mit einem bestimmten Alter • Weitere Restriktionen möglich
	Geographische Regionen	<ul style="list-style-type: none"> • Gesamtbevölkerung • Stadtbevölkerung • Landbevölkerung • Weitere Restriktionen möglich
	Bevölkerungsgruppen	<ul style="list-style-type: none"> • Gesamtbevölkerung • Bevölkerung in abhängiger Beschäftigung • Beschäftigte des sekundären und tertiären Sektors • Weitere Restriktionen möglich

4) Verfügbare Datensätze mit Ungleichheitsdaten

Im Hinblick auf die international vergleichende Forschung der Einkommensverteilung haben in der Vergangenheit zwei Institutionen die Qualität, die Transparenz und damit die internationale Vergleichbarkeit der verfügbaren Verteilungsdaten durch ihre Arbeit verbessert. Zum einen ist es das *Luxembourg Income Study Project* (LIS) und zum anderen das *World Institute for Development Economic Research* (WIDER) der *United Nations University*. Das *Luxembourg Income Study Project* stellt sowohl Verteilungsindizes auf der Grundlage des verfügbaren Einkommens als auch unverarbeitete Einkommensdaten für 30 Länder bereit. Die Einkommensdaten stammen aus unterschiedlichen nationalen repräsentativen Einkommenserhebungen.¹⁴ Bei sechs Ländern liegen für alle sechs Wellen – die erste Welle beginnt im Jahr 1979 und die letzte Welle beginnt im Jahr 2004 – Verteilungsindizes sowie Einkommensdaten vor. Bei anderen Ländern hingegen liegen die Informationen nur in unregelmäßigen Abständen vor. Die Bereitstellung der unverarbeiteten Einkommensdaten durch das *Luxembourg Income Study Project* ermöglicht dem Forscher, ein seinem Forschungsinteresse entsprechendes Einkommenskonzept flexibel zu definieren.

¹⁴ Das LIS Database sowie die LIS Key Figures können über die folgende Internetseite abgerufen werden: <http://www.lisproject.org>, Stand: 11.09.2008.

Weitaus umfangreicher sowohl in der Querschnitts- als auch in der Längsschnittdimension ist die vom *World Institute for Development Economic Research* veröffentlichte *World Income Inequality Database Version 2.c* (WIID), für welche 5313 weltweit existierende Ungleichheitsdaten für insgesamt 159 Länder im Zeitraum von 1867 bis 2006 in einer Datenbank zusammen getragen wurden. Hier sind unter anderem auch die Ungleichheitsdaten zu finden, die von LIS veröffentlicht wurden.¹⁵ Neben der umfangreichen Quer- und Längsschnittdimension besteht das besondere Merkmal der *World Income Inequality Database* darin, dass jeder Ungleichheitswert gemäß einem Kriterienkatalog bewertet und mit einem Qualitätsindikator versehen wurde.

Die *Datenbank Makroindikatoren*, die für das *Graduiertenkolleg Sozialordnung und Lebenschancen im internationalen Vergleich* (GK SOCLIFE) der Wirtschafts- und Sozialwissenschaftlichen Fakultät der *Universität zu Köln* geschaffen wurde, enthält einen Großteil der Daten des *WIID Version 2.0c*. Im Rahmen dieser Datenbank werden allerdings nur jene Daten übernommen, welche auf der Betrachtung von Einkommen basieren und hinsichtlich der betrachteten Altersgruppen sowie bezüglich der geographischen Regionen bevölkerungsweit repräsentativ sind. Von einem Ausschluss von Gini-Koeffizienten, welche im Hinblick auf die Bevölkerungsgruppen nicht repräsentativ sind, wird Abstand genommen, da dies in systematischer Weise die Verfügbarkeit von Gini-Koeffizienten auf der Grundlage des Netto-Einkommens beeinträchtigen würde. Hinsichtlich sonstiger validitätsrelevanter Aspekte der Daten wird keine Vorauswahl getroffen. Weitere Selektionen bleiben also dem Datenbanknutzer überlassen. In der *Datenbank Makroindikatoren* sind 2335 Verteilungsdaten für 119 Länder in einem Zeitraum von 1948 bis 2006 zusammengefasst. An dieser Stelle sei darauf hingewiesen, dass für ein Land in einem Jahr mehrere Verteilungsdaten bei gleicher konzeptioneller Ausgestaltung vorliegen können. Dies ist darauf zurückzuführen, dass verschiedene Institutionen unabhängig voneinander Berechnungen durchführen, bei denen sie auf dieselbe Einkommenserhebung zurückgreifen. Liegen für ein Land mehrere Verteilungsdaten auf gleicher konzeptioneller Basis für ein Jahr vor, ist es für eine Analyse ratsam, Verteilungsdaten zu verwenden, deren veröffentlichte Quelle mit der Quelle übereinstimmt, die die Mehrheit der Verteilungsdaten für ein Land veröffentlicht hat.

5) Aufbau der Datenbank Makroindikatoren

Die *Datenbank Makroindikatoren* (DMI) weist die Ungleichheitskoeffizienten getrennt nach den ihnen zugrunde liegenden Einkommenskonzepten aus. Wie im Abschnitt 3.1.1 bereits

¹⁵ Das *World Income Inequality Database Version 2.c* kann über die folgende Internetseite abgerufen werden: http://www.wider.unu.edu/research/Database/en_GB/database/, Stand: 11.09.2008.

geschildert wurde, sind Einkommenskonzepte bis dato weltweit nicht einheitlich definiert. Es existiert seit 2001 hinsichtlich der für die einzelnen Einkommenskonzepte (Gesamt-/Bruttoeinkommen, Nettoeinkommen, verfügbares Einkommen, adjustiertes verfügbares Einkommen) zu berücksichtigenden Einkommenskomponenten lediglich eine unverbindliche Empfehlung der Canberra Group. Damit ist die Wahrscheinlichkeit bei verfügbaren Verteilungsdaten, welche auf gleichen Einkommenskonzepten beruhen, hoch, dass sie sich in einigen Einkommenskomponenten unterscheiden. Diesem Umstand kann bei der Erstellung der *DMI* allerdings keine Rechnung getragen werden. Für die *DMI* sind die Verteilungsdaten gemäß der ihnen zugrunde liegenden Einkommensdefinitionen auch dann in ein gleiches Einkommenskonzept eingeordnet, wenn sie sich in einzelnen Einkommenskomponenten unterscheiden. Der Nutzer sollte dies bei all seinen Analysen bedenken und es kommentieren.

Die Variable *gini_tot* enthält Gini-Koeffizienten, welche auf dem Gesamteinkommen basieren. Verteilungsdaten, welche auf der Betrachtung des Netto-Einkommens basieren, werden unter der Variablen *gini_net* geführt. In Anlehnung an die englische Bezeichnung des verfügbaren Einkommens (disposable income) weist die Variable *gini_dis* Gini-Koeffizienten auf der Grundlage des verfügbaren Einkommens aus.

Tabelle 4: Verteilungsdaten gemäß des zugrundeliegenden Einkommenskonzeptes in der DMI

	Beobachtungen	Länder	Zeitspanne
Gini-Koeffizienten auf der Grundlage des Bruttoeinkommens	1117	115	1948 bis 2006
Gini-Koeffizienten auf der Grundlage des Nettoeinkommens	117	14	1976 bis 2006
Gini-Koeffizienten auf der Grundlage des verfügbaren Einkommens	1102	51	1958 bis 2006

Die geschilderte Einteilung der Verteilungsdaten in Abhängigkeit ihrer Einkommenskonzepte, erfolgt unabhängig von den weiteren konzeptionellen Aspekten der den Verteilungsdaten zugrundeliegenden Einkommensdaten. D.h. beispielsweise für die Variable *gini_tot* im Hinblick auf den konzeptionellen Aspekt der Analyseeinheit, dass die Gini-Koeffizienten auf der Grundlage des individuellen Gesamteinkommens, des familiären Gesamteinkommens oder des Gesamteinkommens eines Haushaltes berechnet worden sein können. Die Informationen über die konzeptionellen Aspekte und den repräsentativen Eigenschaften sind in vier Variablen abgelegt, die im Folgenden als Informationsvariablen bezeichnet werden. Informationen über die Empfängereinheit sind in der Variablen *incsharu*

abgelegt, während die Analyseeinheit in der Variablen *anaunit* dokumentiert ist. Hinweise über eventuell durchgeführte Bedarfsgewichtungen sind in der Variablen *equisc* abgelegt. Angaben über die Untersuchungspopulation, die in der Einkommenserhebung betrachtet wurde, kann ein Datenbanknutzer der Variablen *popcovr* entnehmen. An dieser Stelle sei noch einmal betont, dass in der *DMI* alle Verteilungsdaten auf Einkommenserhebungen basieren, welche hinsichtlich der betrachteten Altersgruppen sowie bezüglich der geographischen Regionen bevölkerungsweit repräsentativ sind.

6) Handhabung der Datenbank

Wie soll man nun mit der Datenbank umgehen? Generell wird für den Umgang mit Verteilungsdaten empfohlen, in einem ersten Schritt ausgehend von dem persönlichen Forschungsvorhaben für jeden konzeptionellen Aspekt die Kriterien zu definieren, welche die den Verteilungsdaten zugrunde liegenden Einkommensdaten erfüllen sollen. In einem zweiten Schritt sollte dann, ein auf das Forschungsvorhaben zugeschnittener Qualitätsindikator erstellt werden, der transparent vermittelt, ob ein Gini-Koeffizient die zuvor definierten Kriterien erfüllt. In Abschnitt 6.1 wird zur Veranschaulichung eine simple Einstufung der Qualität einzelner Gini-Koeffizienten für jeden konzeptionellen Aspekt (Empfängereinheit, Analyseeinheit, Bedarfsgewichtung, Repräsentativität der Bevölkerungsgruppen) vorgenommen.

Nachdem für jeden konzeptionellen Aspekt ein Qualitätsindikator erstellt wurde, werden in einem dritten Schritt die Verteilungsdaten entsprechend des Qualitätsindikators selektiert. Diese Prozedur trägt dazu bei, die Validität der verwendeten Daten im Hinblick auf die zu bearbeitende Forschungsfrage sicher zu stellen und die konzeptbedingte Variation in den Daten zu minimieren. In Extremfällen, in denen die Selektion der für ein bestimmtes Forschungsvorhaben geeigneten Daten zu einer geringen Datenzahl führt, kann ein Wissenschaftler auf der Basis seines Qualitätsindikators im Nachhinein einen Kompromiss zwischen der relativen Qualität und der Quantität der verwendeten Verteilungsdaten nach eigenem, aber nachvollziehbarem, Ermessen ausloten. In diesem Fall stehen ihm dann für eine Analyse mehrere Daten zur Verfügung, jedoch müssen jegliche Ergebnisse noch kritischer berichtet werden. In Abschnitt 6.2 erfolgt die Selektion der Verteilungsdaten auf Basis des im Abschnitt 6.1 erstellten Qualitätsindikators.

6.1 Kriterien definieren und Qualitätsindikator anlegen

Als Forschungsvorhaben sei für die Demonstration eine international vergleichende Analyse der Verteilung der individuellen Wohlfahrt unterstellt. Für ein solches Analyseziel sollte das verfügbare Einkommen als Einkommenskonzept gewählt werden, da es neben den

staatlichen Abzügen – im Idealfall alle – Transfers mitberücksichtigt, die eine Empfängereinheit erhält. Infolgedessen ist für das Forschungsvorhaben ausschließlich die Variable *gini_dis* von Interesse. Verteilungsdaten, welche auf anderen Einkommenskonzepten beruhen, werden ignoriert.

Für jeden konzeptionellen Aspekt wird zunächst das zu erfüllende Kriterium definiert. Im Anschluss daran werden den Ausprägungen jeder Konzeptionsvariablen in einer weiteren Qualitätsvariablen Werte von '0' bis '2' zugewiesen. Der Qualitätsindikator trägt die Bezeichnung q_x , wobei x für den jeweiligen konzeptionellen Aspekt steht. Eine '0' bedeutet, dass der betreffende Ungleichheitswert das zuvor definierte konzeptionelle Kriterium nicht erfüllt und somit für das Forschungsvorhaben inakzeptabel, bzw. ungeeignet ist. Eine '1' bedeutet, dass das zuvor definierte Kriterium nicht im vollen Maße erfüllt wird und der Ungleichheitswert für das Forschungsvorhaben nur bedingt brauchbar ist. Diese Verteilungsdaten können im Zweifelsfall verwendet werden, wenn ansonsten nicht genügend Daten für eine Analyse zur Verfügung stehen. Hieraus geht auch hervor, dass der Wert '1' nur dann vergeben werden kann, wenn es inhaltlich angebracht erscheint. Eine '2' bedeutet, dass das Kriterium in vollem Maße erfüllt wird, der Ungleichheitswert im Hinblick auf das Forschungsvorhaben folgendermaßen voll geeignet ist.

Tabelle 5: Die Qualitätsindikatoren im Überblick

Konzeptioneller Aspekt	Variable	Qualitätsindikator	Merkmalswerte
Empfängereinheit	<i>incsharu</i>	$q_{incsharu}$	0 „Nicht geeignet“ 1 „Im Bedarfsfall geeignet“ 2 „Geeignet“
Analyseeinheit	<i>uofanala</i>	$q_{uofanala}$	0 „Nicht geeignet“ 1 „Wurde nicht vergeben“ 2 „Geeignet“

Fortsetzung Tabelle 5: Die Qualitätsindikatoren im Überblick

Bedarfsgewichtung	<i>equisc</i>	q_{equisc}	0 „Nicht geeignet“ 1 „Im Bedarfsfall geeignet“ 2 „Geeignet“
Bevölkerungsrepräsentativität	<i>popcovr</i>	$q_{popcovr}$	0 „Nicht geeignet“ 1 „Wurde nicht vergeben“ 2 „Geeignet“

6.1.1 Empfängereinheit

Zunächst stellt sich die Frage nach der optimalen Empfängereinheit (Schritt 1). Hier wird generell zwischen Personen, Familien und Haushalten unterschieden (vgl. Abschnitt 3.1.2). Da Individuen unter anderem in häuslichen Gemeinschaften leben, in denen das zur Verfügung stehende Geld geteilt wird, hängt die persönliche Wohlfahrt nicht nur vom persönlichen Einkommen ab. Damit scheint die Empfängereinheit Person im Hinblick auf das oben genannte Forschungsvorhaben ungeeignet zu sein. Vielmehr sollte in der Einkommenserhebung entweder das innerhalb von Haushalten oder von Familien gemeinsam erzielte Einkommen erhoben worden sein. Welche der beiden Alternativen die bessere Wahl ist, muss von jedem Wissenschaftler individuell beantwortet werden. Die Empfängereinheit Haushalt erlaubt es einerseits auch solche Personen zu berücksichtigen, die gemäß der spezifizierten Familiendefinition nicht in einem Verwandtschaftsverhältnis zu einer weiteren Person im Haushalt stehen, tatsächlich aber ihr Einkommen mit anderen teilen oder vom Einkommen anderer profitieren. Andererseits erscheint die Annahme, dass jede in einem Haushalt lebende Person ihr Einkommen mit anderen im Haushalt lebenden Personen teilt oder von deren Einkommen profitiert, als sehr restriktiv. Diese Annahme scheint eher auf eine Familie zuzutreffen. Je nachdem wie viel Gewicht ein Forscher diesen beiden Argumenten beimisst, verändert dies die Rangordnung hinsichtlich der Eignung eines Verteilungswertes für das Forschungsvorhaben. An dieser Stelle sei, ohne weitere Ausführungen zu machen, der Haushalt die bevorzugte Empfängereinheit. Die Variable $q_{incsharu}$ spiegle die Eignung eines Ungleichheitswertes hinsichtlich des konzeptionellen Aspektes der Empfängereinheit wider. Die Variable $q_{incsharu}$ führt dann den Wert '0' für alle Gini-Koeffizienten, die die Person als Empfängereinheit haben, den Wert '1' für alle Gini-Koeffizienten, die die Familie als Empfängereinheit haben und den Wert '2' für alle Gini-Koeffizienten, die den Haushalt als Empfängereinheit haben (Schritt 2).

6.1.2 Analyseeinheit

Gemäß des oben genannten Forschungsvorhabens ist die Verteilung der individuellen Wohlfahrt von Interesse. Daher sollte unter dem konzeptionellen Aspekt der Analyseeinheit, welcher ebenfalls zwischen Personen, Familien und Haushalten unterscheidet, das Haushaltseinkommen auf die Zahl der Mitglieder der Haushalte herunter gerechnet worden sein (vgl. Abschnitt 3.1.2). Die beiden weiteren Einheiten sind inakzeptabel, da sie die Verteilung der Wohlfahrt von Familien oder Haushalten messen (Schritt 1).

Infolgedessen führt die Variable $q_{uofanala}$, welche die Eignung eines Ungleichheitswertes hinsichtlich des konzeptionellen Aspektes der Analyseeinheit widerspiegelt, den Wert '0' für alle Gini-Koeffizienten, die die Familie oder den Haushalt als Empfängereinheit haben und

den Wert '2' für alle Gini-Koeffizienten, die die Person als Empfängereinheit haben. Der Wert '1' wird hier nicht vergeben (Schritt 2).

6.1.3 Bedarfsgewichtung

Beim gemeinsamen Haushalten liegen in Abhängigkeit der Haushaltsgröße in der Regel Skaleneffekte vor, da die Kosten für notwendige Einmalanschaffungen von mehreren Personen getragen werden. Neben der Haushaltsgröße, entscheidet die demographische Zusammensetzung des Haushalts über die finanziellen Möglichkeiten des einzelnen Haushaltsmitglieds (vgl. Abschnitt 3.1.3). Vor diesem Hintergrund sollte das Haushaltseinkommen unter der Verwendung einer Äquivalenzskala bedarfsgewichtet worden sein. Da für die Bedarfsgewichtung in der gegenwärtigen Literatur zum überwiegenden Teil auf die modifizierte OECD-Skala zurückgegriffen wird, sollte diese bei der Bedarfsgewichtung angewendet worden sein. Alle weiteren Äquivalenzskalen, hierunter fallen auch solche, in denen die Haushaltsgröße ohne Skaleneffekte berücksichtigt wird, werden im Hinblick auf das Forschungsvorhaben als Alternative zweiter Wahl betrachtet. Eine fehlende Bedarfsgewichtung ist für das Forschungsvorhaben inakzeptabel (Schritt 1).

Hieraus resultiert, dass die Variable *q_equisc*, welche die Eignung eines Ungleichheitswertes hinsichtlich des konzeptionellen Aspektes der Bedarfsgewichtung widerspiegeln, den Wert '0' für alle Gini-Koeffizienten trägt, für deren Berechnung keine Bedarfsgewichtung vorgenommen wurde. Sie trägt den Wert '1' für alle Gini-Koeffizienten, für deren Berechnung eine Bedarfsgewichtung mit einer anderen als der modifizierten OECD-Skala durchgeführt wurde. Den Wert '2' trägt sie für alle Gini-Koeffizienten, für deren Berechnung eine Bedarfsgewichtung mit der modifizierten OECD-Skala durchgeführt wurde (Schritt 2).

6.1.4 Repräsentativität der Bevölkerungsgruppen

Für das oben genannte Forschungsvorhaben sollten die den Verteilungsdaten zugrunde liegenden Einkommenserhebungen bevölkerungsweit repräsentativ sein. Basieren die Einkommenserhebungen beispielsweise ausschließlich auf Befragungen der arbeitenden Bevölkerung, sind die hieraus berechneten Verteilungsdaten nicht geeignet. Im Abschnitt 3.2 wurde erläutert, dass man Verteilungsdaten, die ein unterschiedliches Maß an Repräsentativität aufweisen, nicht gemeinsam verwenden sollte (Schritt 1).

Infolgedessen führt die Variable *q_popcovr*, welche die Eignung eines Ungleichheitswertes hinsichtlich seiner Repräsentativität im Hinblick auf die Bevölkerungsgruppen widerspiegeln, den Wert '0' für alle Gini-Koeffizienten, die nicht bevölkerungsweit repräsentativ sind und

den Wert '2' für alle Gini-Koeffizienten, die bevölkerungsweit repräsentativ sind. Der Wert '1' wird hier nicht vergeben (Schritt 2).

6.2 Selektion der Verteilungsdaten

Nachdem nun für jeden konzeptionellen Aspekt ein Qualitätsindikator erstellt wurde, erfolgt die Selektion der Daten (3. Schritt). Zur Erinnerung: die Datenbank DMI besteht aus 2335 Verteilungsdaten, wobei nicht ausgeschlossen ist, dass für ein Land zu einem Jahr mehrere Verteilungsdaten mit denselben konzeptionellen Aspekten vorliegen. Um das Ausmaß an konzeptionsbedingter Variation in den Daten zu minimieren und damit das Risiko empirischer Fehlschlüsse zu minimieren, werden für die Bearbeitung des Forschungsvorhabens nur jene Verteilungsdaten herangezogen werden, die in allen angelegten Qualitätsindikatoren den Wert '2' führen. Dieses Vorgehen führt zu 217 geeigneter Daten, die die Verteilungssituation für über 29 Länder in einem Zeitraum von 1966 bis 2006 in unregelmäßiger Weise beschreiben. Anhand des simplen Beispiels zeigt sich die Problematik der Datenverfügbarkeit. Im äußersten Fall sind dem Forscher Grenzen gesetzt.

Wenn die Datengrundlage für einen Wissenschaftler aus gegebenen Gründen unzureichend sein sollte, kann ein Forscher auf der Grundlage des erstellten Qualitätsindikators überprüfen, bei welchen konzeptionellen Aspekten er zu Kompromissen bereit ist. Geht man für alle konzeptionellen Aspekte einen Kompromiss ein, verwendet man also auch jene Verteilungsdaten, welche in allen angelegten Qualitätsindikatoren den Wert '1' führen, so erhält man 764 Verteilungsdaten für die Analyse. In diesem Fall aber sollte man seinen Ergebnissen weitaus kritischer gegenüber stehen. Die Daten sind bezüglich ihrer konzeptionellen Aspekte heterogener, weshalb Beeinträchtigungen der empirischen Ergebnisse nicht auszuschließen sind. Auf diesen Aspekt sollte stets hingewiesen werden.

Tabelle 6: Für die Analyse verbleibende Verteilungsdaten nach einer Selektion nach strengen Kriterien und nach einer Selektion mit Kompromissen

	Länder	Zahl der Gini-Koeffizienten	Zeitraum	Datensatz
Selektion nach strengen Kriterien	29	217	1966 bis 2006	Unbalanciert
Selektion mit Kompromissen	97	764	1961 bis 2006	Unbalanciert

7) Zusammenfassung

Ziel des Papiers war es, Forschern, welche im Umgang mit Ungleichheitsdaten unerfahren sind, ein Verständnis für die konzeptionelle Vielfalt der ihnen zugrunde liegenden Einkommensdaten zu geben sowie den Umgang mit der *Datenbank Makroindikatoren* zu erläutern. Auf die Probleme, die mit Einkommenserhebungen, aus denen diese Ungleichheitsdaten errechnet werden, verbunden sind, wurde nicht eingegangen. Es wurde ausschließlich die Berechnung und Interpretation des Gini-Koeffizienten erläutert, weil der größte Teil der weltweit verfügbaren Ungleichheitsdaten mit dieser Maßzahl berechnet wurde. Andere Maßzahlen, wie etwa das Atkinson-Maß oder der Theil-Index, blieben unberücksichtigt.

Verteilungsdaten werden auf der Grundlage von Einkommensdaten errechnet, welchen in der Regel ein einjähriger Betrachtungszeitraum zugrunde liegt. Somit gehören Verteilungsdaten zu der Gruppe der aggregierten Makroindikatoren. Die den Verteilungsdaten zugrunde liegenden Einkommensdaten können sich in konzeptionellen und repräsentativen Aspekten erheblich unterscheiden. Aus diesem Kontext heraus lässt sich die erste Implikation ableiten. Es sollten stets Verteilungsdaten verwendet werden, deren zugrunde liegende Einkommensdaten in konzeptioneller Hinsicht im Einklang mit der zu bearbeitenden Forschungsfrage stehen. Wissenschaftliche Abhandlungen, die sich mit der Leistungsgerechtigkeit beschäftigen, erfordern eine andere konzeptionelle Ausgestaltung der den Verteilungsdaten zu Grunde liegenden Einkommensdaten als Forschungsarbeiten, die sich mit der wirtschaftlichen Wohlfahrt oder aber der sozialen Schichtung auseinandersetzen.

Da sich jeder konzeptionelle Aspekt der Einkommensdaten in unterschiedlich starkem Maß auf die Verteilungsdaten auswirkt, ergibt sich für den Umgang mit Verteilungsdaten eine zweite Implikation, die von Wissenschaftlern idealerweise befolgt werden sollte. Für wissenschaftliche Analysen sollten ausschließlich Ungleichheitsdaten verwendet werden, die auf identisch konzipierten Einkommensdaten beruhen. Werden Verteilungsdaten verwendet, welche auf unterschiedlich konzipierten Einkommensdaten beruhen, birgt dies die Gefahr, Fehlschlüssen in der empirischen Analyse zu erliegen, die einzig auf konzeptbedingte Variationen in den Daten zurückzuführen sind. Ergebnisse, welche auf diesem Wege ermittelt werden, bieten eine große Angriffsfläche für Kritik an den verwendeten Daten und entziehen einem Wissenschaftler jede Möglichkeit, der geübten Kritik mit substanziellen Argumenten entgegenzutreten.

Die Befolgung beider Handlungsempfehlungen lässt sich in der Praxis allerdings nicht ohne Abstriche umsetzen. Da die bis dato weltweit verfügbaren Ungleichheitsdaten in Bezug auf die ihnen zu Grunde liegenden Einkommensdaten sehr heterogen sind, kann die Befolgung der Handlungsempfehlungen eine stark verringerte Zahl an verfügbaren

Ungleichheitswerten bedingen. In Abschnitt 6 wurde anhand eines simplen Beispiels demonstriert, dass für ein definiertes Forschungsvorhaben infolge strenger Selektionskriterien lediglich 274 von ursprünglich 2335 Verteilungsdaten verwendet werden konnten. Somit besteht zwischen der konzeptionellen Homogenität der Verteilungsdaten bzw. der Vermeidung von konzeptbedingter Variation in den Daten und der Generalisierbarkeit der Ergebnisse in der Quer- und Längsschnittdimension ein Zielkonflikt der folgenden Art: zu hohe Anforderungen bezüglich der Vergleichbarkeit der Gini-Koeffizienten führen dazu, dass in der empirischen Analyse nur wenige Länder berücksichtigt werden können, sodass letztlich Aussagen nur über wenige Länder und/oder Zeitpunkte gemacht werden können.

Die Ergebnisse können erst dann widerlegt werden, wenn in Zukunft mit Hilfe neuer Einkommensdaten neue Verteilungsdaten berechnet werden, welche in künftigen Forschungsarbeiten zu abweichenden Ergebnissen führen. Vor diesem Hintergrund erscheint es für wissenschaftliche Abhandlungen, in denen Verteilungsdaten verwendet werden, angebracht, stets beide Handlungsempfehlungen zu befolgen. Wird mindestens eine der Handlungsempfehlungen ignoriert, so muss dies in der Forschungsarbeit dokumentiert und Aussagen über die anzunehmenden Auswirkungen auf die Ergebnisse getroffen werden.

Infolge des von der Canberra Group erstellten Rahmenwerks für Einkommensstatistiken ist für die Zukunft zu erwarten, dass Verteilungsdaten hinsichtlich der konzeptionellen Aspekte der ihnen zugrunde liegenden Einkommensdaten homogener werden. Vor diesem Hintergrund ist zu empfehlen, dass die Ergebnisse vergangener Forschungsarbeiten in Zukunft repliziert werden.

Anhang

1. Urliste für Abbildung 3a (n=5)

j	x_{<j>}	p_j (= x _{<1>} +...+ x _{<n>})	q_j (= j*x _{<j>})	u_j (= j/n)	v_j (= p _j /p _n)
0	0	0	0	0	0
1	20	20	20	0.2	0.134
2	20	40	40	0.4	0.268
3	20	60	60	0.6	0.403
4	20	80	80	0.8	0.537
5	69	345	149	1	1.000
	149	545			

Gini- Koeffizient: 0,263; Normierter Gini-Koeffizient: 0,329

2. Urliste für Abbildung 3b (n=5)

j	x_{<j>}	p_j (= x _{<1>} +...+ x _{<n>})	q_j (= j*x _{<j>})	u_j (= j/n)	v_j (= p _j /p _n)
0	0	0	0	0	0
1	19	19	19	0.2	0.068
2	19	38	38	0.4	0.137
3	80	240	118	0.6	0.424
4	80	320	198	0.8	0.712
5	80	400	278	1	1.000
	278	1017			

Gini- Koeffizient: 0,263; Normierter Gini-Koeffizient: 0,329

LITERATURVERZEICHNIS

- Atkinson, Anthony B., Lee Rainwater und Timothy Smeeding, 1995: Income Distribution in OECD Countries. Evidence from the Luxembourg income study. Paris: OECD.
- Atkinson, Anthony B. und Francois Bourguignon, 2000: Income Distribution and Economics. S. 1-58 in: Anthony B. Atkinson und Francois Bourguignon [Hg.]: Handbook of Income Distribution, Amsterdam u.a.: Elsevier.
- Atkinson, A. B. and Brandolini, A. 2001: Promise and Pitfalls in the Use of "Secondary" Data-Sets: Income Inequality in OECD Countries as a Case Study, Journal of Economic Literature, 39: 771-799.
- Becker, Irene und Richard Hauser, 2004: Lebenslagen in Deutschland. Verteilung der Einkommen 1999 – 2003. Bericht zur Studie im Auftrag des Bundesministeriums für Gesundheit und Soziale Sicherung, Online-Dokument: http://www.bmas.de/coremedia/generator/1330/property=pdf/verteilung_der_einkommen_1999_2003.pdf, Stand 01.09.2008.
- Blümle, Gerold, 1975: Theorie der Einkommensverteilung. Berlin u.a.: Springer-Verlag.
- Bomsdorf, Eckhart, 2002: Deskriptive Statistik, Lohmar u.a.: Josef Eul Verlag.
- Buhmann, Brigitte, Lee Rainwater, Guenther Schmaus und Timothy M. Smeeding, 1988: Equivalence Scales, Well-Being, Inequality, And Poverty: Sensivity Estimates Across Ten Countries Using The Luxembourg Income Study (LIS) Database, The review of income and wealth, 34: 115-142.
- Canberra Group, 2001: Expert Group on Household Income Statistics. Final Report and Recommendations, Online-Dokument: <http://www.lisproject.org/links/canberra/finalreport.pdf>, Stand: 26.08.2008.
- Deckl, Silvia, 2006: Indikatoren der Einkommensverteilung in Deutschland 2003. Ergebnisse der Einkommens- Verbraucherstichprobe, Fachserie 15 EVS 2003, Heft 6: 19-27.
- Dennis, Ian und Anna-Catherine Guio, 2004: Monetäre Armut in den neuen Mitgliedstaaten und den Bewerberländern, Statistik kurz gefasst - Bevölkerung und soziale Bedingungen, 12/2004: 1-12. Online-Dokument: http://www.eddestatis.de/de/downloads/sif/nk_04_12.pdf, Stand: 01.09.2008.
- Friedrichs, Jürgen, 1990: Methoden empirischer Sozialforschung. Opladen: Westdeutscher Verlag.
- Gardiner, Karen, John Hills, Jane Falkingham, Holly Sutherland und Valérie Lechene, 1995: The effects of differences in housing and health care systems on international

comparisons of income distribution, Welfare State Programme Discussion Paper No.110, STICERD (London School of Economics), July.

Gobbin, Niko und Glenn Rayp, 2004: Income Inequality Data in Growth Empirics: From Cross-Section to Time Series. Universiteit Gent, Workingpaper 2004/252, Online-Dokument: [http://www.feb.ugent.be/fac/research/WP/Papers/wp_04_252 .pdf](http://www.feb.ugent.be/fac/research/WP/Papers/wp_04_252.pdf), Stand: 05.11.2007.

Gottschalk, Peter and Minh Huynh, 2006: Are Earnings Inequality and Mobility Overstated? The Impact of Non-Classical Measurement Error, Working Paper 649. Online-Dokument: <http://fmwww.bc.edu/ec/gottschalk.php>, Stand: 05.09.2008.

Hauser, Richard, 2002: Zum Einfluss von Äquivalenzskalen auf Ergebnisse zur personellen Einkommensverteilung und zur relativen Einkommensarmut. S. 175-190 in: Lukas Menkhoff und Friedrich L. Sell [Hg.], Zur Theorie, Empirie und Politik der Einkommensverteilung, Berlin u.a.: Springer-Verlag.

Krause, Peter und Daniel Ritz, 2006: EU-Indikatoren zur sozialen Inklusion in Deutschland, Vierteljahrshefte zur Wirtschaftsforschung, 75: 152-173. Online-Dokument: http://www.diw.de/documents/publikationen/73/44351/v_06_1_10.pdf, Stand: 05.09.2008.

Scheuch, Erwin K., 1966: Cross-national comparisons using aggregate data: Some substantive and methodological problems. S. 131-167 in: Richard L. Merritt and Stein Rokkan [Hg.], Comparing nations: The use of quantitative data in cross-national research, New Haven, Conn.: Yale University Press.

Schmid, Friedrich und Karl Mosler, 2004: Wahrscheinlichkeitsrechnung und schließende Statistik. Berlin u.a.: Springer-Verlag.

Smeeding, Timothy M., P. Saunders, J. Coder, S. Jenkins, J. Fritzell, A. Hagenaars, R. Hauser und M. Wolfson, 1993: Poverty, inequality, and family living standard impacts across seven nations: The effect of noncash subsidies, Review of Income and Wealth 39 (3): 229-256.

World Institute for Development Economics Research WIID User Guide and Data Sources, 2008. Online-Dokument: <http://www.wider.unu.edu/wiid/wiid.htm>, Stand: 19.08.2008.