

Rendtel, Ulrich; Pannenberg, Markus; Daschke, Stefan

**Article — Digitized Version**

## Die Gewichtung der Zuwanderer-Stichprobe des Sozio-oekonomischen Panels (SOEP)

Vierteljahrshefte zur Wirtschaftsforschung

**Provided in Cooperation with:**

German Institute for Economic Research (DIW Berlin)

*Suggested Citation:* Rendtel, Ulrich; Pannenberg, Markus; Daschke, Stefan (1997) : Die Gewichtung der Zuwanderer-Stichprobe des Sozio-oekonomischen Panels (SOEP), Vierteljahrshefte zur Wirtschaftsforschung, ISSN 0340-1707, Duncker & Humblot, Berlin, Vol. 66, Iss. 2, pp. 271-286

This Version is available at:

<https://hdl.handle.net/10419/141184>

**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.*

# Die Gewichtung der Zuwanderer-Stichprobe des Sozio-oekonomischen Panels (SOEP)\*

Von Ulrich Rendtel\*\*, Markus Pannenberg\*\*\* und Stefan Daschke\*\*\*\*

## Zusammenfassung

*Generell sehen sich Wiederholungsbefragungen von Haushalten oder Unternehmen (Panel-Design) mit dem Problem konfrontiert, daß sie exogen verursachte Bestandsdynamiken wie z.B. Zuwanderung von Haushalten oder Neugründung von Unternehmen nicht ohne weiteres nachbilden können. Diese unzureichende Erfassung exogener Bestandsdynamiken ist aus inhaltlichen Überlegungen heraus unbefriedigend. Ein Ausweg ist die Ergänzung der laufenden Panelbefragung um eine Ergänzungsstichprobe. Der folgende Artikel beschreibt das Design einer derartigen Stichprobe, die Berechnung von (isolierten) Hochrechnungsfaktoren und die Integration der berechneten Gewichte in ein bestehendes Hochrechnungskonzept einer laufenden Panelbefragung am Beispiel des Sozio-ökonomischen Panels (SOEP).*

## 1. Einleitung

Generell sehen sich Wiederholungsbefragungen von Haushalten oder Unternehmen (Panel-Design) mit dem Problem konfrontiert, daß sie exogen verursachte Bestandsdynamiken wie z.B. Zuwanderung von Haushalten oder Neugründung von Unternehmen nicht ohne weiteres nachbilden können. Das Beispiel der Migration von Haushalten verdeutlicht dies: Die Zuwanderung in das Gebiet, dessen Wohnbevölkerung durch eine Panelbefragung repräsentiert wird, wird nur erfaßt, wenn diese Haushalte unmittelbar oder aber im Laufe der Zeit nach der Zuwanderung in bestehende Befragungseinheiten einziehen. Ansonsten wird diese (neue) Bevölkerungsgruppe durch eine laufende Panelbefragung nicht erfaßt.

Wenngleich diese Eigenschaft laufender Panelbefragungen stichprobentheoretisch keine Probleme aufwirft, solange man ein Panel als Kohorten-Studie interpretiert, ist die unzureichende Erfassung exogener Bestandsdynamiken jedoch aus inhaltlichen Überlegungen heraus unbefriedigend. Zentrales Anliegen von Haushaltsbefragungen ist die Modellierung von sozialem Wandel im jeweiligen Erhebungsgebiet. Dies schließt selbstverständlich Zuwanderung von Haushalten ein. Ein naheliegender Ausweg aus dem skizzierten Dilemma ist die Ergänzung der laufenden Panelbefragung um eine Ergänzungsstichprobe (in unserem Beispiel eine Zuwandererstichprobe). Der folgende Artikel beschreibt das Design einer derartigen Stichprobe, die Berechnung von (isolierten) Hochrechnungsfaktoren und die Integration der berechneten

Gewichte in ein bestehendes Hochrechnungskonzept einer laufenden Panelbefragung am Beispiel des Sozio-ökonomischen Panels<sup>1</sup>.

## 2. Die Erfassung von Zuwanderern im Sozio-oekonomischen Panel (SOEP)

Das SOEP ist eine Erhebung, die Auskunft über das Leben und den gesellschaftlichen Wandel in Deutschland gibt. Die Grundgesamtheit des SOEP umfaßt die in Privathaushalten lebende Wohnbevölkerung. Die Besonderheit einer Panel-Befragung besteht darin, daß Personen, die einmal in die Stichprobe gekommen sind, immer wieder

\* Wir danken Gert Wagner und Bernhard v. Rosenblatt für kritische Anmerkungen und konzeptionelle Ideen.

\*\* Johann Wolfgang Goethe-Universität, Frankfurt am Main.

\*\*\* Projektgruppe SOEP im Deutschen Institut für Wirtschaftsforschung, Berlin.

\*\*\*\* Student der Volkswirtschaftslehre, Universität Bielefeld; Praktikant bei der Projektgruppe des Sozio-oekonomischen Panels im DIW.

<sup>1</sup> Der Artikel setzt damit die Beschreibung der von der Projektgruppe Sozio-ökonomisches Panel am DIW bereitgestellten Längsschnittdaten und ihrem Nutzen für wissenschaftliche Analysen in den Vierteljahrsheften für Wirtschaftsforschung fort. Vgl. z.B. Projektgruppe Sozio-ökonomisches Panel (1995) oder Rendtel (1987). Einen Überblick über ausgewählte Analysen geben Burkhauser/Wagner (1994) oder Dunn/Schwarze (1997).

befragt werden. Bei der jährlich durchgeführten Wiederholungsbefragung des SOEP werden alle im jeweiligen Panel-Haushalt lebenden Personen registriert. Mit allen Personen im Alter von mindestens 16 Jahren werden persönliche Interviews geführt (vgl. dazu z.B. Projektgruppe „Das Sozio-oekonomische Panel“, 1995).

Seit der Startwelle des SOEP im Jahr 1984 haben sich die Panel-Haushalte wie ein lebender Organismus gewandelt. Kinder sind mit Vollendung des sechzehnten Lebensjahres in den Befragungsmodus gelangt. Geburten und Todesfälle haben für eine natürliche Veränderung der Haushaltszusammensetzungen gesorgt. Des weiteren haben Personen ihren Stammbau verlassen und einen eigenen Haushalt gegründet. Andere Personen sind vorübergehend oder dauerhaft ins Ausland gezogen. Für alle Personen, die jemals ein Interview gegeben haben, werden im Falle eines Umzugs im Inland die neuen Adressen ermittelt. Sie bleiben im SOEP und werden weiter befragt (zu den Weiterverfolgungsregeln des SOEP vgl. Wagner et al., 1994).

Ein konzeptionelles Problem bereiteten aber die Zuwanderer, die seit Bestehen des SOEP nach Deutschland gekommen sind (vgl. Schulz et al., 1993). Das SOEP-Konzept erfaßte nur diejenigen Zuwanderer, die direkt in bereits bestehende Panel-Haushalte aufgenommen wurden oder später dorthin umzogen. Folglich hatten Zuwanderer, die sofort oder nach einem Aufenthalt in einem Übergangsheim einen eigenen Haushalt gründeten, keine Möglichkeit, in eine der Teilstichproben des SOEP zu gelangen. Gleiches galt für Zuwanderer, die nur für eine kurze Zeit zwischen den Befragungen in einem Panel-Haushalt lebten und dann einen eigenen Haushalt gründeten.

Bedingt durch historische Ereignisse und die dadurch veränderten politischen Rahmenbedingungen sind zahlreiche Personen, teilweise sogar ganze Dorfgemeinschaften, von Ost- nach Westeuropa gewandert. Die größte Population bilden Aussiedler, bei denen es sich um deutschstämmige Personen aus osteuropäischen Staaten handelt. Übersiedler, also Zuzügler aus der DDR, sind bis Juni 1990 ebenfalls nicht erfaßt worden (Erst ab Juli 1990 spiegelt die Teilstichprobe C (Ost) die innerdeutschen Wanderungsbewegungen wider). Weiterhin hat es noch zahlreiche Zuwanderungen aus anderen Staaten innerhalb und außerhalb Europas gegeben, darunter Deutsche, die aus dem Ausland zurückkehrten, sowie Flüchtlinge, Asylbewerber und übrige Ausländer.

Versucht man, den Umfang an Zuwanderung nach Deutschland zu quantifizieren, so fällt auf, daß zuverlässige Zahlen der amtlichen Statistik für die in Privathaushalten lebende Zuwandererpopulation nicht vorliegen. Insbesondere existiert für diese verstreut lebende Spezialpopulation kein Zentralregister, aus dem Adressen gezogen werden können. Zwar kann man auf das Ausländerzentralregister, das vom Statistischen Bundesamt ausgewertet wird, und auf die Wanderungsstatistik, deren Erhebung auf

Daten der Einwohnermeldeämter basiert, zurückgreifen. Beide erfassen aber die Zuwanderungen nur unvollständig, da z.B. viele der Zuwanderer keine Ausländer, sondern Deutsche im Sinne des Grundgesetzes sind. Noch schwieriger ist die Erfassung der Fortzüge dieses Personenkreises. Des weiteren verweigerten beim Mikrozensus von 1992 mehr als ein Drittel der Befragten die (freiwillige) Antwort auf die Frage nach ihrer Herkunft, so daß auch diese Erhebung keine zuverlässigen Informationen über den Bestand der Zuwandererpopulation in Privathaushalten liefert (vgl. Schupp/Wagner 1995). Schätzungen des DIW ergaben bis Ende 1993 einen Bestand an Personen in Privathaushalten, die im Zeitraum 1984 bis 1993 zugewandert sind, von weit mehr als 5 Millionen. Dies bedeutet, daß im Jahre 1994/1995 in rund 7 vH der westdeutschen Privathaushalte Zuwanderer lebten (vgl. Schulz 1994).

Vor dem Hintergrund dieser (geschätzten) quantitativen Bedeutung von Zuwanderer-Haushalten seit Mitte der achtziger Jahre und ihrer unzureichenden Erfassung in den Teilstichproben A, B und C des SOEP, wurde von der Projektgruppe „Das Sozio-oekonomische Panel“ beschlossen, eine Zuwanderer-Stichprobe für das SOEP — die Teilstichprobe D — zu konzipieren. Diese sollte, wie auch die Teilstichproben B und C, überproportional gezogen werden, um eigenständige Analysen der Zuwandererpopulation zu ermöglichen. Das Erhebungsgebiet beschränkte sich auf die westdeutschen Bundesländer und auf das ehemalige West-Berlin, da in den neuen Bundesländern zum Erhebungszeitpunkt kaum Zuwanderer lebten. Erhebungseinheiten sind weiterhin die Privathaushalte, wodurch Bewohner von Übergangsheimen unberücksichtigt bleiben.

Im folgenden wird in einem ersten Schritt kurz die Zielung der Teilstichprobe D dargestellt. Im Hauptabschnitt wird die Gewichtung der Hochrechnung dieser Teilstichprobe konzeptionell vorgestellt. Daran anschließend werden die Hochrechnungsergebnisse mit der DIW-Schätzung und mit entsprechenden Schätzungen aus der ALLBUS-Erhebung 1994 verglichen. Im Anhang wird die praktische Nutzung der Hochrechnungs- (bzw. Gewichtungs-) Faktoren erläutert.

### 3. Die Erhebung der Teilstichprobe D des SOEP

Die Grundgesamtheit der Teilstichprobe D bestand aus allen Haushalten, in denen Personen leben, die seit 1984 nach Westdeutschland gezogen sind. Hiervon sind Haushalte ausgenommen, die seit Juni 1990 von Ost- nach Westdeutschland zogen, da dieser Teil der Zuwanderungen bereits über die Teilstichprobe C erfaßt wird. Die Zielung der Zielhaushalte von verstreut lebenden Teilpopulationen stellt ein erhebungstechnisches Problem dar. Verschiedene Testerhebungen führten schließlich dazu, für die Zuwanderer-Stichprobe D die Methode der Adressenkumulation im Rahmen regulärer bevölkerungsrepräsentativer

tativer Mehrthemenbefragungen (Buseinschaltungen)<sup>2</sup> zu benutzen. Testweise wurde auch eine einstufige „Schneeball-Ergänzung“ verwendet<sup>3</sup>, um die Zahl der Befragungshaushalte bei geringen Feldkosten aufzustocken. Die Erhebung wurde von Infratest Sozialforschung durchgeführt, das seit Beginn des SOEP mit der Durchführung der Feldarbeit betraut ist.

Die „Busumfragen“ erfolgten einerseits mittels Ziehung der Haushalte nach einem modifizierten ADM-Design<sup>4</sup> und face-to-face-Interviews und zum anderen durch Telefon-Interviews mit automatischen Wählalgorithmen (InfraScope). Allerdings dürfte die Zahl der Zuwanderer bei allgemeinen Bevölkerungsumfragen eher unterschätzt werden, insbesondere die Zahl der ausländischen Haushalte. Diese potentielle Untererfassung erklärt sich in erster Linie durch das ADM-Stichprobendesign. Das ADM-Ziehungsband basiert auf den Stimmbezirken der jeweiligen Bundestagswahl. Bei dem zweistufigen Auswahlverfahren werden in der ersten Stufe in systematischer Zufallsauswahl mit größenproportionalen Auswahlwahrscheinlichkeiten und „Schichtung durch Anordnung“ einzelne Stimmbezirke gezogen. Das ADM-Design schätzt dabei die Anzahl der Privathaushalte der Primäreinheit (Stimmbezirk) proportional zur Anzahl der Wahlberechtigten im Stimmbezirk. Folglich werden Wohngebieten mit einer starken Häufung von Zuwandererhaushalten beim ADM-System unterrepräsentiert. In der zweiten Stufe werden über ein „Random-Route“-Begehungsverfahren die Haushalte ausgewählt. Hier werden bei den üblichen Bevölkerungsumfragen nur Personen mit deutscher Staatsangehörigkeit ausgewählt. Die Interviewer sind also darauf geschult, Ausländer aus der Befragung auszuschließen. Wenn sie dies wie im Fall der D-Stichprobe ausnahmsweise anders handhaben sollen, so ist das im Rahmen der standardisierten Abläufe bei den Busumfragen ein nicht unproblematischer Eingriff in die Verfahrensregeln. Nach Einschätzung des Befragungsinstituts ist diese Umstellung erst beim zweiten Durchlauf für die Zuwanderer-Stichprobe des SOEP, d.h. bei der Teilstichprobe D2, befriedigend gelungen.

Im Jahre 1992 wurden jeweils zwei Mehrthemenbefragungen mit Infratest-Bus und InfraScope durchgeführt. Von jeweils 4000 Haushalten wurden 141 bzw. 189 Haushalte identifiziert, in denen mindestens eine seit 1984 aus dem Ausland zugewanderte Person lebte. Von diesen Haushalten gaben 111 bzw. 84 ihr Einverständnis zur Speicherung der Adresse. Die Zustimmung zur Speicherung der Adresse zum Zwecke späterer Befragung war bei den persönlich geführten Interviews mit 78,8 vH deutlich höher als bei der Telefonumfrage (44,4 vH). In 138 Haushalten wurden Interviews realisiert. Diese Haushalte wurden um die Angabe von ihnen bekannten weiteren Zuwandererhaushalten gebeten (einstufige Schneeball-Ergänzung). In 98 der 141 genannten Schneeball-Haushalte wurden Interviews realisiert<sup>5</sup> (vgl. Tabelle 1). Diese 1994 befragten Haushalte bilden zusammen die Sub-Stichprobe D1.

Um die angestrebte Größe (ca. 500 Haushalte) der Zuwanderer-Stichprobe D zu erreichen, sind im Jahre 1994 fünf weitere Mehrthemenbefragungen (Infratest-Bus) für ein Adressen-Screening genutzt worden. Dabei wurden insgesamt 725 Zuwanderer-Haushalte identifiziert. Davon gaben 548 Haushalte ihr Einverständnis zur Speicherung der Adresse (Brutto-D2). Aufgrund finanzieller Restriktionen wurde jedoch nur ein Bestand von 400 Adressen für die Befragung ausgewählt. Die Auswahl erfolgte über eine Quotierung nach Herkunftsländern<sup>6</sup>.

Es sprechen einige Anzeichen dafür, daß die Sub-Stichprobe „Brutto-D2“ der „wahren“ Verteilung der Zuwanderer-Haushalte nach Herkunftsländern besser entspricht als die der Haushalte der Sub-Stichprobe D1. So kommt eine Schätzung des DIW (vgl. Schulz 1994), die ihre Aussagen auf amtliche Daten stützt, zu ähnlichen Aussagen über den Umfang und die Verteilung nach Herkunftsländern, wie die Spalte Brutto-D2 in Tabelle 2. Außerdem ist der Anteil der identifizierten Zuwanderer-Haushalte in Brutto-D2 mit 6,6 vH des Stichprobenumfangs deutlich höher als in D1 mit 3,5 vH (Infratest-Bus) und 4,7 vH (InfraScope), vgl. Tabelle 1.

Die Gesamtstichprobe D (D1 einschließlich der Schneeball-Haushalte sowie D2) soll die Zuwandererbevolkerung in Privathaushalten repräsentieren. Die Anteile verschiedener Bevölkerungsgruppen (Übersiedler, Aussiedler, sonstige Zuwanderer) sind in den einzelnen Sub-Stichproben sehr unterschiedlich, vgl. Tabelle 2. So ist z.B. der Anteil der Aussiedler in D1 im Vergleich zu Brutto-D2 unverhältnismäßig hoch und der Anteil der sonstigen Zuwanderer zu niedrig. Die Quotierung des Brutto-D2 wurde von Infratest so gewählt, daß die um D2 vervollständigte Gesamtstichprobe die Verteilung von Brutto-D2 nach Herkunftsländern reproduziert. Das bedeutet vor allem, daß Zuwanderer, die nicht die deutsche Staatsbürgerschaft besitzen, in D2 überproportional vertreten sind. Die Einordnung der Zuwanderer in die drei Teil-Populationen (Übersiedler, Aussiedler, Sonstige) ist allerdings nicht unproblematisch. So

<sup>2</sup> Es wurde die Frage gestellt: „Gibt es unter den erwachsenen Haushaltsmitgliedern in diesem Haushalt jemanden, der vor zehn Jahren — also im Jahr 1984 — noch nicht in der Bundesrepublik gelebt hat, sondern im Ausland oder in der ehemaligen DDR?“

<sup>3</sup> Vgl. zu den verschiedenen Auswahlverfahren Schulz et al. (1993), S. 20 ff.

<sup>4</sup> Arbeitskreis Deutscher Marktforschungsinstitute e.V. (ADM), vgl. dazu Krug et al (1994), S. 248 ff.

<sup>5</sup> Bei der Gewichtung wurde die Schneeballergänzung der D1-Haushalte nicht berücksichtigt, vgl. die Ausführungen in Abschnitt 2.1.

<sup>6</sup> Ein weiteres Auswahlkriterium war die regionale Nähe zu bestehenden Infratest-Sample-Points. Die 295 Haushalte der Teilstichprobe D2 kommen aus 172 Sample-Points der zugrunde liegenden Bevölkerungsstichprobe. Im Höchstfall gibt es 4 Zuwandererhaushalte in einem Sample-Point (7 Fälle), 3 Haushalte gibt es in 10 Fällen. In der großen Mehrzahl entfallen auf einen Sample-Point also nur 1 oder 2 Zuwandererhaushalte. Die Klumpung der Stichprobe ist damit gering.

Tabelle 1

## Basisdaten für die Designgewichtung der Zuwanderer-Stichprobe D

	Teilstichprobe D1		Teilstichprobe D2
	Infratest Bus	Infratest Scope	Infratest Bus
Erhebungsjahr	1992	1992	1994
Anzahl der Privathaushalte in Westdeutschland / Zahl der gezogenen Haushalte = Auswahlatz	28 776 000 / 4 000 = 7 194	28 776 000 / 4 000 = 7 194	29 562 000 / 11 000 = 2 687
Identifizierte Zuwanderer-HH (Anteil in vH)	141 (3.5)	189 (4.7)	725 (6.6)
Einverständnis zur Speicherung der Adresse (Anteil in vH)	111 (78.7)	84 (44.4)	548 (75.6)
	Realisierungsquote 1994 in vH nach Herkunftsgebiet:		
Übersiedler	77.1		
Aussiedler	77.9		
Sonstige	41.1		
	Realisierungsquote 1995 in vH nach Herkunftsgebiet:		
Übersiedler	93.6		
Aussiedler	90.8		
Sonstige	95.4		
			Quotierung der D2-Stichprobe in vH des Ausgangsbestands nach Herkunftsgebieten:
Übersiedler			52.7
Aussiedler			69.6
Sonstige			100.0
			Realisierungsquote 1995 in vH nach Herkunftsgebiet:
Übersiedler			76.6
Aussiedler			83.3
Sonstige			65.5
<i>Quelle:</i> Infratest Methodenbericht; eigene Berechnungen.			

ist gerade die Abgrenzung zwischen „Aussiedlern“ und „Sonstigen“ eine „weiche“ Abgrenzung. Hier kann der Fall auftreten, daß Aussiedler als Nationalität die Nationalität ihres Herkunftslandes angeben, obwohl sie mit dem Zeitpunkt ihrer Einreise nach Deutschland als Deutsche (Aussiedler) gezählt werden. Diese Ungenauigkeit tritt beispielsweise bei der Wanderungsstatistik auf Basis der Angaben der Zuwanderer bei den Einwohnermeldeämtern auf (vgl. Schulz 1994). Aus diesem Grund ist mit Unge-

nauigkeiten zwischen der Gruppe der Aussiedler und der sonstigen Zuwanderer zu rechnen.

#### 4. Die Generierung der Hochrechnungsfaktoren

##### 4.1 Konzeptionelle Überlegungen

Die Hochrechnung der Teilstichprobe D geschah unter den folgenden konzeptionellen Vorgaben:

Tabelle 2

**Zusammensetzung der Teilstichproben nach Herkunftsgruppen**  
**Auswahl: Haushalte mit Interview in Welle L (1995)**  
 Prozentangaben in Klammern

	D1 Bus, Scope	Schneeball	D2	D	Brutto-D2
Übersiedler	61 (46.2)	31 (32.6)	59 (20.0)	151 (28.9)	146 (26.9)
Aussiedler	60 (45.5)	52 (54.7)	104 (35.3)	216 (41.4)	217 (40.0)
Sonstige	11 (8.3)	12 (12.6)	132 (44.8)	155 (29.7)	179 (33.1)
Summe	132 (100.0)	95 (100.0)	295 (100.0)	522 (100.0)	542 (100.0)

*Quelle:* Methodenbericht Infratest; eigene Berechnungen.

- Zum einen soll es möglich sein, die Teilstichprobe D eigenständig auszuwerten. Ebenso sollen die alten Teilstichproben A, B und C weiterhin eigenständig und ohne Strukturbruch in den Eckzahlen auswertbar bleiben. Zum zweiten sollen die Teilstichproben des SOEP (A, B, C und D) zusammen als Stichprobe *aller* Privathaushalte ausgewertet werden können.
- Für den Gesamtbestand der Zuwanderungen seit 1984 stehen aus anderen Quellen nur grobe Schätzungen zur Verfügung. Aus diesem Grunde sollte die Hochrechnung zunächst auf einer Design-Gewichtung ohne Randanpassung basieren, um nicht völlig auf den Annahmen der Schätzungen des Gesamtbestands zu basieren.

Diese Vorgaben erzwangen die folgende Vorgehensweise:

Die Teilstichprobe D enthält einige Haushalte, die sowohl über die bisherigen Teilstichproben des SOEP als auch über das Auswahlverfahren für D1 und D2 in das SOEP gelangen konnten. Hierunter fallen sogenannte „gemischte Haushalte“, in denen Zuwanderer und „Alt“-Bürger gemeinsam leben. Im bisherigen SOEP treten derartige gemischte Haushalte als Haushalte mit Zuzügen aus Ostdeutschland bzw. aus dem Ausland auf. Insgesamt lassen sich in der D-Stichprobe 70 gemischte Haushalte (13 vH) identifizieren. Der Umfang dieser Gruppe ist also nicht vernachlässigbar.

Zahlenmäßig nicht ignorierbar sind ebenfalls gewisse Überlappungen mit der Teilstichprobe C (Ostdeutsche). Die Stichprobe D enthält 43 Haushalte (= 8 vH), die erst nach Juni 1990 nach Westdeutschland gezogen sind. Diese Haushalte gehören nicht zu der Auswahlpopulation der D-Stichprobe. Die Auswahl dieser Haushalte erfolgte vorwiegend in der Stichprobe D1 aufgrund von Ungenauigkeiten bei der Stichprobenabgrenzung im Feld: Ausschlusskriterium war dort „Übersiedler aus der ehemaligen DDR, die nach 1990 in die alten Bundesländer gekommen sind“. Damit waren Zuwanderungen von Juli bis Dezember 1990, die bereits in der C-Stichprobe erfaßt wurden, ebenfalls in

der D1 Stichprobe enthalten. Bei der Gewichtung wird diese ungenaue Stichprobenabgrenzung nachträglich dadurch korrigiert, daß diese Haushalte generell den Gewichtungsfaktor 0 erhalten. Möchte man die Stichprobe D separat als Zuwanderer-Stichprobe auswerten, so werden die Zuwanderer der übrigen SOEP-Teilstichproben nicht berücksichtigt. Umgekehrt wären jedoch die gemischten Haushalte bei einer gemeinsamen Auswertung aller Teilstichproben überrepräsentiert. Dieses Problem läßt sich offensichtlich nur durch die Verwendung von zwei Gewichtungsvariablen lösen: Eine Gewichtungsvariable bei Verwendung von D als separater Stichprobe<sup>7</sup> und eine Gewichtungsvariable bei gemeinsamer Verwendung aller SOEP-Teilstichproben<sup>8</sup>. Um die bisherigen SOEP-Gewichtungsvariablen nicht ändern zu müssen, wurden die gemischten Haushalte der D-Stichprobe bei einer gemeinsamen Auswertung mit den übrigen SOEP-Stichproben mit dem Faktor „Null“ gewichtet<sup>9</sup>.

Überlegungen hinsichtlich der Bereitstellung einer Designgewichtung legen den Ausschluß der Schneeballstichprobe nahe, die auch wegen ihrer starken Klumpung keine guten Stichprobeneigenschaften hat. Wie in Schulz et al. (1993, S. 28 ff.) gezeigt wurde, benötigt man für die Bestimmung der Ausfallwahrscheinlichkeiten nach dem Schneeballverfahren die Anzahl der Haushalte, die einen anderen Haushalt im Rahmen des Schneeballverfahrens nennen. Diese Information ist zwar für die D1-Haushalte verfügbar, nicht jedoch für die D2-Haushalte. Somit kann für die D2-Haushalte nicht die Eintrittswahrscheinlichkeit in D bestimmt werden, die sich näherungsweise als

<sup>7</sup> Diese wird bei der Namensvergabe durch das Anhängen des Buchstabens „D“ gekennzeichnet, also „LPHRFD“ für Personenquerschnitte in Welle L (1995) und „LHHRFD“ für Haushaltsquerschnitte in Welle L.

<sup>8</sup> Diese Variable folgt bei der Namensvergabe der üblichen Konvention im SOEP, also „LPHRF“ für Personenquerschnitte in Welle L und „LHHRF“ für Haushaltsquerschnitte in Welle L.

<sup>9</sup> Alternativ hätte man für alle gemischten Haushalte die Auswahlwahrscheinlichkeiten über die Stichproben A, B, C oder D schätzen müssen.

Summe der Inklusionswahrscheinlichkeiten über D1 und D2 ergibt. Auf dem Konstruktionsprinzip der Gewichte als Kehrwert der Auswahlwahrscheinlichkeiten basiert aber die gesamte Gewichtung des SOEP, so daß allein aus Gründen der methodischen Einheitlichkeit bei der Hochrechnung auf diesen Teil der D-Stichprobe verzichtet werden sollte.

Es läßt sich jedoch noch ein zweiter Grund anführen, auf immerhin 95 der insgesamt 522 Zuwandererhaushalte (18 vH) bei der Hochrechnung zu verzichten. Dieser liegt in der starken Quotierung der D2 Bruttostichprobe nach den drei Zuwanderergruppen. Grundsätzlich existieren methodologische Alternativen zu der Gewichtung über den Kehrwert der Auswahlwahrscheinlichkeit. Im Falle der D-Stichprobe bestehen vier Möglichkeiten, in die Gesamtstichprobe zu gelangen: über die D1 Buseinschaltung, über die darauf aufbauende Schneeballstichprobe, über die D1 Scope Stichprobe und schließlich über die D2 Stichprobe. Für jede der vier Einzelstichproben  $j = 1, \dots, 4$  sind Inklusionswahrscheinlichkeiten  $\pi_{ij}$  zumindest näherungsweise bekannt. Hierbei gibt  $\pi_{ij}$  die Wahrscheinlichkeit an, daß Haushalt  $i$  über die Teilstichprobe  $j$  die Stichprobe D erreicht. Bezeichnet  $C_{ij}$  den entsprechenden Auswahlindikator, so kann man den sogenannten Konvex-Schätzer für Merkmalssummen definieren; vgl. z.B. Brick und Kalton (1985) sowie Rendtel (1997). Dieser schätzt das Gesamtaufkommen

$$T = \sum_{i \in G} y_i$$

eines Merkmals  $y$  in einer Grundgesamtheit  $G$  durch

$$\hat{T} = \sum_{i \in G} \sum_{j=1}^4 \lambda_{ij} \frac{C_{ij}}{\pi_{ij}} y_i$$

Hierbei sind  $\lambda_{ij}$  ( $j=1, \dots, 4$ )  $\geq 0$  Konvexgewichte, deren Summe für jedes  $i$  den Wert 1 ergibt. Dabei ist es wesentlich, daß die Konvexgewichte  $\lambda_{ij}$  nicht von den Auswahlindikatoren  $C_{ij}$  abhängen. Wie man leicht sieht, ist  $\hat{T}$  eine erwartungstreue Schätzfunktion für  $T$ .

Die Quotierung der D2 Bruttostichprobe äußert sich in einer Koppelung des Auswahlindikators  $C_{i4}$  mit einem der wichtigsten Merkmale der D-Stichprobe, der Zugehörigkeit zu den Zuwanderergruppen. Würde man den Konvex-Schätzer  $\hat{T}$  für die Gewichtung von D verwenden, so erhielte man eine Konvexkombination der Ergebnisse über die Teilstichproben von D. Da die Zusammensetzung der Zuwanderergruppen jedoch systematisch zwischen den Teilstichproben schwankt, lassen sich für unterschiedliche Werte von  $\lambda_{ij}$  alle möglichen Anteilswerte der Zuwanderergruppen als Kombination der Werte in Tabelle 2 erzeugen. Somit entscheidet die Wahl der  $\lambda_{ij}$  jedoch automatisch über den Schätzwert wichtiger Kenngrößen der Zuwandererpopulation. Diese Eigenschaft spricht gegen die Verwendung von  $\hat{T}$ .

Schließlich wurde entschieden, die Sub-Stichprobe D1 lediglich als Stichprobe von Aus- und Übersiedlern zu nutzen, da sonstige Zuwanderer nur in einem sehr geringen Anteil (8,3 vH, vgl. Tabelle 2) ermittelt wurden und somit eine offensichtliche Untererfassung bei der Feldarbeit vorlag.

Die Hochrechnung der Teilstichproben A, B und C wird von der Existenz der Teilstichprobe D in marginaler Weise berührt, da bis zur Erhebung der Zuwandererstichprobe D für das SOEP die Annahme gemacht werden mußte, daß die alten Teilstichproben A, B und C auch alle Zuwanderergruppen repräsentieren. Bezüglich der Zuwanderer, die in bereits bestehende Haushalte zogen (direkt nach der Zuwanderung oder später) ist diese Annahme auch richtig. Sie gilt jedoch nicht für Zuwanderer, die neue Haushalte gründeten (und in die später keine „Alt-Bürger“ hineinziehen). Trotz der Verfügbarkeit der Teilstichprobe D sind für die Erhebungsjahre 1984 bis 1994 keine neuen Informationen über Zuwanderer hinzugekommen, die die SOEP-Daten und SOEP-Hochrechnungen für diesen Zeitraum verändern könnten.

Deswegen bleiben die Hochrechnungsfaktoren der Wellen A (1984) bis K (1994) unverändert. Ab Welle L (1995) werden die Hochrechnungsfaktoren für die Haushalte der Teilstichproben A, B und C, die in Westdeutschland leben, jedoch um die Masse der Zuwanderer vermindert, die nun von Teilstichprobe D repräsentiert wird. Mangels weiterer Informationen geschieht dies durch eine gleichmäßige Reduktion um den Faktor 0.949<sup>10</sup>. Mit anderen Worten: Die Hochrechnungsfaktoren der westdeutschen Haushalte in den Teilstichproben A, B und C werden an die Bevölkerungseckzahl ohne Zuwanderer angepaßt. Die gleichmäßige Verminderung der Hochrechnungsfaktoren hat den Vorteil, daß diese durch eine einfache Multiplikation der Hochrechnungsfaktoren für die Teilstichproben A und B um den reziproken Wert  $1/0.949 = 1.053$  dem bisherigen Gewicht entsprechen. D.h. daß insbesondere methodisch konsistente Querschnitts-Vergleiche auf Basis der Teilstichproben A und B von 1984 bis 1994 und 1995 sowie darüber hinaus ohne weiteres möglich sind.

#### 4.2 Die Designgewichtung der D-Stichprobe

Unter der Designgewichtung wird hier die näherungsweise Bestimmung der Inklusionswahrscheinlichkeiten verstanden. Hierbei bezeichnen  $\pi_1$  die Inklusionswahrscheinlichkeit von Haushalten über den Infratest-Bus in D1,

<sup>10</sup> Zu diesem Faktor gelangt man folgendermaßen: Die Anzahl der in Westdeutschland lebenden Haushalte, die durch die Teilstichproben A, B und C nicht repräsentiert werden, erhält man als Summe der Hochrechnungsfaktoren LHHRF. Man erhält 1530 Tsd. Haushalte. Bezogen auf diese Grundgesamtheit von ca. 28625 Tsd. Privathaushalten in Westdeutschland sind dies 5.3 vH. Folglich muß die Eckzahl für die Haushaltsgesamtheit um den Faktor 0.949 verringert werden.

$\pi_2$  die Inklusionswahrscheinlichkeit über Infratest-Scope in D1 und  $\pi_3$  die Inklusionswahrscheinlichkeit über den Infratest-Bus in D2.

Die Auswahlwahrscheinlichkeit für die Auswahl in Stichprobe D<sup>11</sup>,  $\pi_D$ , ergibt sich dann, unter Vernachlässigung der gemeinsamen Auswahlwahrscheinlichkeiten, als Summe der Einzelwahrscheinlichkeiten, d.h.  $\pi_D \approx \pi_1 + \pi_2 + \pi_3$ . Lediglich für die Gruppe der sonstigen Zuwanderer ergibt sich nur eine Möglichkeit, in die Stichprobe zu gelangen: Über den Infratest-Bus D2. Folglich gilt für diese Gruppe  $\pi_D = \pi_3$ .

Ein Haushalt gilt als Mitglied der Stichprobe D, wenn er dem Nettobestand in Welle L (1995) angehört. Diese Festlegung ist insofern wichtig, als die Haushalte der D1-Stichprobe schon seit der Welle K (1994) befragt werden. Es erscheint aber vorteilhaft, die Inklusionsentscheidung der Gesamtstichprobe D auf einen Zeitpunkt zu datieren.

Die Berechnung der Inklusionswahrscheinlichkeiten wird getrennt für die Gruppen Übersiedler, Aussiedler und sonstige Zuwanderer durchgeführt. Lediglich für diese Gruppenabgrenzungen sind die notwendigen Informationen über alle Auswahlstufen vorhanden. Diese Abgrenzung erscheint auch im Sinne der angestrebten Analysen sinnvoll.

Tabelle 3 dokumentiert für diese drei Gruppen die einzelnen Faktoren, deren Produkt näherungsweise die Inklusionswahrscheinlichkeiten  $\pi_1$ ,  $\pi_2$ ,  $\pi_3$  ergeben. Unter der Zeile „Design“ sind zunächst die Auswahlwahrscheinlichkeiten aufgeführt, die sich aus dem ADM-Design ergeben. Hierbei kann näherungsweise für jeden Haushalt dieselbe Auswahlwahrscheinlichkeit unterstellt werden, so daß sich diese als Anzahl n der Haushalte im Screening-Verfahren, dividiert durch die Anzahl N aller Privathaushalte ergibt, vgl. Tabelle 1.

Für die Telefonbefragung über InfraScope wurde die Anzahl der über Telefon erreichbaren Haushalte mit der Anzahl der Privathaushalte gleichgesetzt, was bei einer Telefonausstattung von über 95 vH in Westdeutschland gerechtfertigt erscheint. Die Auswahl der Telefonnummern wurde unter Berücksichtigung regionaler Schichtungskriterien über eine direkte Auswahl von Telefonnummern aus dem Gesamtverzeichnis der Telefonanschlüsse vorgenommen. Deshalb erscheint die Annahme gleicher Auswahlwahrscheinlichkeiten für Haushalte in guter Näherung erfüllt zu sein.

Die nächste Zeile von Tabelle 1 dokumentiert die unterschiedlichen Anteile an Zuwanderer-Haushalten, die von 3,5 vH (D1-Bus) über 4,7 vH (D1-Scope) und 6,6 vH (D2-Bus) reichen. Die niedrigeren Anteile für die D1-Stichprobe sind unter anderem auch dem früheren Auswahltermin 1992 (gegenüber 1994) geschuldet. Trotzdem verbirgt sich hinter dem gestiegenen Anteilswert für den D2-Bus auch eine verbesserte Feldarbeit bei der Ermittlung der Zuwandererhaushalte. Zunächst wurde diese Ermittlung bei der Berechnung der Auswahlwahrscheinlichkeiten

Tabelle 3

**Die Berechnung der Inklusionswahrscheinlichkeiten für die Zuwanderer — Teilstichprobe D**

$\pi_1$ : Inklusionswahrscheinlichkeiten über Infratest-Bus in D1		
	Übersiedler	Aussiedler
Design	1/7194	1/7194
Einverständnis	0.787	0.787
Netto 1994	0.771	0.779
Netto 1995	0.936	0.908
Produkt	1/12667	1/12923

$\pi_2$ : Inklusionswahrscheinlichkeiten über Infratest-Scope in D1		
	Übersiedler	Aussiedler
Design	1/7194	1/7194
Einverständnis	0.444	0.444
Netto 1994	0.771	0.779
Netto 1995	0.936	0.908
Produkt	1/22452	1/22906

$\pi_3$ : Inklusionswahrscheinlichkeiten über Infratest-Bus in D2			
	Übersiedler	Aussiedler	Sonstige
Design	1/2687	1/2687	1/2687
Einverständnis	0.756	0.756	0.756
Quotierung	0.527	0.696	1.00
Netto 1995	0.766	0.833	0.655
Produkt	1/8804	1/6130	1/5426
<i>Quelle:</i> Infratest Methodenbericht; eigene Berechnungen.			

nicht berücksichtigt. Im Gegensatz zu den nachfolgenden Randomisierungsschritten (Einwilligung zur Adressenspeicherung, Realisierung des Interviews) ist bei dem Identifizierungsanteil die Basis aller zu identifizierenden Haushalte unbekannt. Dies hat zur Folge, daß die Gesamtzahl der Zuwandererhaushalte in der Tendenz unterschätzt wird.

Bei der Einverständnisquote für die Adressenspeicherung (Zeile 4 in Tabelle 1 bzw. 2. Faktor in Tabelle 3) liegt der Wert bei der Telefonstichprobe (44,4 vH) naturgemäß deutlich unter den Werten bei den persönlichen Interviews. Für eine Differenzierung dieser Quoten nach Zuwanderergruppen fehlen die erforderlichen Informationen.

<sup>11</sup> Bei Ausschluß der Schneeballstichprobe.

Die nächsten beiden Faktoren bei der Bestimmung der Inklusionswahrscheinlichkeiten sind in den Sub-Stichproben D1-Bus und D1-Scope die Chancen für die Realisierung eines Interviews in Welle K (1994) und in Welle L (1995), vgl. die Zeilen „Netto 1994“ und „Netto 1995“ in Tabelle 3. Diese Erfolgsquoten wurden nach den drei Zuwanderergruppen differenziert. Allerdings wurde hierbei nicht mehr zwischen D1-Bus und D1-Scope unterschieden. Dies wurde für die Welle K (1994) durch fehlende weitergehende Angaben über den Bruttobestand erzwungen. Für die Welle L (1995) hätte hier durchaus die Möglichkeit zu einer differenzierten Ausfallanalyse bestanden, da nun alle Informationen aus der vorhergehenden Welle K für eine Ausfallanalyse bereitstehen. Angesichts der relativ groben Unterteilung auf den übrigen Auswahlritten erscheint jedoch eine feine Differenzierung der Ausfallrisiken auf dieser letzten Auswahlstufe kaum angebracht. Es zeigt sich zudem, daß die Ausfallrisiken für die drei Zuwanderergruppen nur unwesentlich voneinander verschieden sind, vgl. Tabelle 1, Zeilen unter „Realisierungsquote 1995 in vH“.

Wie bereits erwähnt, wurde der Adressenbestand von D2, der im folgenden als „Brutto-D2“ bezeichnet wird, nur zu zwei Drittel ausgeschöpft. Die Ausschöpfungsquoten variieren hier systematisch über die Zuwanderergruppen<sup>12</sup>, da die Sub-Stichprobe D2 die Sub-Stichprobe D1 — inklusive der Schneeballstichprobe und der Gruppe der sonstigen Zuwanderer — so „ausfüllen“ sollte, daß insgesamt die prozentuale Aufteilung nach Herkunftsländern in Brutto-D2 reproduziert wird. Wegen der Nichtberücksichtigung der Schneeballstichprobe mit ihrem hohen Anteil an Aussiedler-Haushalten (54,7 vH der Schneeballstichprobe sind Aussiedler-Haushalte, vgl. Tabelle 2) reproduziert die für die Gewichtung berücksichtigte Stichprobe diese prozentuale Aufteilung nicht mehr exakt<sup>13</sup>, vgl. Tabelle 2.

Bei der Gewichtung wurde die Quote wie eine zufällige Auswahl aus dem Adressenbestand von Brutto-D2 behandelt, wobei der Auswahlatz über die Zuwanderergruppen schwankt, vgl. Zeilen unter „Quotierung der D2-Stichprobe“ in Tabelle 1 und „Quotierung“ in Tabelle 3.

Schließlich ist noch die Realisierung von Interviews in Welle L (1995) für die Sub-Stichprobe D2 zu berücksichtigen. Hier liegt Information zunächst nur auf der Ebene der Zuwanderergruppen vor. Es zeigt sich, daß die Feldarbeit bei den Übersiedlern (Realisierungsquote 76,6 vH) und erst recht bei den Aussiedlern (Realisierungsquote 83,3 vH) deutlich über der — durchaus üblichen — Realisierungsquote von 65,5 vH bei den sonstigen Zuwanderer-Haushalten liegt, vgl. Zeilen unter „Realisierungsquote 1995 in vH“ in Tabelle 1 bzw. Zeile „Netto 1995“ für  $\pi_3$  in Tabelle 3.

Dieses Resultat ist insofern bemerkenswert, als bei den sonstigen Zuwanderer-Haushalten die größten Sprachprobleme bei der Durchführung der Interviews zu erwarten waren. Der Gefahr, in erster Linie bereits „etablierte“, d.h. relativ integrierte Zuwanderer-Haushalte, zu treffen, wurde durch die bei Erstbefragungen im SOEP vorgeschriebene hohe Zahl von 10 Kontaktversuchen begegnet. Falls Sprachprobleme bei den Zuwanderern auftraten, durften gegebenenfalls auch Begleitpersonen als Dolmetscher fungieren. Frühere Erfahrungen haben gezeigt, daß gut

<sup>12</sup> Für die Haushalte aus Brutto-D2 ist lediglich das Zugangsland dokumentiert. Für die Abgrenzung der Aussiedler wurden die folgenden Herkunftsländer ausgewählt: Polen, die Nachfolgestaaten der UDSSR sowie Rumänien.

<sup>13</sup> Auch weicht die designgewichtete D-Stichprobe hinsichtlich der Zuwanderergruppen deutlich von der ungewichteten D-Stichprobe ab.

Tabelle 4

**Deutschkenntnisse der Kontaktperson bei realisierten Haushalten:  
Auskunftsperson für Haushaltsfragebogen**  
Fallzahlen (vH-Werte in Klammern)

	D1		D2	
	Brutto	Netto	Brutto	Netto
Sehr gut	26 (13)	13 (9)	163 (41)	128 (43)
Gut	21 (11)	16 (12)	119 (30)	88 (30)
Es geht	18 (9)	15 (11)	80 (20)	55 (19)
Schlecht/gar nicht	2 (1)	0 (0)	38 (9)	24 (8)
Keine Angabe	101 (52)	94 (68)	—	—
Trifft nicht zu	27 (14)	0 (0)	—	—
<b>Summe</b>	<b>195 (100)</b>	<b>138 (100)</b>	<b>400 (100)</b>	<b>295 (100)</b>

Quelle: Infratest Methodenbericht.

Tabelle 5

## Berechnung der Inklusionswahrscheinlichkeiten für die Zuwandererstichprobe D

$\pi_D$ : Auswahl eines Haushalts in die Stichprobe D (Netto 1995)			
	Übersiedler	Aussiedler	Sonstige
Berechnung	$\pi_D \approx \pi_1 + \pi_2 + \pi_3$	$\pi_D \approx \pi_1 + \pi_2 + \pi_3$	$\pi_D = \pi_3$
Wert	0.000237 = 1/4218	0.000284 = 1/3519	0.00018 = 1/5426
<i>Quelle:</i> Eigene Berechnungen.			

geschulte und motivierte Interviewer trotz eines deutschsprachigen Fragebogens befriedigende Ergebnisse erzielen (vgl. Schupp/Wagner 1995, S. 19). In der Tat zeigt ein Vergleich der Verteilung nach Deutschkenntnissen der Kontaktperson, daß schlechte Sprachkenntnisse vermutlich kaum eine Rolle bei der Realisierung der Interviews spielten, vgl. Tabelle 4. Die Aussagekraft wird in D1 etwas eingeschränkt durch die hohe Zahl derjenigen, die keine Angabe zu dieser Frage gemacht haben.

Insgesamt ergeben sich die Inklusionswahrscheinlichkeiten als Produkt der vorhergehenden Faktoren. Sie sind in der Zeile „Produkt“ von Tabelle 3 dokumentiert.

Abschließend gibt Tabelle 5 die Wahrscheinlichkeit wieder, über eine der drei Sub-Stichproben von D in die Zuwandererstichprobe von D zu gelangen. Für die Bestimmung der Auswahlwahrscheinlichkeiten ist es dabei unerheblich, in welcher der drei Sub-Stichproben der Haushalt tatsächlich auftritt. Für die Gruppe der sonstigen Zuwanderer kann dies allerdings nur die Sub-Stichprobe D2 sein. Die aus den Ziehungswahrscheinlichkeiten resultierenden Hochrechnungsfaktoren (Übersiedler: 4218, Aussiedler: 3519 und Sonstige: 5426) liegen für Aussiedler etwas unter dem durchschnittlichen Startwert für die Stichprobe A (4500 in Welle A, 1984). Dies zeigt, daß Aussiedler in der SOEP-Gesamtstichprobe überproportional vertreten sind.

Umgekehrt sind die sonstigen Zuwanderer in der SOEP-Gesamtstichprobe unterproportional vertreten. Die Gruppe der Übersiedler bis 1990 entspricht in ihren Ziehungswahrscheinlichkeiten in etwa dem Niveau der Stichprobe A.

## 5. Die externe Validierung der Zuwandererstichprobe

### 5.1 Möglichkeiten zur Validierung der Zuwandererstichprobe

Bei der Berechnung der externen Validierung einer Stichprobe werden Schätzergebnisse auf Basis der Design-Gewichtung mit den Ergebnissen der amtlichen Statistik sowie mit Schätzergebnissen aus anderen Erhebungen verglichen. Wie bereits erwähnt, liefert die amtliche Statistik keine genauen Zahlen über die Zahl der

Zuwanderer seit 1984, die in Privathaushalten leben.

Auf Basis unterschiedlicher Quellen wurde aber vom DIW (vgl. Schulz, 1994) eine Schätzung vorgelegt, die unter Zuhilfenahme von plausiblen Annahmen über Altersverteilung, Haushaltsgröße, Überwecheln vom Anstaltsbereich sowie Rückkehr in die Herkunftsregion Eckzahlen für die Anzahl der Zuwandererhaushalte im Jahr 1993 und deren Mitglieder präsentiert. Diese Zahlen werden im folgenden als „DIW-Schätzung“ zitiert.

Eine weitere unabhängige Quelle zur Abschätzung von Eckzahlen auf Personenebene stellt die ALLBUS-Erhebung aus dem Jahre 1994 dar.

Nicht zuletzt kann die Bruttostichprobe von D2 für eine Hochrechnung der Zuwandererhaushalte nach Herkunftsländern genutzt werden, da laut Aussage von Infratest dieser Bruttostichprobe eine im Vergleich zu D1 höhere Qualität bei der Feldarbeit bescheinigt werden kann.

## 5.2. Vergleiche auf Haushaltsebene

Tabelle 6 vergleicht auf Haushaltsebene die Hochrechnungsergebnisse der D-Stichprobe mit der DIW-Schätzung und einer Hochrechnung auf Basis der Brutto-D2-Stichprobe<sup>14</sup>.

Zunächst einmal stimmen die DIW-Schätzung und die Schätzung über Brutto-D2 hinsichtlich des Gesamtumfangs der Zuwandererhaushalte gut überein (1980 bzw. 1926 Tausend Zuwandererhaushalte).

Größere Abweichungen ergeben sich aber hinsichtlich der Aufteilung nach den Zuwanderergruppen. Hier liegen die Anteile der Übersiedler und Aussiedler in Brutto-D2 um jeweils ca. 10 Prozentpunkte über den entsprechenden Anteilswerten in der DIW-Schätzung. Als Konsequenz liegt der Anteil der Haushalte der sonstigen Zuwanderer um fast 20 Prozentpunkte unter dem entsprechenden Anteil in der DIW-Schätzung.

<sup>14</sup> Der Hochrechnungsfaktor ergibt sich dabei aus dem Auswahlatz der Stichprobe D2 multipliziert mit dem Kehrwert der Einwilligungsquote, also  $2687/0.756=3554$ .

Tabelle 6

**Vergleich der Hochrechnung der Zuwanderer-Haushalte mit der DIW-Schätzung und der Brutto-D2 Stichprobe**  
Prozentwerte in Klammern

	Zuwanderer-Stichprobe D		DIW-Schätzung	Brutto-D2	
	ungewichtet	gewichtet			
	Anzahl	HR in Tsd.*	in Tsd.	Anzahl	in Tsd.
Übersiedler	82 (21.7)	346 (21.1)	370 (18.7)	146	519 (27.0)
Aussiedler	164 (43.4)	577 (35.2)	580 (29.3)	217	771 (40.0)
Sonstige	132 (34.9)	716 (43.7)	1 030 (52.0)	179	636 (33.0)
Summe	378 (100.0)	1 639 (100.0)	1 980 (100.0)	542	1 926 (100.0)

\* HR in Tsd.: Hochrechnung in Tausend.  
Quelle: Infratest Methodenbericht; Schulz 1994; eigene Berechnungen.

Die Hochrechnung der D-Stichprobe unterschätzt mit 1636 Tsd. den Richtwert von ca. 1950 Tsd. um ca. 300 Tsd.. Die Gründe für diese Unterschätzung liegen, wie bereits erwähnt, in einer ungenügenden Erfassung des Merkmals „Zuwandererhaushalt“, was eine gewisse Kooperationsbereitschaft der Haushalte voraussetzt. Diese mangelnde Erfassung konnte bei der Designgewichtung nicht berücksichtigt werden. Weiterhin spielte auch der frühe Start der Stichprobe D1 im Jahre 1992 eine Rolle, da zu diesem Zeitpunkt weniger Zuwanderer in Deutschland lebten als im Jahre 1994.

Hinsichtlich der Anteile der drei Zuwanderergruppen weicht die ungewichtete D Stichprobe (ohne Schneeballstichprobe von D1) nur mäßig von der Brutto D2 ab. Diese Übereinstimmung basiert auf der bereits erwähnten Quotierung der D2-Stichprobe nach den Zuwanderergruppen. Der Ausschluß der D1-Schneeballstichprobe bewirkt hier in erster Linie einen Rückgang des Anteils der Übersiedler und eine Vergrößerung der Anteile der Aussiedler und der übrigen Zuwanderer.

Die Designgewichtung der Zuwanderer-Stichprobe bewirkt eine deutliche Veränderung der Anteilswerte in Richtung auf die Werte der DIW-Schätzung. Die designgewichtete Stichprobe liegt hinsichtlich aller Anteilswerte am dichtesten an dem entsprechenden Wert der DIW-Schätzung.

Hinsichtlich der Absolutzahlen stimmen die Ergebnisse der Design-Hochrechnung für die Gruppen der Übersiedler und der Aussiedler bemerkenswert gut mit der DIW-Schätzung überein: Übersiedler: 346 Tsd. (Hochrechnung) vs. 370 Tsd. (DIW); Aussiedler: 577 Tsd. (Hochrechnung) vs. 580 Tsd. (DIW). Lediglich für die Gruppe der sonstigen Zuwanderer ergibt sich eine Abweichung von 30 Prozent. Da sich diese Gruppe ausschließlich über die Teilstichprobe D2 rekrutiert, können die Feldprobleme, die in D1 bei der Identifikation ausländischer Zuwandererhaushalte auf-

treten, nicht für diese Abweichungen herangezogen werden. Es verbleiben jedoch noch eine Reihe von weiteren Gründen, die derartige Abweichungen erzeugen können:

- Die Untererfassung von Sample-Points mit hohem Ausländeranteil durch das ADM-Design, das die Basis der D2-Stichprobe war.
- Eine potentielle Überschätzung der Gruppe der sonstigen Zuwanderer im Rahmen der DIW-Schätzung.

Mögliche Gründe für eine Überschätzung der Haushalte der übrigen Zuwanderer in der DIW-Schätzung sind:

- Annahme zu kleiner Haushaltsgrößen.
- Unterschätzung des Anteils der noch im Anstaltsbereich lebenden Zuwanderer.
- Mögliche Abgrenzungsprobleme mit der Gruppe der Aussiedler, wenn Personen als Nationalität ihr Auswanderungsland angeben.

Eine abschließende Diskussion dieser Abweichungen wird nach einem Vergleich der entsprechenden Personenzahlen durchgeführt.

### 5.3 Vergleiche auf Personenebene

#### 5.3.1 Die Vergleichs-Stichprobe ALLBUS 1994<sup>15</sup>

Beim ALLBUS 1994 kam erstmalig eine Gemeindestichprobe mit anschließender Ziehung von Adressen aus den Einwohnermelderegistern zum Einsatz. Im Unterschied zu dem bisher benutzten ADM-Stichprobensystem wird eine

<sup>15</sup> Der ALLBUS wird durchgeführt und konzeptionell betreut vom Zentrum für Umfragen, Methoden und Analysen in Mannheim (ZUMA) in Zusammenarbeit mit dem Zentralarchiv für empirische Sozialforschung der Universität zu Köln (ZA); vgl. Koch et al. (1994).

Stichprobe auf Personenebene gezogen, die allen Personen dieselbe Auswahlwahrscheinlichkeit zuordnet. Die Grundgesamtheit bildet die in Privathaushalten lebende erwachsene Bevölkerung Deutschlands. Befragt wurden Personen im Alter von mindestens 18 Jahren. Dabei wurde einheitlich ein deutschsprachiger Fragebogen verwendet. Diese Stichprobe kann als außerordentlich zuverlässig gelten. Allerdings ist die Präzision der Schätzungen für Zuwanderer aufgrund der kleinen Fallzahl im ALLBUS recht ungenau.

Bei der Identifizierung der Zuwanderer im ALLBUS waren einige Besonderheiten zu beachten. So lassen sich Personen nur dann als Übersiedler identifizieren, wenn sie die deutsche Staatsangehörigkeit besitzen, auf dem Gebiet der heutigen östlichen Bundesländer oder Ostberlins geboren sind und nach Westdeutschland übergesiedelt sind. Problematisch ist hier vor allem die Charakterisierung über den Geburtsort. Diese schließt bestimmte Personen von der Identifikation als Übersiedler aus. So werden zum einen Vertriebene aus den ehemaligen ostdeutschen Gebieten (Schlesien, Ostpreußen, Pommern), die sich anschließend auf dem Gebiet der DDR niedergelassen haben, nicht als Übersiedler erfaßt, wenn sie später nach Westdeutschland ziehen. Weiterhin bleiben alle Personen unberücksichtigt, die auf westdeutschem Gebiet geboren sind und zum Zeitpunkt der Gründung der DDR auf DDR-Gebiet lebten und später das Gebiet der DDR wieder verlassen haben. Die Zahl der Übersiedler wird aus diesem Grunde vermutlich etwas unterschätzt.

### 5.3.2 Die DIW-Schätzung

Die amtlichen Statistiken, auf denen die DIW-Schätzung basiert, beruhen auf Personenangaben. Die Schätzung der Haushaltszahlen basiert lediglich auf einer Division dieser Personenzahlen durch eine ebenfalls geschätzte durchschnittliche Haushaltsgröße. Dies verleiht den Schätzwerten auf Personenebene eine etwas höhere Verlässlichkeit. Die geschätzten Zuwandererzahlen auf Personenebene sind in Tabelle 6 von Schulz (1994) zusammengefaßt. Allerdings ist dort die Altersabgrenzung (0-16 Jahre, 16 Jahre und älter) verschieden von der Abgrenzung beim ALLBUS (18 Jahre und älter), so daß für einen gemeinsamen Vergleich mit dem ALLBUS lediglich die Anteile der drei Zuwanderergruppen herangezogen werden.

### 5.3.3 Ergebnisse des Vergleichs

Die Tabelle 7 vergleicht die Ergebnisse der Design-Hochrechnung für Personen ab 18 Jahren mit den hochgerechneten Ergebnissen des ALLBUS<sup>16</sup> und den Gruppenanteilen der DIW-Schätzung.

<sup>16</sup> Für die Hochrechnungen des ALLBUS wurde eine Eckzahl von 53257 Tsd. Personen über 18 Jahren verwendet. Diese Eckzahl ergibt sich als die Anzahl der Personen in Privathaushalten laut Mikrozensus 1993 (63934) multipliziert mit 0,833. Letzterer Faktor wurde für den Anteil der über 18jährigen für den ALLBUS benutzt. Auf Basis dieser Eckzahlen erhält man einen Auswahl-satz von  $53257000/2342=22740$ , der als Hochrechnungsfaktor benutzt wurde.

Tabelle 7

**Vergleich der unkorrigierten Design-Gewichtung der D-Stichprobe für Personen ab 18 Jahren mit den Resultaten des ALLBUS 1994 und mit der DIW-Schätzung**  
Prozentwerte in Klammern

	Stichprobe D		ALLBUS 1994		DIW-Schätzung*
	Anzahl	HR in Tsd.	Anzahl	HR in Tsd.	Anteile der Gruppen
Übersiedler	136 (19.0)	574 (18.9)	23	523 (20.2)	(15.1)
Aussiedler	357 (49.8)	1 256 (41.2)	48	1 091 (42.1)	(29.9)
Sonstige	224 (31.2)	1 215 (39.9)	43	977 (37.7)	(53.5)
Summe	717 (100.0)	3 045 (100.0)	114	2 591 (100.0)	(100.0)
Zuzugsjahr missing			20	453	
Summe			134	3 044	

\* Personen 16 Jahre und älter.  
Quelle: Eigene Berechnungen; ZUMA-Arbeitsbericht zum ALLBUS 1994; Schulz 1994.

Insgesamt können im ALLBUS 114 Zuwanderer in Westdeutschland identifiziert werden, die nach 1983 zugezogen sind und der Zuwandererpopulation des SOEP angehören. Die Verteilung dieser Personen nach den drei Zuwanderergruppen stimmt mit der Design-Gewichtung der D-Stichprobe sehr gut überein. Auch hinsichtlich der Absolutzahlen stimmen beide Erhebungen gut miteinander überein.

Erwartungsgemäß liegt die geschätzte Zahl der Übersiedler im ALLBUS etwas geringer als in der D-Stichprobe. Doch können relative Schwankungen von gut 10 Prozent auch rein zufälliger Natur sein.

Zu beachten ist weiterhin, daß im ALLBUS 20 Personen befragt werden, deren Zuwanderungsjahr fehlt. Hochgerechnet ergibt dies 453 Tsd. Personen, von denen ein unbekannter Anteil der Zuwandererpopulation des SOEP angehört. Maximal kann also die Zuwandererzahl laut ALLBUS den Wert von 3044 Tsd. erreichen, was exakt mit der Design-Hochrechnung von D übereinstimmt.

Hinsichtlich der Anteile der Zuwanderergruppen ergeben sich deutliche Abweichungen zur DIW-Schätzung. Derartige Unterschiede können nicht durch die geringfügig andere Altersabgrenzung erklärt werden<sup>17</sup>. Wie bei den Haushalten liegt die DIW-Schätzung für den Anteil der sonstigen Zuwanderer deutlich über den entsprechenden empirischen Befunden. Auch zeigen sich bei der DIW-Schätzung kaum Unterschiede beim Übergang von der Haushalts- zur Personenebene. Bei der designgewichteten Stichprobe D verschieben sich jedoch die Anteile zwischen Aussiedlern und den übrigen Zuwanderern bedeutend: So steigt der Anteil der Aussiedler von 35,2 vH (Haushalte) auf 41,2 vH (Personen über 18 Jahre) während er bei den übrigen Zuwanderern von 43,7 vH (Haushalte) auf 39,9 vH (Personen über 18 Jahre) absinkt. Eine derartige Umkehrung der Anteile ist nur bei unterschiedlichen Haushaltsgrößen möglich. In der Tat weist die DIW-Schätzung (vgl. Tabelle 6 von Schulz (1994)) für die Aussiedler und für die ausländischen Zuwanderer<sup>18</sup> die gleiche geschätzte durchschnittliche Haushaltsgröße von 2,7 aus. Die empirischen Befunde sprechen jedoch für höhere Zahlen bei den Aussiedlern bzw. niedrigere Zahlen bei den übrigen Zuwanderern. So liegt im SOEP beispielsweise die durchschnittliche Haushaltsgröße der Aussiedlerhaushalte bei 3,5 und bei übrigen Zuwanderern bei 2,9.

Allerdings sind diese empirischen Haushaltsgrößen nicht direkt mit der Haushaltsgröße der DIW-Schätzung vergleichbar. Diese ergibt sich als Summe der echten Zuwanderer, also als Haushaltsgröße abzüglich der nicht zugewanderten Haushaltsmitglieder. Die nicht zugewanderten Haushaltsmitglieder bestehen zum einen aus erwachsenen Personen, die schon 1984 im Gebiet der alten BRD gelebt haben, und Kindern, die ebenfalls im Gebiet der alten BRD — zum Teil nach der Zuwanderung — geboren wurden.

Die Abgrenzung dieser Kinder ist in der D-Stichprobe problematisch, da diese letztendlich nur über das Zuwan-

derungsjahr der Mutter bestimmt werden kann. Leichter ist die Abgrenzung der Erwachsenen, die keine Zuwanderer sind. Es zeigt sich, daß die Haushaltszahlen sensibel auf derartige Modifikationen reagieren. Man erhält beispielsweise für die durchschnittliche Haushaltsgröße unter Ausschluß der erwachsenen Nicht-Zuwanderer die folgenden Durchschnittswerte (Tabelle 8).

Betrachtet man die Werte in Tabelle 8, so erscheint die empirisch bestätigte Zunahme des Anteils der Aussiedler beim Übergang zur Personenebene durchaus plausibel. Die Rolle der Haushaltsgröße wird bei der Korrektur der Designhochrechnung im folgenden Abschnitt abschließend diskutiert.

Die folgende Tabelle 9 vergleicht die DIW-Schätzung mit der Hochrechnung der D-Stichprobe auf niedrigerem Aggregationsniveau. Hierbei ist die Gruppe der sonstigen Zuwanderer aufgespalten nach Flüchtlingen/Asylbewerbern, sonstigen Deutschen und sonstigen Ausländern. Diese Untergliederung wurde auch für die DIW-Schätzung benutzt, vgl. Tabelle 12 bei Schulz (1994).

Tabelle 9 zeigt eine bemerkenswert gute Übereinstimmung der geschätzten Zahlen der Übersiedler und der Aussiedler mit den DIW-Angaben (Übersiedler: 586 Tsd. (Stichprobe D) vs 622 Tsd. (DIW); Aussiedler: 1209 Tsd. (Stichprobe D) vs 1232 Tsd. (DIW)). Wie auch auf Haushaltsebene wird die Anzahl der übrigen Zuwanderer deutlich unterschätzt.

Die Aufspaltung nach den drei Teilgruppen übrige Deutsche, Flüchtlinge/Asylbewerber und übrige Ausländer gibt die Möglichkeit, eine in der Sozialforschung häufig geäußerte Hypothese zu prüfen. Diese Hypothese besagt, daß Flüchtlinge und Asylbewerber bei der Feldarbeit nur unzureichend erfaßt werden. Gründe hierfür seien einerseits Sprachprobleme und andererseits eine gewisse „Behördenscheuheit“, die dieser Gruppe unterstellt wird. Wie die empirischen Ergebnisse der Feldarbeit zeigten, scheinen Sprachprobleme nur eine untergeordnete Rolle im Teilnehmerverhalten gespielt zu haben, vgl. Tabelle 4. Vergleicht man weiterhin den Grad der Unterschätzung der DIW-Angaben über die drei Teilgruppen, so wird die Gruppe der übrigen Deutschen am deutlichsten unterschätzt (103/260 = 0.396). Die „Risikogruppe“ der Flüchtlinge/Asylbewerber weist eine noch höhere Nachweisquote (412/680 = 0.605) auf als die der übrigen Ausländer (733/1326 = 0.552). Diese empirischen Befunde widersprechen eindeutig der o.g. Hypothese.

## 6. Revision der Design-Gewichtung

Die Ergebnisse der externen Validierung der Designgewichteten Stichprobe sind widersprüchlich in bezug auf

<sup>17</sup> Beim Übergang zu den 16jährigen von Tabelle 7 nach Tabelle 8 vergrößert sich der Anteil der sonstigen Übersiedler von 39,9 vH (18 Jahre) auf 41,0 vH (16 Jahre).

<sup>18</sup> Dies machen 88 Prozent der Gruppe der sonstigen Zuwanderer aus.

Tabelle 8

**Vergleich der durchschnittlichen Haushaltsgröße von Zuwanderer-Haushalten**

	Durchschnittliche Haushaltsgröße		
	unkorrigiert	korrigiert um erwachsene Nicht-Zuwanderer	DIW-Schätzung
Übersiedler	2.7	2.3	2.0
Aussiedler	3.5	3.2	2.7
Sonstige Zuwanderer	2.9	2.4	2.6*

\* Dieser Wert ergibt sich als gewichtetes Mittel der Haushaltsgröße der übrigen Deutschen (Wert = 1.9) und der Ausländer (Wert = 2.7).  
*Quelle:* Tabelle 6 von Schulz (1994); eigene Berechnungen.

die Gruppe der sonstigen Zuwanderer. Bei den Aus- und Übersiedlern ergibt sich eine sehr gute Übereinstimmung hinsichtlich der Absolutzahlen mit der DIW-Schätzung auf Haushaltsebene (Tabelle 6) sowie mit dem ALLBUS und der DIW-Schätzung auf Personenebene.

Die Abweichungen hinsichtlich der Anteilswerte der drei Zuwanderergruppen sind lediglich das Resultat der höheren Basiszahlen für die sonstigen Zuwanderer bei der DIW-Schätzung. Die Anteile lagen auf Haushaltsebene bei 52 vH und auf Personenebene bei 55 vH (Personen über 16 Jahre). Die empirischen Ergebnisse von Brutto-D2 weisen diesen Anteil als 33,0 vH (Haushaltsebene) bzw. 37,7 vH (ALLBUS) aus. Die Design-Gewichtung liegt jeweils zwischen diesen Werten: 43,7 vH (Haushaltsebene) und 39,9 (Personen über 18) bzw. 41,0 vH (Personen über 16). Betragsmäßig wird die DIW-Schätzung für die sonstigen Zuwanderer um den Faktor  $716/1030 = 0.695$  auf Haushaltsebene und um den Faktor  $1248/2266 = 0.550$  auf Personenebene (älter als 16) unterschätzt.

Forscht man nach den Gründen für eine mögliche Untererfassung der sonstigen Zuwanderer, die in ihrer großen Mehrheit Ausländer sind, so bietet sich folgendes Bild:

**Brutto-D2:** Mögliche Untererfassung von Sample-Points mit Ausländern durch das ADM-Design. Aufgrund der Ergebnisse der Feldkontrolle sind Sprachprobleme eher unwahrscheinlich

**ALLBUS 1994:** Eine Untererfassung über das Auswahl-design kann aufgrund der Benutzung der Einwohnermelde-register ausgeschlossen werden. Allerdings ist hier die Feldkontrolle nicht so gut dokumentiert, daß Sprachprobleme ausgeschlossen werden können.

Umgekehrt basiert die DIW-Schätzung der Haushaltszahlen auf Annahmen über die durchschnittliche Haushaltsgröße, die sich nicht mit den empirischen Befunden decken, vgl. Tabelle 8. Man kann nun versuchen, anhand der korrigierten Haushaltsgröße (Haushaltsgröße — Nicht-Zuwanderer) die DIW-Schätzung auf Haushaltsebene zu

Tabelle 9

**Vergleich der Design-Hochrechnung der D-Stichprobe für Personen im Befragungsalter (16 Jahre und älter) und Vergleich mit der DIW-Schätzung**  
 Prozentangaben in Klammern

	D-Stichprobe		DIW-Schätzung
	Anzahl	HR in Tsd.	HR in Tsd.
Übersiedler	139 (18.7)	586 (19.3)	622 (15.1)
Aussiedler	368 (50.0)	1 209 (39.7)	1 232 (29.9)
Übrige Deutsche	19 (2.6)	103 (3.4)	260 (6.3)
Flüchtlinge/Asylbewerber	76 (10.3)	412 (13.5)	680 (16.5)
Sonstige Ausländer	133 (18.3)	733 (24.1)	1 326 (32.2)
Summe	735 (100.0)	3 043 (100.0)	4 120 (100.0)

Tabelle 10

**Alternative Berechnung der Anzahl der Haushalte für die DIW-Hochrechnung  
in Tausend**

	DIW-Schätzung	
	bisher	alternativ
Übersiedler	740/2.0 = 370 (18.7)	321 (16.7)
Aussiedler	1 560/2.7 = 578 (29.3)	488 (25.4)
Sonstige Zuwanderer	2 670/2.6 = 1 027 (52.0)	1 113 (57.9)
Summe	1975 (100.0)	1 922 (100.0)

„verbessern“. Dies führt in diesem Fall jedoch auf eine weitere Vergrößerung des Anteils der Gruppe der sonstigen Zuwanderer, da hier die korrigierte durchschnittliche Haushaltsgröße mit 2.4 unter dem Wert von 2.6 der DIW-Schätzung liegt, während bei den beiden anderen Gruppen das Verhältnis genau umgekehrt liegt. Man erhält die folgenden Werte (Tabelle 10).

Eine derartige Alternative wäre also noch weiter von den empirischen Befunden entfernt und scheidet damit als unplausibel aus.

Da alle Ergebnisse relativ gut mit der DIW-Schätzung für die Aus- und Übersiedler in Einklang stehen, wurde dies als Bestätigung für die Güte der DIW-Schätzung gewertet, die hinsichtlich dieser beiden Gruppen mit den meisten ad-hoc Annahmen behaftet war. Hinsichtlich der Gruppe der übrigen Zuwanderer ist die DIW-Schätzung auf relativ präzise amtliche Angaben gestützt, da es sich bei diesem Personenkreis im wesentlichen um Ausländer handelt. Auch die Abschätzung des Anteils der Personen, die noch im Anstaltsbereich lebt, dürfte wegen der Länge des Referenzzeitraums (Zuwanderung seit 1983) eher von untergeordneter Bedeutung sein.

Für die Revision der Design-Gewichtung wurde der folgende Kompromiß gewählt. Die Design-Gewichtung wurde an die Verteilung der DIW-Schätzung auf Haushaltsebene angepaßt. Dies verändert die Haushaltszahlen in der Gruppe der Aus- und Übersiedler nur unwesentlich. Die Haushalte der sonstigen Zuwanderer-Haushalte werden mit dem Faktor  $1030/716 = 1.438$  hochgewichtet. Andererseits wird die Design-Gewichtung auf Personenebene nicht an die DIW-Schätzung angepaßt. Mit der revidierten Design-Gewichtung erhält man für die Gruppe der über 16jährigen (Tabelle 11).

Unter dieser revidierten Design-Hochrechnung beträgt der Anteil der Gruppe sonstiger Zuwanderer 46,5 vH (Personen über 16 Jahre) gegenüber 41,0 vH bei der Design-Gewichtung. Es wird also nicht ganz der Anteil von 55 vH der DIW-Schätzung erreicht. Da die Hochrechnungsfaktoren auf Haushaltsebene identisch für alle Haushaltsmitglieder sein sollten, kann eine gleichzeitige Anpassung an die DIW-Schätzung auf Personenebene nur erreicht werden, wenn größere Haushalte stärker als kleinere Haushalte gewichtet werden. Dies ist aber bei einer Interpretation der Gewichte als Kehrwert der Ziehungswahrscheinlichkeit

Tabelle 11

**Vergleich der revidierten Design-Hochrechnung mit der DIW-Schätzung für die Gruppe der über 16jährigen  
Schätzwerte in Tausend**

	D-Stichprobe Design-Hochrechnung		DIW-Schätzung
	ursprünglich	revidiert	
Übersiedler	586 (19.3)	630 (17.1)	622 (15.1)
Aussiedler	1 209 (39.7)	1 342 (36.4)	1 232 (29.9)
Übrige Deutsche	103 (3.4)	135 (3.7)	260 (6.3)
Flüchtlinge/Asylbewerber	412 (13.5)	576 (15.6)	680 (16.5)
Sonstige Ausländer	733 (24.1)	1 004 (27.2)	1 326 (32.2)
Summe	3 043 (100.0)	3 687 (100.0)	4 120 (100.0)

lichkeit extrem unplausibel, da man implizit schließen würde, daß größere Haushalte schwieriger als kleinere Haushalte zu erreichen sind.

## 7. Zusammenfassung

Anhand der Zuwanderer-Stichprobe des Sozio-ökonomischen Panels (SOEP) wird exemplarisch gezeigt, wie Design und Gewichtung einer Ergänzungsstichprobe einer laufenden Panelbefragung entwickelt werden und die Integration in das vorliegende Hochrechnungskonzept vorgenommen werden kann. Der dabei verfolgte Ansatz fühlt sich im wesentlichen der klassischen Stichprobentheorie verbunden, die Hochrechnungsfaktoren über den Kehrwert der Auswahlwahrscheinlichkeiten bestimmt. Eine detaillierte Anpassung an amtliche Eckzahlen verbietet sich im Falle der Ziehung von Ergänzungsstichproben für spezifische Bevölkerungsgruppen wie im vorliegenden Fall häufig von selbst, da keine verlässlichen amtlichen Zahlen über den Umfang der jeweiligen Bevölkerungsgruppe vorliegen.

## Anhang

Hochrechnungsvariablen in der SOEP-Datenbank und deren praktische Anwendung.

L P HRFD Hochrechnungsfaktor für *Personen*querschnitte in Welle L (1995) bei separater Auswertung der Zuwanderer-Stichprobe D. Dieser Hochrechnungsfaktor

schließt Haushalte mit Zuwanderern ein, die mit Nicht-Zuwanderern zusammenleben.

L H HRFD Hochrechnungsfaktor für *Haushalts*querschnitte in Welle L (1995) bei separater Auswertung der Zuwanderer-Stichprobe D. Dieser Hochrechnungsfaktor schließt auch Haushalte mit Zuwanderern ein, die mit Nicht-Zuwanderern zusammenleben.

L P HRF Hochrechnungsfaktor auf *Personenebene* in Welle L (1995) bei gemeinsamer Verwendung aller Teilstichproben (A, B, C und D). Bei diesem Hochrechnungsfaktor sind die Haushalte, in denen Zuwanderer und Nicht-Zuwanderer zusammenleben, mit einem Gewicht von Null versehen, da diese Gruppen bereits von den Teilstichproben A, B und C repräsentiert werden.

L H HRF Hochrechnungsfaktor auf *Haushaltsebene* in Welle L (1995) bei gemeinsamer Verwendung aller Teilstichproben (A, B, C und D). Bei diesem Hochrechnungsfaktor sind die Haushalte, in denen Zuwanderer und Nicht-Zuwanderer zusammenleben, mit einem Gewicht von Null versehen, da diese Gruppen bereits von den Teilstichproben A, B und C repräsentiert werden.

Sollen für einen Zeitvergleich nur die Teilstichproben A und B ausgewertet werden, so sind die Hochrechnungsfaktoren LPHRF und LHHRF mit Faktor 1.053 zu multiplizieren. Dieser Faktor repräsentiert die Zuwanderer, die nicht in bestehende Haushalte zugezogen sind. Ab Welle L (1995) wurden die Hochrechnungsfaktoren für die Teilstichproben A und B entsprechend reduziert, da die Zuwanderer explizit in der Teilstichprobe D erfaßt werden.

## Literaturverzeichnis

- Burkhauser, Richard und Gert Wagner* (1994): Proceedings of the First International Conference of SOEP Users, Vierteljahrshefte zur Wirtschaftsforschung, 63 (1).
- Dunn, Thomas A. und Johannes Schwarze* (1997): Proceedings of the Second International Conference of SOEP Users, Vierteljahrshefte zur Wirtschaftsforschung, 66 (1).
- Infratest Sozialforschung* (1994): SOEP '94, Zuwanderer-Befragung (Teilstichprobe D1), Methodenbericht, München.
- Kalton, Graham und Michael Brick* (1995): Weighting Schemes for Household Panel Surveys, Survey Methodology, 21, p. 33-44.
- Koch, Achim, Siegfried Gabler und Michael Braun* (1994): Konzeption und Durchführung der „Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften“ (ALLBUS) 1994, ZUMA-Arbeitsbericht 94/11, Mannheim.
- Krug, Walter, Martin Nourney und Jürgen Schmidt* (1994): Wirtschafts- und Sozialstatistik. Die Gewinnung von Daten, 3. Auflage, Oldenbourg Verlag, München, Wien.
- Pannenberg, Markus, Ulrich Rendtel* (1996): Dokumentation des Sozio-oekonomischen Panel (SOEP): Erhebungsdesign, Fallzahlen und erhebungsbedingte Ausfälle sowie die Schätzung von Ausfallwahrscheinlichkeiten bis Welle 12 (1984 bis 1995), DIW Diskussionspapier Nr. 137, Berlin.
- Pischner, Rainer* (1994): Quer- und Längsschnittgewichtung des Sozio-oekonomischen Panels, in: Gabler et al. (Hrsg.): Gewichtung in der Umfragepraxis.
- Projektgruppe Panel* (1993): Zehn Jahre Sozio-oekonomisches Panel (SOEP). Vierteljahrshefte zur Wirtschaftsforschung, Heft 1/2, S. 27-42.
- Projektgruppe Panel* (1995): Das Sozio-oekonomische Panel (SOEP) im Jahre 1994, Vierteljahrshefte zur Wirtschaftsforschung, Heft 1, S. 7-17.
- Rendtel, Ulrich* (1987): Methodische Konzepte für die Hochrechnung von Panel-Daten. Vierteljahrsheft zur Wirtschaftsforschung, Heft 4, S. 278-290.
- Rendtel, Ulrich* (1995): Panelausfälle und Panelrepräsentativität, Band 6 der Reihe: Sozioökonomische Daten und Analysen für die Bundesrepublik Deutschland, Campus Verlag, Frankfurt am Main, New York.
- Rendtel, Ulrich* (1997): Konvexe Gewichtung von Paneldaten. Erscheint in: Hujer, R.; Rendtel, U.; Wagner, G. (Hrsg.): Sozial- und wirtschaftswissenschaftliche Panel-Studien in Deutschland, Sonderheft zum Allgemeinen Statistischen Archiv, Göttingen 1997.
- Schulz, Erika, Ulrich Rendtel, Jürgen Schupp, Gert Wagner* (1993): Das Zuwanderer-Problem in Wiederholungsbefragungen am Beispiel des Sozio-oekonomischen Panels (SOEP), DIW Diskussionspapier 71, Berlin.
- Schulz, Erika* (1994): Zuwanderungen der letzten zehn Jahre und Abschätzung des Bestandes zum 31.12.93, DIW Diskussionspapier 99, Berlin.
- Schupp, Jürgen und Gert Wagner* (1995): Die Zuwanderer-Stichprobe des Sozioökonomischen Panels (SOEP), Vierteljahrshefte zur Wirtschaftsforschung, Heft 1, S. 16-25.
- Wagner, Gert, Jürgen Schupp und Ulrich Rendtel* (1994): Das Sozio-oekonomische Panel — Methoden der Datenproduktion und -aufbereitung im Längsschnitt, in: R. Hauser et al. (Hrsg.), Mikroanalytische Grundlagen der Gesellschaftspolitik, Bd. II, Berlin, Akademie Verlag, S. 70-111.

## Summary

### The weighting procedure of the immigrant sample of the German Socio-Economic Panel (GSOEP)

*In general, panel surveys face the problem that they cannot cover exogenously caused changes of their target population, e.g. the flux of immigrants or the foundation of new enterprises, occurring after the sampling has been conducted. One solution is to add a 'refreshment' sample to the existing panel survey, which covers the group of interest. Using the German Socio-Economic Panel (GSOEP), the paper describes the design of such a 'refreshment' sample, the weighting procedure employed and the incorporation of the new sample into an existing panel survey.*