

Projektgruppe "Das Sozio-ökonomische Panel" (DIW Berlin)

**Article — Digitized Version**

## Das Sozio-ökonomische Panel für die Bundesrepublik Deutschland nach fünf Wellen

Vierteljahrshefte zur Wirtschaftsforschung

**Provided in Cooperation with:**

German Institute for Economic Research (DIW Berlin)

*Suggested Citation:* Projektgruppe "Das Sozio-ökonomische Panel" (DIW Berlin) (1990) : Das Sozio-ökonomische Panel für die Bundesrepublik Deutschland nach fünf Wellen, Vierteljahrshefte zur Wirtschaftsforschung, ISSN 0340-1707, Duncker & Humblot, Berlin, Vol. 59, Iss. 2/3, pp. 141-151

This Version is available at:

<https://hdl.handle.net/10419/140993>

**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.*

# Das Sozio-ökonomische Panel für die Bundesrepublik Deutschland nach fünf Wellen

von der Projektgruppe „Das Sozio-ökonomische Panel“<sup>\*)</sup>

## Einleitung

Das Sozio-ökonomische Panel<sup>1</sup> ist eine Stichprobenerhebung im Rahmen der Bevölkerungs- und Erwerbsstatistik. Mit dem Sozio-ökonomischen Panel werden seit 1984 durch wiederholte Befragungen regelmäßig repräsentative Mikro-Längsschnittdaten über Personen, Familien und Haushalte in der Bundesrepublik Deutschland bereitgestellt. Für die Ausdehnung der Panelerhebung auf das Gebiet der DDR im Juni 1990 vgl. den Aufsatz von Schupp und Wagner in diesem Heft.

Die Untersuchung legt besonderen Wert auf ein breites Themenspektrum, damit Stabilität und Wandel der Lebenslagen der Menschen nicht nur in einzelnen Bereichen isoliert untersucht, sondern auch die Abhängigkeiten, die zwischen Veränderungen in verschiedenen Lebensbereichen existieren, ermittelt werden können. Die wichtigsten Themenbereiche der Studie betreffen die Veränderungen der Haushaltszusammensetzung, die Erwerbsbeteiligung und die berufliche Mobilität, Einkommensverläufe sowie die Wohnsituation und die regionale Mobilität. Darüber hinaus werden aber z. B. auch Informationen zur Bildung und Weiterbildung, Gesundheit, Zeitverwendung, Zufriedenheit und zu Werteinstellungen erhoben. Für die meisten Themenbereiche wird dabei angestrebt, sowohl Angaben über objektive Lebensbedingungen als auch subjektive Wahrnehmungen zu erhalten.

Der vorliegende Aufsatz gibt zunächst einen knappen Überblick über das Befragungskonzept des Panels (Abschnitt 1). Da dieser Bereich in der Literatur schon dokumentiert ist (vgl. Hanefeld 1987) und erste Ergebnisse vorliegen (vgl. Krupp und Hanefeld 1987, Krupp und Schupp 1988, Hauser 1990, Wagner 1990)<sup>2</sup>, wird er hier nicht vertieft. Ausführlich wird hingegen auf die Entwicklung der Stichprobe (Abschnitt 2) und die Datenaufbereitung und -weitergabe eingegangen (Abschnitt 3). Im folgenden werden die Begriffe „Sozio-ökonomisches Panel“ und „Panel“ synonym gebraucht. Als Akronym wird SOEP benutzt<sup>3</sup>.

Zur Durchführung der Erhebung des SOEP wurde Infratest Sozialforschung, München, beauftragt. Neben der Kooperation bei der endgültigen Gestaltung der Erhe-

bungsinstrumente, erfolgt durch Infratest die Datenerhebung einschließlich der Datenaufbereitung, der Dokumentation bis zum Erstellen des geprüften Datenbandes, welches an das DIW übergeben wird.

---

<sup>\*)</sup> Der Projektgruppe „Das Sozio-ökonomische Panel“ im Deutschen Institut für Wirtschaftsforschung (DIW) gehören Gert Wagner, Joachim Frick, Elke Holst, Peter Krause, Herbert Lahmann, Christiane Paparoditis, Rainer Pischner, Ulrich Rendtel, Jürgen Schupp und Johannes Schwarze an. Der vorliegende Text basiert auch auf Arbeiten aus dem Jahre 1988, als noch Ute Hanefeld, Stefan Jäkel-Hanefeld, Ingo Pfeiffer und James Witte Mitarbeiter im Projekt waren (vgl. auch Fußnote 1).

<sup>1</sup> Das Projekt ‚Das Sozio-ökonomische Panel‘ (SOEP) wurde in der ersten Forschungsphase des Sonderforschungsbereichs 3 (Sfb 3; vgl. dazu Hujer und Wagner 1985) im Rahmen des Teilprojekts B-1 ‚Integrierte Mikrodatenfiles (IMDAF)‘ an der Universität Frankfurt mit Prof. Dr. H.-J. Krupp als Projektleiter vorbereitet. Unter seiner Leitung wurde es seit 1982 als eigenständiges Teilprojekt B-5 des Sfb 3 von der Deutschen Forschungsgemeinschaft (DFG) gefördert. Nach der Berufung Prof. Krupps zum Senator für Finanzen der Freien und Hansestadt Hamburg wurde das Projekt vom Juni 1988 bis zum Juni 1989 von Prof. Dr. W. Zapf, Präsident des Wissenschaftszentrums Berlin, geleitet. Seit dem 1. Juli 1989 ist Dr. Gert Wagner als wissenschaftlicher Leiter des Projekts tätig. Seit 1983 wird es in Kooperation mit dem Deutschen Institut für Wirtschaftsforschung in Berlin durchgeführt, wo Dr. Horst Seidler bis Juni 1989 der zuständige Abteilungsleiter für die Projektgruppe war (vgl. Sonderforschungsbereich 3 1988, S. 7-96). Planmäßig endete 1989 die Anlauffinanzierung der DFG für das SOEP. Für die Weiterförderung erwies sich nach Prüfung mehrerer Optionen eine Projektförderung im Rahmen einer Bund-Länder-Finanzierung, zunächst bis zum Ende des Jahres 1995, als beste Lösung. Die erforderlichen Finanzmittel werden der DFG als Sonderzuwendungen des Bundes und der Länder zur Verfügung gestellt, da sie die fachliche, organisatorische und finanzielle Betreuung weiter übernimmt. Der Finanzierungsschlüssel für Bund und Länder beträgt 50:50; 25 vH des Länderanteils trägt Berlin als Sitzland. Für die Zeit nach 1995 werden rechtzeitig folgende Möglichkeiten geprüft werden: Fortsetzung des Projektes, Beendigung mit abschließender Auswertungsphase oder Fortführung mit neuer Panelkonzeption.

<sup>2</sup> Für einen nahezu vollständigen Nachweis aller auf Basis der Panel-Daten bisher entstandenen Publikationen vgl. DIW (1989, S. 83 ff).

<sup>3</sup> SOEP ist als „SÖP“ gelesen die deutsche Abkürzung; im englischen ergibt sich SOEP aus Socio Economic Panel.

## 1. Das Grundkonzept

### 1.1 Fragebogenkonzeption des Sozio-ökonomischen Panels

Auch für das SOEP gelten die bekannten Regeln des Untersuchungsdesigns von Panel-Studien, daß nämlich bei denselben Personen bestimmte Informationen in bestimmten Zeitabständen (Wellen) wiederholt erfragt werden (vgl. Lazarsfeld et al. 1957). Der Zeitbezug der Fragen des SOEP reicht jedoch bei vielen Indikatoren über den jeweiligen Befragungszeitpunkt hinaus bzw. in die Vergangenheit hinein. So werden beispielsweise Informationen — insbesondere bei objektiven Lebensbedingungen — auch *direkt* durch retrospektive Fragen ermittelt (zu Einzelheiten vgl. Hanefeld 1987, S. 225 ff.). Diese Fragen bilden zusammen mit den reinen zeitpunktbezogenen Einstellungsfragen das Standarderhebungsprogramm zur Ermittlung von Stabilität bzw. intraindividuelle Veränderung. Neben diesem Standardteil gibt es wellenspezifische Zusatzfragen. Diese sind aber kein Instrument, mit dem umfassende Informationen zum Schwerpunktthema gewonnen werden können, sondern in jeder Welle kann nur das Standardprogramm in einem bestimmten Bereich vertieft werden. Ein Sechstel bis ein Fünftel der Gesamtbefragungsdauer ist solchen Schwerpunktfragen gewidmet.

Da im SOEP der variable Erhebungsteil im Umfang sehr begrenzt ist, wurde bei der Auswahl dieser Fragen darauf geachtet, daß diese einer der folgenden panelspezifischen Kategorien zugeordnet werden können (vgl. zur generellen Klassifizierung von Frageinhalten Holm 1982, S. 32 ff):

- *Fragen, die als Hintergrundvariable einen wichtigen Erklärungsbeitrag für Richtung und Stärke des beobachtbaren Wandels liefern, der mit dem Grundprogramm untersucht wird. Für die Themen des SOEPs sind dies meist demographische Fragen, welche in der Regel retrospektiv zu erhebende Inhalte haben.*
- *Fragen zu Tatbeständen, welche sich — wenn überhaupt — in der Regel nur selten, d.h. zwischen mehreren Befragungswellen, ändern und die abhängigen Variablen des Standardprogramms beeinflussen. Die geringe Veränderungsintensität solcher Variablen macht eine jährliche Wiederholung verzichtbar.*
- *Fragen zu gegenwärtigem, vergangenem oder zukünftigem Handeln, die vor dem Hintergrund tatsächlich eintretender Entwicklungen oder bereits erfaßter Veränderungen untersucht werden können. Zwar liegt hier in der Regel eine hohe Veränderungsintensität vor, was eine jährliche Wiederholung rechtfertigen könnte, jedoch verbietet der knappe Rahmen der Befragungsdauer eine jährliche Erhebung.*
- *Einstellungs- und Bewertungsfragen, welche sich in Querschnittstudien methodisch bewährt haben, mit*

*anderen Variablenkomplexen des Längsschnitt-Standardprogramms in einem Zusammenhang stehen, aber ebenfalls wegen der knappen Befragungsdauer nicht in das Standardprogramm einbezogen werden können.*

*Übersicht 1 zeigt die Befragungsschwerpunkte in den Wellen 1 bis 8*

#### Übersicht 1

<i>Welle 1 1984</i>	<i>Erwerbsbiographie seit dem 15. Lebensjahr</i>
<i>Welle 2 1985</i>	<i>Ehe- und Familienbiographie</i>
<i>Welle 3 1986</i>	<i>Soziale Herkunft, Wohnumfeld und Berufseinstieg</i>
<i>Welle 4 1987</i>	<i>Soziale Sicherung, Übergang in den Ruhestand, Kinderbetreuung</i>
<i>Welle 5 1988</i>	<i>Vermögen und Vermögensbildung</i>
<i>Welle 6 1989</i>	<i>Weiterbildung und Qualifizierung im Beruf</i>
<i>Welle 7 1990</i>	<i>Zeitverwendung und -präferenzen</i>
<i>in Vorbereitung:</i>	
<i>Welle 8 1991</i>	<i>Familie und soziale Dienste</i>

Neben den wissenschaftlichen Kriterien, die zur Auswahl und thematischen Festlegung bei der Fragebogenkonstruktion berücksichtigt werden, wird stets versucht, der Grundregel von Panelstudien gleichfalls Rechnung zu tragen: „... je kürzer, interessanter und abwechslungsreicher der Fragebogen gestaltet ist, desto höher wird die Teilnahmebereitschaft von Befragter und Befragtem sein“ (Kaplitza 1982, S. 176).

## 1.2 Die Erhebung

Im Jahre 1984 wurde die erste Datenerhebung bei fast 6 000 Haushalten durchgeführt, in denen alle Personen im Alter von mindestens 16 Jahren durch ausgewählte Interviewer von Infratest Sozialforschung befragt wurden. Neben Haushalten mit einem deutschen Haushaltsvorstand sind auch solche mit einem ausländischen Haushaltsvorstand in der Stichprobe enthalten. Um zumindest über die fünf in der Bundesrepublik Deutschland am stärksten vertretenen Ausländergruppen getrennte Aussagen treffen zu können, wurden sie mit insgesamt 1 400 Haushalten durch einen disproportionalen Stichprobennansatz in der Untersuchung überrepräsentiert.

Aus erhebungstechnischen Gründen wurden zwei Teilgruppen gebildet, für die jeweils getrennt Zufallsstichproben gezogen wurden:

Stichprobe A: Personen in Privathaushalten, deren Haushaltsvorstand nicht die türkische, griechische, jugoslawische, spanische oder italienische Staatsangehörigkeit besitzt;

Stichprobe B: Personen in Privathaushalten, deren Haushaltsvorstand die türkische, griechische, jugoslawi-

sche, spanische oder italienische Staatsangehörigkeit besitzt.

In der Stichprobe A sind somit neben den deutschen auch Haushalte mit einem ausländischen Haushaltsvorstand enthalten, z. B. solche von Polen, Österreichern, Niederländern. Der disproportionale Stichprobenansatz wurde im Rahmen der Hochrechnung berücksichtigt, so daß repräsentative Aussagen über die gesamte Bevölkerung getroffen werden können.

Bei der Studie werden neben dem vom Interviewer bearbeiteten Adressen- und Kontaktprotokoll zwei Erhebungsinstrumente eingesetzt. Der Haushaltsfragebogen enthält insbesondere Fragen zur Wohnung, zum Haushaltseinkommen und zur Vermögenssituation sowie zu den haushaltsbezogenen Transferzahlungen. Er soll vom Haushaltsvorstand beantwortet werden. Der Personenfragebogen, den alle Befragungspersonen erhalten, umfaßt dagegen individuelle Merkmale und subjektive Einschätzungen<sup>4</sup>.

Überwiegend wird die Befragung mit Hilfe von Interviewern durchgeführt, wobei diese ein mündliches Interview führen oder die Befragten die Fragebogen selbst ausfüllen lassen. Nur in Ausnahmefällen finden telefonische oder schriftliche Befragungen statt. Für die Haushalte der Stichprobe B werden zweisprachige Erhebungsinstrumente eingesetzt, bei denen zusätzlich zum Fragetext in der jeweiligen Heimatsprache der deutsche Wortlaut eingedruckt ist. Außerdem haben die Interviewer bei dieser Teilstichprobe die Möglichkeit, sich eine Begleitperson der entsprechenden Nationalität zu suchen, die ihnen bei eventuellen Kontakt- und Sprachschwierigkeiten behilflich sein kann.

### 1.3 Das Weiterverfolgungskonzept

In die der ersten Welle folgenden Erhebungswellen, die im jährlichen Abstand durchgeführt werden, sind alle Personen einbezogen, die an der ersten Befragung teilgenommen haben, unabhängig davon, ob sie mit dem gesamten Haushalt innerhalb der Bundesrepublik Deutschland umziehen oder sich von dem ursprünglichen Haushalt trennen, um alleine oder zusammen mit anderen einen neuen Haushalt zu gründen. Nur Personen und Haushalte, die ins Ausland ziehen, werden nicht weiterverfolgt. In den Panel-Haushalten werden alle 16 Jahre und älteren Personen interviewt, so daß fortlaufend die nachwachsende Generation berücksichtigt wird. Darüber hinaus werden aber auch Personen in die Befragung einbezogen, die im Verlauf der Untersuchung in die ‚Panelhaushalte‘ ziehen; sie wurden jedoch bisher nicht weiter in die Erhebungen einbezogen, wenn sie sich von diesen Haushalten wieder trennen.

Personen, die in einer Welle nicht interviewt werden konnten, werden als sogenannte temporäre Ausfälle (auch „vorläufige Ausfälle“ genannt) weiterverfolgt. Es

handelt sich dabei aber nur um solche Personen, die nicht definitiv erklärt haben, nicht mehr bei der Befragung mitmachen zu wollen. Definitive Verweigerungen werden aus Datenschutzgründen aus der Stichprobe genommen. Ein typischer Grund für einen temporären Ausfall ist hingegen ein Auslandsaufenthalt, der zu einer Nichterreichbarkeit der Zielperson führt.

Der technische Ablauf der Weiterverfolgung ist in *Infra-test Sozialforschung* (1986, S. 8 ff.) dokumentiert. Bisher wurden sechs Befragungswellen vollständig abgeschlossen; hier werden Auswertungen für die ersten fünf Wellen vorgelegt (vgl. dazu auch *Deutsches Institut für Wirtschaftsforschung* 1989, S. 35 ff).

### 1.4 Panelpflege

Zentrales Ziel der Panelpflege ist die Aufrechterhaltung des Stichprobenbestandes. Die Qualität der Panelstichprobe ist wesentlich von der Bereitschaft der ausgewählten Haushalte abhängig, sich regelmäßig befragen zu lassen. Nur bei sehr hohen Ausschöpfungsraten über mehrere Wellen hinweg ist der Bestand der Stichprobe als Basis für repräsentative Aussagen über die Grundgesamtheit gesichert.

Die Panelpflege beim SOEP gliedert sich damit in zwei wichtige Bereiche:

- die Kontaktaufnahme und -pflege, um die befragten Personen für weitere Interviews zu gewinnen (Motivation),
- das Wiederauffinden von Haushalten, die zwischen zwei Befragungen innerhalb der Bundesrepublik Deutschland umziehen (Aktualisierung des Adressenbestandes).

Die Panelpflege begann bereits mit der ersten Befragungswelle. Um bei den Befragten in populärer Form auf die Inhalte der Untersuchung hinzuweisen, wurde anstelle des allgemein nur schwer verständlichen Projektstitels „Sozio-ökonomisches Panel“ für die Datenerhebung die Bezeichnung „Leben in Deutschland“ gewählt. Die direkte Kontaktaufnahme zu den Befragten beginnt mit einem Brief und der Überreichung einer Informationsbrochure. Beide werden für die ausländischen Haushalte der Stichprobe B in die jeweilige Landessprache übersetzt. Während der Feldzeit werden die Befragten dann in der Regel durch den Interviewer betreut. Der Befragung folgen Dankschreiben und als kleines Dankeschön ein

<sup>4</sup> Bei den Personenfragebogen gibt es für besondere Subgruppen zusätzliche Varianten: ausländerspezifische Teile; Nacherhebung der biographischen Schwerpunktthemen der ersten drei Wellen (siehe Übersicht 1) für neu in die Stichprobe kommende Personen (z.B. durch Heirat einer „Stammperson“); Nacherhebung von Veränderungen nach „temporären Ausfällen“, d.h. für Personen, die nach einer Befragungspause in der nächsten Welle wieder interviewt werden konnten (vgl. dazu Abschnitt 1.3 unten).

Los für die Lotterie „Der große Preis“. Darüber hinaus werden Informationen über Ergebnisse der Befragungen mit Hinweis auf die folgende Welle übersandt. Diese Aktivitäten dienen zugleich der Aktualisierung des Adressenbestandes im Falle von Umzügen der Befragungshaushalte. Insbesondere durch den Versand der Lose und Broschüren können die Adressen rechtzeitig vor Feldbeginn noch einmal auf den neuesten Stand gebracht werden.

Die vom Interviewer beim ersten Kontakt überreichte Broschüre erläutert Sinn und Zweck der Untersuchung; es wird zum Beispiel erklärt, daß der Haushalt zufällig ausgewählt wurde und warum seine Teilnahme wichtig ist. Außerdem enthält die Broschüre Informationen zum Datenschutz und die Ankündigung des Lotterie-Loses. Institutionen, Mitarbeiterinnen und Mitarbeiter, die das Projekt durchführen, werden kurz vorgestellt.

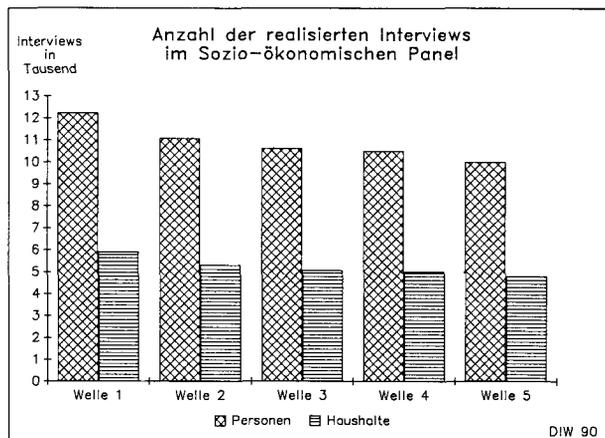
Seit der zweiten Welle werden die Befragten mit einer Informationsbroschüre angesprochen, die vor jeder Welle anhand der neuesten Panel-Auswertungen inhaltlich neu gestaltet wird. Diese Informationsbroschüre wird dann mit einem Schreiben, in dem der Interviewerbesuch angekündigt wird, etwa sechs Wochen vor der Befragung den Haushalten übersandt.

Seit der vierten Welle erhalten alle Panelhaushalte neben dem Los ein kleines Geschenk. Die Interviewer hatten zuvor zum Teil beklagt, daß sie beim Besuch keine Aufmerksamkeit als Dank für die Teilnahme persönlich überreichen konnten. Der Wert des Geschenks wurde mit 5,— bis 10,— DM per Welle und Haushalt niedrig angesetzt, so daß von einer Bezahlung der Teilnehmer nach wie vor nicht gesprochen werden kann.

Im übrigen wird der Motivation der Interviewer erhebliche Bedeutung beigemessen. Die Interviewer stellen in den meisten Fällen den Kontakt zu den Haushalten her und gestalten die Befragungssituation. Von daher hat der Interviewer auch einen erheblichen Einfluß auf die Teilnahmebereitschaft der Befragten. Nach den bisherigen Erfahrungen ist für die Motivation der Interviewer vor allem wichtig, deren Anregungen aufzunehmen und bei der Gestaltung der Befragung die Kritik zu berücksichtigen sowie ihnen gegenüber Anerkennung und Dank für ihre zum Teil recht schwierige Arbeit auszudrücken. Darüberhinaus bilden eine gute Schulung, ausreichende Informationen über das Projekt und den Fragenkatalog, eine übersichtliche Gestaltung der Erhebungsinstrumente und die Information über Forschungsergebnisse<sup>5</sup> die Grundlage für eine erfolgreiche Interviewertätigkeit.

Im Projekt wird die Verantwortung gegenüber den Befragten auch beim Datenschutz sehr ernst genommen. Es wurden spezielle Verträge mit Datennutzern entwickelt (vgl. Abschnitt 3.4 unten), die nur eine genau umrissene empirische Nutzung zulassen. Zu jeder Befragungswelle werden über das gesetzliche Maß hinaus Informationen zum Datenschutz an die befragten Haushalte gegeben.

Schaubild 1



Nur durch Transparenz über die ausschließlich wissenschaftliche Verwendung der Angaben und durch Aufklärung über Vorkehrungen zum Schutz vor Reidentifizierung der Befragten können eventuell auftretende Bedenken ausgeräumt und das Projekt erfolgreich durchgeführt werden.

## 2. Die Panelpopulation

Der Ausgangsbestand des SOEPs ist durch die Nettostichprobe der in Welle 1 realisierten Haushalts-Interviews festgelegt. Im Jahre 1984 beteiligten sich 5 921 Haushalte an der Untersuchung, in denen 16 205 Haushaltsmitglieder erfaßt wurden. 12 290 Haushaltsmitglieder gehörten zu den Befragungspersonen mit einem Mindestalter von 16 Jahren; mit 12 245 Befragungspersonen konnte ein Interview realisiert werden.

### 2.1 Ausgewählte Ergebnisse zur Stichprobenentwicklung

Aus Schaubild 1 geht hervor, wie sich die Fallzahlen der realisierten Personen- und Haushaltsinterviews bis zur Welle 5 entwickelt haben. Insgesamt zeigt sich eine deutliche Stabilisierung der Ausfallraten und der Fall-

<sup>5</sup> Die Bedeutung der Interviewermotivation wurde im Volkszählungsjahr 1987 deutlich. Obwohl die Panelfeldarbeit im Vorfeld des Zensus auf dem Höhepunkt der öffentlichen Diskussion durchgeführt werden mußte, war die Ausschöpfung sehr hoch (danach stellte sich heraus, daß mit 95,3 vH sogar der vorläufige Gipfel der Ausschöpfung im Jahre 1987 erreicht wurde). Der Interviewer-Einsatzleiter von Infratest führte dies u.a. darauf zurück, daß im Vorfeld der Volkszählung kaum normale Querschnitts-Befragungen durchgeführt wurden, da potentielle Auftragnehmer negative Effekte erwarteten. Dadurch waren die freiberuflichen Interviewer tendenziell unterbeschäftigt, und sie konzentrierten sich — im Interesse ihres Einkommens — noch stärker auf den Erfolg der Panelinterviews.

Tabelle 1

**Ausfallraten aufgrund von Nichterreichbarkeit  
in vH**

	Welle 2		Welle 3		Welle 4		Welle 5	
	N	vH	N	vH	N	vH	N	vH
Insgesamt	6051	1,9	5814	1,4	5465	1,2	5342	1,1
davon:								
HH ohne Adressenänderung	5413	0,8	5039	0,4	4808	0,1	4683	0,1
verzogene Einpersonen-Haushalte	119	21,0	180	13,9	142	11,2	143	10,5
N = Anzahl der Fälle insgesamt (alte Haushalte + neue Haushalte – durch Tod aufgelöste Einpersonenhaushalte – ins Ausland verzogene Haushalte) vH = Prozentsatz der Haushalte ohne Kontaktaufnahme (Basis N) Datenbasis: Das Sozio-ökonomische Panel, Welle 1–5.								

zahlen des Panels. In Welle 5 wurde mit 4814 Haushalten und 10 023 Befragungspersonen ein Interview realisiert. Dies entspricht 81,3 vH bzw. 81,8 vH der Ausgangsgröße von Welle 1. Deutlich wird auch, daß die Entwicklung des Haushalts- und des Personenbestandes annähernd gleichförmig verläuft.

Die Stichprobenentwicklung im SOEP resultiert einmal aus demographischem Wandel (Abgänge durch Tod oder Umzug ins Ausland, Zugänge durch Erreichen des Befragungsalters oder Entstehen neuer Haushalte). Zum anderen wird der Stichprobenbestand wesentlich durch die Feldarbeit bestimmt. Ihr Erfolg wird im folgenden unter zwei Aspekten beurteilt: a) Gelingt es, mit den Haushalten aus der vorhergehenden Befragungswelle wieder Kontakt aufzunehmen? b) Erhält man von den Haushalten mit Kontaktaufnahme ein Interview? (vgl. auch Rendtel 1988b).

Tabelle 1 zeigt, daß die Ausfälle aufgrund von Nicht-Erreichbarkeit insgesamt eine untergeordnete Rolle spielen. Zudem konnte diese Ausfallquote von 1,9 vH in Welle 2 auf 1,1 vH in Welle 5 gesenkt werden. Allerdings ist die Kontaktaufnahme mit den Haushalten der vorhergehenden Befragungswelle unterschiedlich schwer:

In der Gruppe der fortgezogenen Einpersonen-Haushalte ist die Ausfallrate wesentlich höher als bei Haus-

halten ohne Adressenänderung. Dies belegen in Tabelle 1 die deutlich verschiedenen Ausfallraten in den jeweiligen Haushaltsgruppen. In beiden Gruppen konnte jedoch die Feldarbeit erheblich verbessert werden.

Der Erfolg der nächsten Stufe der Feldarbeit wird daran gemessen, ob es gelingt, von den wieder erreichten Haushalten ein Interview zu erhalten. Hierbei werden die Haushalte in drei Untergruppen eingeteilt: Teilnehmer des Vorjahres, vorläufige Ausfälle im Vorjahr und die Gruppe der hinzugekommenen Haushalte. Die in Tabelle 2 ausgewiesenen Ausfallquoten zeigen den unterschiedlichen Erfolg der Feldarbeit bei den einzelnen Teilgruppen. Erwartungsgemäß wird das beste Ergebnis bei den Haushalten erzielt, die sich auch im Vorjahr am Panel beteiligten. Erfreulich ist hierbei, daß der Erfolg der Feldarbeit wesentlich gesteigert werden konnte.

Schwieriger zu befragen sind dagegen neue Haushalte, die häufig von jungen Personen nach der Trennung von den Eltern gegründet werden. Doch auch hier konnte die Ausfallrate von Welle 2 auf Welle 4 beträchtlich gesenkt werden. In Welle 5 steigt hier die Ausfallrate wieder etwas. Der negative Effekt des Befragungsschwerpunktes „Vermögen und Vermögensbildung“ ist in der Gruppe der neuen Haushalte im Vergleich zu den Mitgliedern wiederholt befragter Haushalte wesentlich größer gewesen.

Tabelle 2

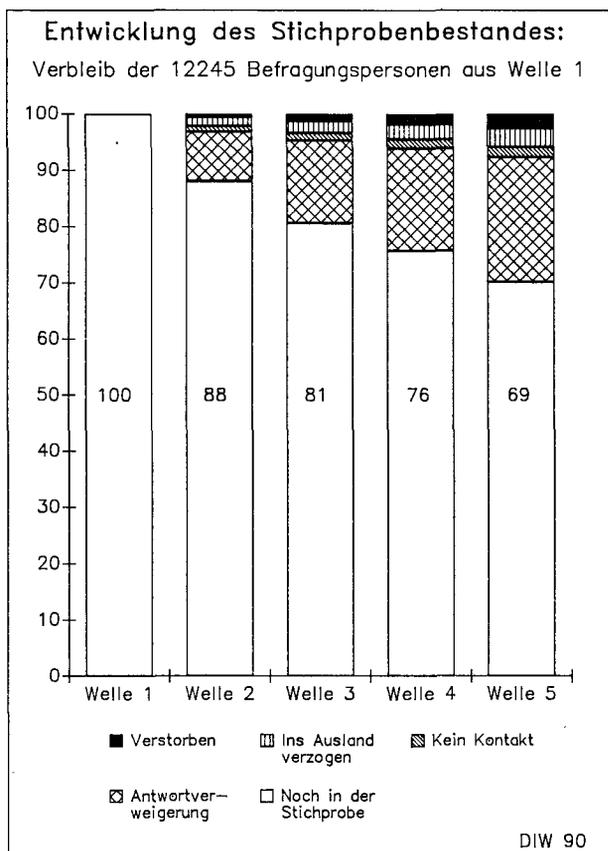
**Ausfallraten der errichteten Haushalte nach der Teilnahme im Vorjahr  
in vH**

	Welle 2		Welle 3		Welle 4		Welle 5	
	N	vH	N	vH	N	vH	N	vH
Teilnehmer des Vorjahrs	5742	9,7	5235	8,3	5019	4,7	4937	6,4
Neue Haushalte	195	29,2	238	21,4	182	17,6	194	23,7
Vorläufige Ausfälle des Vorjahres	—	—	259	59,5	197	52,8	154	71,4
N = Anzahl der Fälle insgesamt (Haushalte mit Kontaktaufnahme) vH = Prozentsatz der Haushalte ohne realisiertes Interview (Basis N) Datenbasis: Das Sozio-ökonomische Panel, Welle 1 bis 5.								

Eine erneute Befragung von (vorläufigen) Ausfällen ist bei anderen Panel-Erhebungen nicht üblich. Durch den systematischen Versuch, Personen, die in einer Welle nicht an der Befragung teilnahmen, wieder zum Mitmachen beim SOEP zu bewegen, konnten immerhin 40 vH der vorläufigen Ausfälle für die Befragung wieder gewonnen werden. Diese Haushalte wären andernfalls endgültig aus der Untersuchung ausgeschieden. Pro Welle wird dadurch eine Reduzierung der Stichprobe um rund 100 Haushalte vermieden. Durch die ‚Nacherhebung von Längsschnittmerkmalen bei lückenhaften Verläufen‘ wird ein Teil der in der Ausfallwelle erfragten Informationen zu Personen nachträglich erfaßt, so daß sie in viele Längsschnittauswertungen einbezogen werden können.

Für Längsschnittauswertungen ist weniger die in Schaubild 1 ausgewiesene Entwicklung des Querschnittbestandes des Panels von Belang als vielmehr der Bestand an Personen, die während der gesamten Laufzeit des Panels im Stichprobenbestand verbleiben. Naturgemäß nimmt dieser Bestand durch demographische Bewegungen (Tod, Fortzug ins Ausland) und erhebungsbedingte Ausfälle (kein Kontakt, Antwortverweigerung) von Welle zu Welle ab. In Schaubild 2 wird die Entwicklung des Bestands der 12 245 Befragungspersonen aus Welle 1 dargestellt. Für 69 vH des Ausgangsbestandes stehen Längsschnittinformationen über die ersten fünf Befragungswellen zur Verfügung. Bis Welle 5 belaufen

Schaubild 2



sich die Verluste durch demographische Bewegungen auf 5,8 vH des Ausgangsbestandes. Die Verluste durch erhebungsbedingte Ausfälle sind 4,3 mal so groß (24,8 vH des Ausgangsbestandes) wie die Verluste durch demographische Veränderungen.

Die Beschränkung auf Personen, die über die gesamte Dauer des Panels im Befragungsbestand verblieben sind, ist sicher die restriktivste Auswahl für eine Längsschnittanalyse. Bei vielen Untersuchungen ist die Analysegesamtheit jedoch erheblich höher. Beispielsweise können alle gestorbenen oder ins Ausland verzogenen Befragungspersonen bei Analysen zur Mortalität und zum Mobilitätsverhalten in die Untersuchungen mit einbezogen werden. Hinzu kommen Gewinne für den Längsschnittbestand, die dadurch entstehen, daß Kinder von Stammpersonen das Befragungsalter von 16 Jahren erreichen oder daß neue Personen in einen befragten Haushalt ziehen.

Der Befragungsbestand hat sich damit kumulativ bis Welle 5 um 1 157 Jugendliche, die 1984 das Befragungsalter noch nicht erreicht hatten, und um 719 zugezogene Befragungspersonen vergrößert. Je nach Analysethema kann dieser Personenkreis ganz oder teilweise in die Längsschnittauswertung einbezogen werden. Dies wird insbesondere dadurch erleichtert, daß bei zugezogenen Befragungspersonen die biographischen Angaben nach erhoben werden. Eine Obergrenze des Analysebestands ist durch die Anzahl aller Personen gegeben, für die mindestens ein Interview vorliegt. Dies sind bis zur fünften Welle 14 121 Befragungspersonen (= 115 vH des Ausgangsbestands der ersten Welle).

### 3. Datenaufbereitung und -weitergabe

#### 3.1 Datenbank

##### 3.1.1 Grundsätze

Für die Aufbereitung, Haltung, Pflege und Weitergabe galt es Datenstrukturen aufzubauen, die dem komplexen, sich ständig vergrößernden Datenbestand des Panels gerecht werden und den Datenschutz unterstützen (vgl. für das Datenschutzkonzept Wagner und Lutterbeck 1985). Voraussetzung hierfür ist ein Datenbanksystem, das unterschiedlich aufgebaute Datensätze zu einem Gesamtbestand zusammenfassen kann. Dateiteile, die in mehreren Dateien auftreten, sowie Informationen, die für unterschiedliche Individuen gleichermaßen existieren — im Panel sind dies zum Beispiel die Haushaltsinformationen, die für jede Person im Haushalt zumindest für eine Welle gleich sind —, sollten nur einmal abgespeichert werden. Dadurch wird eine Datenbank zu einer systematischen Ordnung von Dateien, die — gemeinsamen Hauptordnungs-begriffen folgend — zentrale Zugriffe, Suchoperationen und Veränderungen zuläßt.

Vor allem unter dem Gesichtspunkt der Datenweitergabe entschloß sich die Projektgruppe, eine rechnerunabhängige Datenbanksoftware zu nutzen (vgl. Schupp 1985). Das Programmpaket SIR/DBMS wird in vielen Punkten den oben genannten Anforderungen an ein Datenbanksystem gerecht. Das System ist stark anwendungsorientiert und für empirische Sozial- und Wirtschaftsforscher relativ leicht und schnell zu erlernen. Es ist speziell für statistische Zwecke konzipiert, was sich vor allem an seinen komfortablen Schnittstellen zu den gängigen Statistik-Softwarepaketen SPSS, SAS und BMDP zeigt. Das Datenbanksystem ist in der Bundesrepublik weit verbreitet, zwei Drittel der derzeit 60 externen Datennutzer des Panels verwenden eine Version der Panel-Datenbank. Die anderen Datennutzer erhalten Rohdaten einschließlich eines Dictionary-Files mit Variablenliste, Variablenbeschreibung und Formatspezifikationen.

### 3.1.2 Die SIR-Paneldatenbank

Abgesehen von der erwähnten Speicherplatzersparnis gibt es weitere Anforderungen, die an ein System zur Haltung der Paneldaten zu stellen sind. Die Paneldaten werden von einer wachsenden Zahl sehr unterschiedlicher Datennutzer bearbeitet und ausgewertet. Diesem heterogenen Nutzerkreis gilt es durch einfache und übersichtliche, für alle Auswertungswünsche offene Datenstrukturen Rechnung zu tragen, die die Portabilität der Daten erleichtern. Bei den möglichen Analyseanforderungen sind vor allem drei Gruppen zu nennen:

- reine Querschnittanalysen,
- Auswertungen im Längsschnitt und
- eine Mischform von beidem, in der den Querschnittinformationen einer Welle Hintergrundinformationen aus anderen Wellen zugespielt werden.

Der Datenbestand, der jährlich in die Panel-Datenbank aufgenommen wird, gliedert sich in zwei Teile, den Brutto- und den Nettobestand. Im Nettobestand sind nur Personen enthalten, die tatsächlich in der jeweiligen Welle ein auswertbares Interview gewährt haben. Der Bruttobestand umfaßt insbesondere auch erfolglose Interviewversuche; leitet sich aus dem Adressenprotokoll sowie den Ergebnissen der Feldarbeit ab und bildet damit die sogenannte Paneldatei, die in einem besonderen Codebuch dokumentiert ist (vgl. Infratest Sozialforschung 1987). Die Paneldatei dient Infratest Sozialforschung als Instrument für die Weiterverfolgung und hilft den Anwendern bei Auswertungen zur Verknüpfung der Informationen über mehrere Panelwellen und über alle Haushaltsmitglieder.

Basis für den Bruttobestand sind alle Haushalte und Personen, die mit dem Panel in der aktuellen sowie letzten Welle zu tun hatten, einschließlich der noch nicht zu befragenden Kinder. Komplette Haushaltsausfälle, die

in der vorhergehenden Welle aufgetreten sind, sind indes im aktuellen Bruttobestand nicht vorhanden, während für Personenausfälle eine „Merkvariable“ im Haushaltsrecord enthalten ist.

Im Gegensatz zum Bruttobestand sind im Nettobestand lediglich die Informationen der realisierten Interviews der jeweiligen Befragungswelle abgespeichert. Der Nettodatensatz gliedert sich ebenfalls in einen Haushaltsbestand, der alle Angaben aus dem Haushaltsfragebogen enthält sowie einen Personendatensatz mit den Informationen des Personenfragebogens. Der Datenbestand des Panels setzt sich also aus verschiedenen Dateien mit unterschiedlicher Anzahl von Sätzen zusammen.

Der Datennutzer soll im Idealfall allein mit Hilfe der Fragebögen sowie einiger Zusatzinformationen in die Lage versetzt werden, mit den Paneldaten zu arbeiten. Die aufwendigen Vorarbeiten vor allem am Data-Dictionary, das die Verbindung zwischen dem Bspaltungsplan und den Fragebogen herstellt, zahlen sich später bei Analysen außerordentlich aus. Gerade der Umstand, daß alle Variablen und alle Variablenwerte — soweit sinnvoll — mit Beschreibungen versehen sind und diese auch für Auswertungen zum Beispiel mit dem Statistik-Programmpaket SPSS zur Verfügung stehen, ist für den Anwender sehr hilfreich.

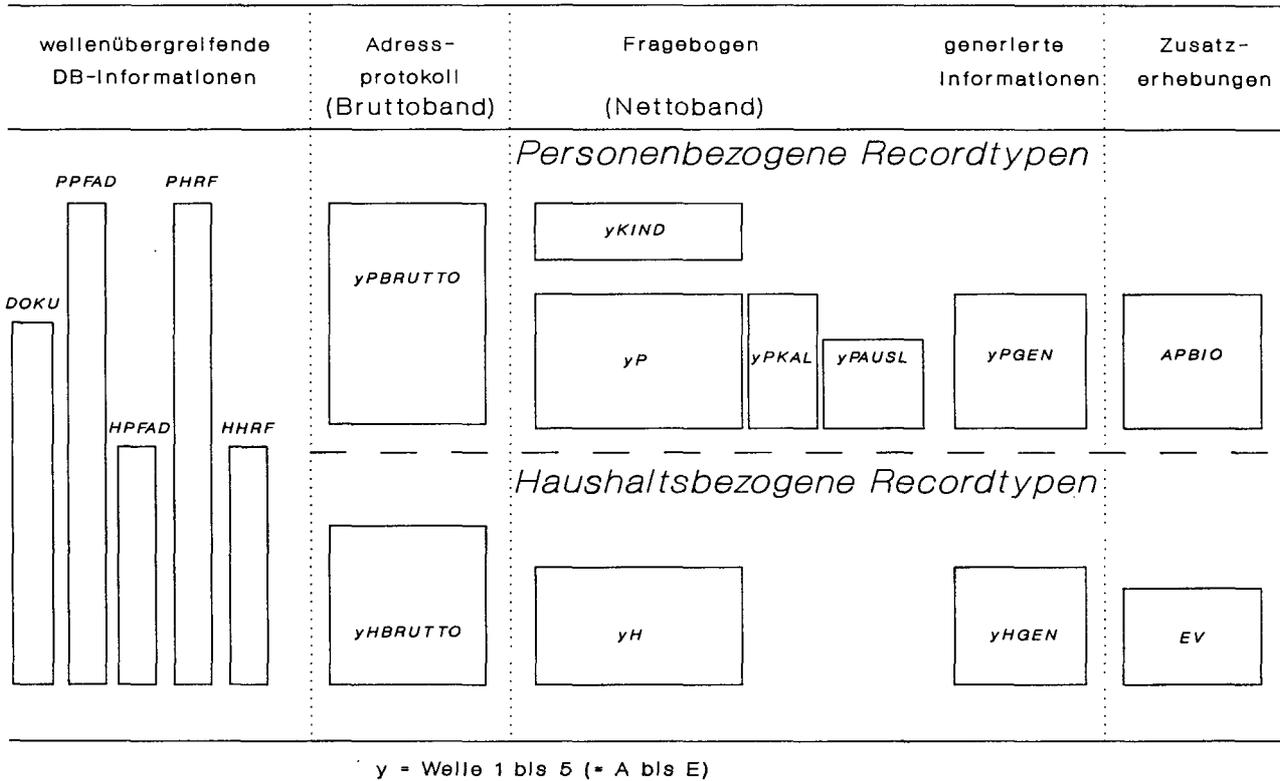
Die Namen der in der Datenbank befindlichen Dateien, der sogenannten Recordtypen, sowie die der Variablen sind so gewählt, daß der jeweilige Nutzer leicht folgende Merkmale erkennen kann:

- das Jahr der Befragung,
- den Unterschied zwischen Haushalts-, Personen- und Kinderinformationen,
- die laufende Frage- und Itemnummer im Erhebungsinstrument.

Die Forderung nach Benutzerfreundlichkeit und Effektivität steht auch beim Zugriff im Längsschnitt im Vordergrund. Es wurde eine Struktur entwickelt, die sich das Weiterverfolgungskonzept des Panels zunutze macht. Aufbauend auf einem Stammbaumkonzept wurden folgende Identifikationsmerkmale als Datenbankschlüssel installiert:

- Haushaltsnummer der ersten Welle. Sie ist das oberste Sortierkriterium, der sogenannte Primärschlüssel der Datenbank (unveränderliche Case-ID).
- Haushaltsnummer der laufenden Welle. Sie ist hierarchisch unter die Haushaltsnummer der ersten Welle gelegt und bildet somit den ersten sogenannten Sekundärschlüssel für die Verknüpfung der verschiedenen Recordtypen. Für alte Haushalte ist diese Nummer identisch mit der Nummer des Ursprungshaushalts. Neue Haushalte bekommen neue Nummern zugewiesen, sind aber datenbanktechnisch wei-

Recordtypen der Panel-SIR-Datenbank



DIW 1990.

terhin Bestandteil des Ursprungshaushalts, aus dem sie sich abgespalten haben.

- Feststehende Personennummer (Pers-ID). Dies ist die Nummer, die einer Person zugewiesen wird, sobald sie erstmals im Panel auftaucht. Sie ist hierarchisch unter den Haushaltsnummern als zweiter Sekundärschlüssel abgelegt. Diese Nummer behält eine Person auch dann bei, wenn sie in einen anderen Haushalt wechselt.

Anhand dieser Schlüssel lassen sich jede Entwicklung, die ein Haushalt vollzieht, sowie jede Bewegung von Personen zwischen unterschiedlichen Haushalten nachvollziehen. Über die Wellen hinweg spielen sich alle Veränderungen datenbankintern innerhalb eines Falles — des Ursprungshaushalts — ab und stehen somit beim Zugriff auf diesen Fall direkt zur Verfügung. Alle weiteren Informationen sind den Schlüsseln untergeordnet und können mit deren Hilfe beliebig verknüpft werden.

Anhand des Schaubildes 3, das die derzeitige Struktur des Datenbankaufbaus schematisch wiedergibt, wird deutlich, daß jede Panelwelle in einen logischen Block von Recordtypen aufgespalten wird, der zwar in sich abgeschlossen ist, über die Datenbankschlüssel aber zusammengeführt werden kann.

### 3.2 Datenprüfung und Fehlerdokumentation

Die Prüfung und Bereinigung der Daten des SOEPs ist ein mehrstufiger Prozeß, der in Zusammenarbeit mit Infratest durchgeführt wird (vgl. auch Infratest Sozialforschung 1985, S. 40 ff.). Erste manuelle Prüfungen werden von der Infratest ‚Editinggruppe‘ vorgenommen. Dieses sogenannte Editing hat zum Ziel, vor allem solche Fehler zu beseitigen, die nach der Datenerfassung auf EDV von maschinellen Prüfprogrammen nicht oder nur schwer als Fehler erkennbar sind (z.B. unzulässige Mehrfachnennungen oder Kommastellen bei Zahlenwerten). Anschließend wird eine EDV-gestützte Basisprüfung durchgeführt, mit der die Haushaltszusammensetzung und die richtige Zuordnung aller Fragebögen, die zu einem Haushalt gehören, überprüft werden<sup>6</sup>. Danach werden die Variablen der Haushalts- und Personendatensätze im Hinblick auf die Richtigkeit der Filterführung und des Wertebereiches geprüft sowie Konsistenz- und Plau-

<sup>6</sup> Da neu in einen Haushalt gezogene Personen mit einem anderen Fragebogen interviewt werden als die anderen Haushaltsmitglieder und zudem die Sprachfassungen der Personenfragebögen eines Haushaltes unterschiedlich sein können, ist die Prüfung der korrekten Zusammenführung der Fragebögen nicht überflüssig.

sibilitätschecks durchgeführt. Bei dieser Stufe der Datenprüfung werden sechs Fehlertypen identifiziert:

- 1) Filterfehler,
- 2) unzulässige Mehrfachnennung,
- 3) unzulässiger Code,
- 4) numerischer Wert außerhalb eines plausiblen Wertebereichs,
- 5) Aufsummierung mehrerer numerischer Werte ergibt unmöglichen oder unplausiblen Wert,
- 6) Angabe steht im Widerspruch zu anderen Angaben im Fragebogen.

Die Fehlertypen 1 bis 3 sind rein formale Fehler und können meistens unproblematisch bereinigt werden. Die Fehlertypen 4 bis 6 betreffen die inhaltliche Plausibilität.

Wenn nach Heranziehung aller relevanten Informationen im Fragebogen eine Aufklärung eindeutig oder mit hoher Sicherheit möglich ist, wird eine Datenkorrektur vorgenommen. In allen anderen Fällen wird der unplausible Wert mit einem ‚Fehlervermerk‘ versehen, wobei zwischen vom Prüfprogramm monierten, jedoch möglichen und tatsächlich unplausiblen Angaben unterschieden wird. Außerdem wird von Welle 2 an von Infratest eine begrenzte Längsschnittprüfung der Daten durchgeführt. Diese bezieht sich auf die Variablen Geburtsjahr, Geschlecht, Familienstand und Nationalität.

Bei den Arbeiten mit den Daten hat sich herausgestellt, daß es sinnvoll ist, zusätzlich zu den Standardprüfungen gezielt weitergehende Prüfungen vorzunehmen. Vorrangig werden dabei die Variablen aus dem Demographiebereich ergänzenden Prüfungen unterzogen. Die Fehler bzw. Unplausibilitäten werden Infratest gemeldet, und dort wird versucht, anhand der Originalunterlagen die jeweilige Unstimmigkeit zu klären. Fälle, die sich nicht eindeutig bereinigen lassen, werden in einem separaten Datensatz dokumentiert, die übrigen werden im Datenbestand korrigiert. Auf diese Weise wurden bereits über 1 000 einzelne Angaben der ersten zwei Panelwellen berichtigt oder ergänzt.

Während die Standardprüfungen mit der Lieferung der Daten an das DIW so gut wie abgeschlossen sind, ist ein Ende der sich daran anschließenden weitergehenden Datenprüfungen nicht eindeutig determiniert. Dies liegt grundsätzlich daran, daß im Rahmen von laufenden Forschungsarbeiten immer wieder neue Variablenkombinationen gebildet und auf deren Grundlage neuerliche Unstimmigkeiten gefunden werden.

Entsprechend der Mehrstufigkeit der Datenprüfungen findet die Dokumentation dieser Prüfungen auch auf mehreren Ebenen statt. Alle Datenprüfungen und -bereinigungen bei Infratest werden dort schriftlich erfaßt und nach Abschluß der Arbeiten dem DIW übermittelt. Die Fehlervermerke werden in eine spezielle Datenbank eingegeben. Zusätzlich zu dieser Form der Fehlerdokumentation ist ein erweiterter Dokumentationsrecordtyp ent-

wickelt worden, mit dessen Hilfe sowohl Informationen über Unplausibilitäten, die bei Infratest gefunden wurden, als auch solche, die im Laufe der weitergehenden Datenprüfung entdeckt werden, in die SIR-Datenbank aufgenommen werden (vgl. in Schaubild 3 den Recordtype DOKU). Der Dokumentations-Record hat zwei Aufgaben zu erfüllen:

- es soll eine einfache Identifizierung unplausibler bzw. fehlerhafter Fälle ermöglicht werden;
- es soll die Möglichkeit gegeben werden, Korrekturwerte anstelle der unplausiblen Werte in etwaige Analysen aufzunehmen.

### 3.3 Hochrechnung und Repräsentativitätsprüfung

Die Hochrechnung von Mikrodaten in Längsschnitten ermöglicht wichtige Auswertungen des SOEPs, weil anderweitig kaum Datensätze existieren, aus denen sich derartige Längsschnittinformationen gewinnen lassen. Die Kenntnis der Anzahl von Personen mit einem bestimmten Verhalten im Zeitverlauf ist vielfach per se ein Untersuchungsinteresse.

Bei anderen Auswertungen wird lediglich der Prozentsatz von interessierenden Längsschnittereignissen innerhalb der Längsschnittgrundgesamtheit geschätzt. Für derartige Auswertungen bedarf es nicht unbedingt der Kenntnis eines Hochrechnungsfaktors. Allerdings ist die Panel-Stichprobe disproportional angelegt, so daß für gemeinsame Auswertungen der Teilstichproben A und B auf jeden Fall eine Designgewichtung für die Schätzung von Prozentsätzen durchgeführt werden muß (vgl. auch Galler 1987). Auch hat die kumulierte Panelmortalität mittlerweile in kleinen Subgruppen ein Ausmaß erreicht, das es auch bei der Schätzung von Prozentzahlen ratsam erscheinen läßt, auf Hochrechnungsfaktoren zurückzugreifen, die den gemeinsamen Effekt von Stichprobendesign und Panelmortalität berücksichtigen. Für Querschnittauswertungen korrigieren die Hochrechnungsfaktoren zudem die durch das Weiterverfolgungskonzept des Panels implizierte Überrepräsentation von Haushalten mit Zuzügen (vgl. Rendtel 1987).

Gerade bei der Hochrechnung von Längsschnitten ist es wichtig, die Genauigkeit dieser Ergebnisse beurteilen zu können. Anders als bei Querschnitten liegen in diesen Fällen keine Vergleichsdaten aus dem Mikrozensus oder anderen Erhebungen vor, anhand derer sich eine Abschätzung der Fehlergröße vornehmen ließe. Genauigkeitsschätzungen können über eine Bestimmung der Varianz der Hochrechnungsergebnisse vorgenommen werden. Diese Schätzung der Varianz kann über drei Ansätze geschehen:

- Bestimmung der gemeinsamen Auswahlwahrscheinlichkeit von je zwei Stichprobenelementen (vgl. Raj 1968, S. 53).

- Zufällige Unterteilung der Panelstichprobe in Unterstichproben (vgl. Wolter 1985, Kapitel 2, 3, 4).
- Bestimmung von generalisierten Varianzfunktionen (vgl. Wolter 1985, Kapitel 5).

Es existieren erste empirische Erfahrungen mit dem ersten Ansatz (vgl. Rendtel 1988a). Im Rahmen der Umsetzung des zweiten Ansatzes wurde mittlerweile eine Unterteilung des Panels in acht Teilpanels in der SIR-Datenbank installiert. Über diese Teilpanels erhält man acht Schätzwerte für den gesuchten Populationswert. Anhand der Spannweite dieser Werte ist eine unmittelbare Schätzung für die Varianz der Hochrechnungsergebnisse möglich (vgl. Sonderforschungsbereich 3 1989, Kapitel K).

### 3.4 Datenweitergabe und Datenschutz

Die Daten des SOEPs werden für wissenschaftliche Zwecke Universitätsinstituten und ähnlichen Einrichtungen kostenlos zur Verfügung gestellt. Das Interesse an den Daten hat ständig zugenommen. Bis zum Dezember 1989 wurden insgesamt 60 Datenweitergabeverträge geschlossen, wobei hinter einem Vertrag häufig mehrere Wissenschaftler stehen, die mit den Daten arbeiten<sup>7</sup>.

Die externen Datennutzer müssen aus rechtlichen Gründen einen schriftlichen Antrag auf Datenweitergabe stellen, in dem sie ihren Analysezweck benennen. Die Prüfung der Datenschutzmaßnahmen bei den externen Nutzern nimmt die Datenschutzbeauftragte des DIW vor. Insbesondere bei einer Datenweitergabe in das Ausland können sich deswegen Verzögerungen ergeben. Bisher mußte aber noch kein Auftrag abgelehnt werden. Zudem werden die externen Datennutzer auch im Hinblick auf den Datenschutz unterstützt. Hier geht es hauptsächlich darum, auf datenbankinterne Sicherungsmöglichkeiten wie auch auf zusätzliche Vorkehrungen auf der Betriebssystemebene hinzuweisen und die in diesem Bereich gesammelten Erfahrungen des Panelprojektes weiterzugeben. Nach den erforderlichen Prüfungen wird ein Datenweitergabevertrag geschlossen, der u. a. datenschutzrechtliche Regelungen enthält.

Die eigentliche Datenweitergabe geschieht dann mittels SIR-Exportfiles oder durch Übermittlung von Rohdaten. Die Größe des Datensatzes führt naturgemäß zu einer immer schwerer handhabbaren Vielzahl an Variablen, die nur zu einem Teil vom Nutzer für die individuellen Analysen gebraucht werden. Akute Speicherplatzprobleme müssen hier durch geeignete Selektionen gelöst werden.

Eine Möglichkeit zur Reduzierung von Speicherplatz besteht in der Auswahl einer geeigneten Stichprobe aus dem Datenbestand. Die in Programmpaketen standardmäßig angebotene Möglichkeit zur Ziehung eines Samples würde aufgrund des Stichprobendesigns im Panel zu gravierenden Informationsverlusten führen. Statt dessen kann hierzu auf eine neu gebildete Variable zugegriffen

werden, die die gesamte Panelstichprobe nach dem Zufallsprinzip in acht Teilpanels zerlegt; dadurch werden Verbindungen zwischen Haushalten (d.h. Stammhaushalte und abgespaltene Haushalte) nicht zerrissen.

### 3.5 Dokumentation und Nutzerschulung

Zentrales Instrument der Dokumentation ist das Benutzerhandbuch (vgl. Sonderforschungsbereich 3 1989). Es enthält im wesentlichen Hinweise und Erläuterungen für die praktische Arbeit mit den Daten. Das Handbuch ist eine Loseblatt-Sammlung; durch Nachlieferungen wird es jeweils um die Informationen über die weiteren Erhebungswellen und Modifikationen des Datenbestandes ergänzt. Es wird gegenwärtig geprüft, inwieweit — zunächst Teile — der Informationen des Handbuches als eigene Datenbank zur Verfügung gestellt werden können. Ein mögliches Konzept stellt das auf dBASE-Basis installierte Programmpaket „FRAVIS“ dar (vgl. Gräf 1988), das Variablen nach Themenbereichen zusammengefaßt und deren Dokumentation samt Grundauszählung über alle Wellen abrufbar bereit hält.

Es wird jedoch immer wieder deutlich, daß der Datensatz — selbst für versierte Empiriker — sehr komplex ist. Dies liegt einmal daran, daß viele Datennutzer bisher entweder nur Haushalts- oder Personendatensätze analysiert haben, nicht aber — wie beim Panel möglich — gleichzeitig Haushalts- und Personeninformationen. Zum anderen wird diese Datenstruktur im Längsschnitt noch komplizierter, weil in Abhängigkeit vom jeweiligen Analyseziel Haushalts- und Personenmerkmale unterschiedlich verknüpft werden müssen.

Regelmäßige Schulungen sind notwendig, wenn vermieden werden soll, daß einzelne Datennutzer resignieren, weil ihnen die Einarbeitung in den Datensatz zu kompliziert und zeitaufwendig erscheint. Nachdem zuletzt im Frühjahr 1990 für externe Datennutzer ein Workshop am DIW durchgeführt worden war, ist geplant, derartige Schulungen regelmäßig durchzuführen. Darüber hinaus beteiligt sich die Panel-Projektgruppe seit 1987 auch an den entsprechenden Veranstaltungen der ZUMA<sup>8</sup>-Arbeitsgruppe Panelforschung.

Zusätzlich wurde ein beispielorientiertes Benutzer-Tutorial entwickelt, das als Ergänzung zum deskriptiven Benutzerhandbuch gedacht ist. Es baut auf den von der Panelgruppe am DIW veranstalteten Workshops für externe Datennutzer auf und wird künftig als Schulungsmaterial dienen (vgl. Frick 1990).

<sup>7</sup> Hervorgehoben sei, daß auch im Ausland reges Interesse an den Paneldaten besteht. So wurde vom Sfb 3-Teilprojekt C-1 „Soziale Sicherung und personelle Einkommensverteilung“ ein Teildatensatz für die Luxemburg Income Study (LIS) erstellt, ein Projekt, bei dem Einkommensdaten aus verschiedenen Ländern für internationale Vergleiche aufbereitet werden.

<sup>8</sup> Zentrum für Umfragen, Methoden und Analysen, Mannheim.

## Literaturverzeichnis

- Alemann, Heinz von, 1977, Der Forschungsprozeß, Stuttgart.
- Beckett, Sean et al., 1988, The Panel Study of Income Dynamics After Fourteen Years — An Evaluation, in: Journal of Labor Economics, Vol. 6, No. 4, S. 472-492.
- Deutsches Institut für Wirtschaftsforschung (DIW), 1989, Projekt „Das Sozio-ökonomische Panel“ — Antrag auf Förderung der Forschungsphase 1990 und 1991, Berlin.
- Frick, Joachim, 1990, Die SIR-Datenbank des SOEP, Berlin (DIW).
- Galler, Heinz P., 1987, Zur Längsschnittgewichtung des Sozio-ökonomischen Panels, in: Krupp und Hanefeld, S. 295-317.
- Gräf, Lorenz, 1988, Erfahrungen mit einer PC-Datenbank zur Verwaltung von Indikatoren in Panelerhebungen, in: ZANachrichten, Heft 22, S. 31-44.
- Hanefeld, Ute, 1987, Das Sozio-ökonomische Panel — Grundlagen und Konzeption, Frankfurt — New York.
- Hauser, Richard, 1990, Der Sonderforschungsbereich ‚Mikroanalytische Grundlagen der Gesellschaftspolitik‘, in: H. Rapin (Hrsg.), Der private Haushalt im Spiegel sozioempirischer Erhebung, Frankfurt, New York, (in Druck).
- Holm, Kurt, 1982, Die Frage, in: Holm, K. (Hrsg.): Die Befragung, Bd. 1, 2. Auflage, München, S. 136-186.
- Hujer, Reinhard und Gert Wagner, 1985, Grundlagenforschung für eine aktive Gesellschaftspolitik, in: K. Stackmann und A. Streiter (Hrsg.), Sonderforschungsbereiche 1969-1984, Weinheim, S. 37-56.
- Infratest Sozialforschung, 1985, Das Sozio-ökonomische Panel, Welle 1, Methodenbericht zur Haupterhebung, München.
- Infratest Sozialforschung, 1986, Das Sozio-ökonomische Panel, Welle 2, Methodenbericht zur Haupterhebung, München.
- Infratest Sozialforschung, 1987, Das Sozio-ökonomische Panel — Paneldatei, Codebuch Welle 2-4, München.
- Infratest Sozialforschung, 1989a, Das Sozio-ökonomische Panel, Welle 5, Methodenbericht zur Haupterhebung, München.
- Infratest Sozialforschung, 1989b, Das Sozio-ökonomische Panel — Vorerhebung zu Welle 7 „Zeitverwendung und -präferenzen“, München.
- Kaplitza, Gabriele, 1982, Die Stichprobe, in: Holm, K. (Hrsg.): Die Befragung, Bd. 1, 2. Auflage, München, S. 136-186.
- Krupp, Hans-Jürgen und Ute Hanefeld (Hrsg.), 1987, Lebenslagen im Wandel — Analysen 1987, Frankfurt, New York.
- Krupp, Hans-Jürgen und Jürgen Schupp, (Hrsg.), 1988, Lebenslagen im Wandel — Daten 1987, Frankfurt, New York.
- Lazarsfeld, Paul et al., 1957, Die Panel-Befragung, in: R. König (Hrsg.), Praktische Sozialforschung 1, Das Interview, 2. Aufl., S. 253-268.
- Raj, Des, 1968, Sampling Theory, New York.
- Rendtel, Ulrich, 1987, Methodische Konzepte für die Hochrechnung von Panel-Daten. In: Vierteljahrshefte zur Wirtschaftsforschung, Heft 4-1987, S. 278-290.
- Rendtel, Ulrich, 1988a, Repräsentativität und Hochrechnung der Datenbasis, in: Krupp, Schupp (Hrsg.) 1988: Lebenslagen im Wandel: Daten 1987, Frankfurt, New York.
- Rendtel, Ulrich, 1988b, Panelmortalität — Eine Analyse der Antwortausfälle beim Sozio-ökonomischen Panel in der 2. und 3. Befragungswelle. In: Vierteljahrshefte zur Wirtschaftsforschung, Heft 1-2/1988, S. 37-59.
- Rendtel, Ulrich, 1989, Über den Einfluß der Panelselektivität auf Längsschnittanalysen. In: Vierteljahrshefte zur Wirtschaftsforschung, Heft 1-1989, S. 45-61.
- Schupp, Jürgen, 1985, Erfahrungen mit dem Aufbau einer SIR-DBMS-Datenbank beim Sozio-ökonomischen Panel, in: Status GmbH (Hrsg.), Berichtsband SIR/DBMS Benutzerkonferenz, Berlin, S. 117-142.
- Schupp, Jürgen, 1988, Erwerbsbeteiligung und Arbeitsmarkt, in: Krupp, Hans-Jürgen und Jürgen Schupp (Hrsg.); Lebenslagen im Wandel: Daten 1987, Frankfurt-New York, S. 88-113.
- Sonderforschungsbereich 3, 1982, Antrag auf Förderung des Projektes B-5, Das Sozio-ökonomische Panel, Frankfurt — Mannheim.
- Sonderforschungsbereich 3, 1987a, Bericht für die Forschungsphase 1985-1987, Frankfurt, Mannheim.
- Sonderforschungsbereich 3, 1987b, Antrag auf Förderung für die vierte Forschungsphase 1988-1990, Frankfurt, Mannheim.
- Sonderforschungsbereich 3, 1988, Teilprojekt B-5. Das Sozio-ökonomische Panel, Bericht über die Forschungstätigkeit 1985-1988, Antrag auf Förderung der Forschungsphase 1989, Frankfurt, Mannheim, Berlin.
- Sonderforschungsbereich 3, (Hrsg.) 1989, Benutzerhandbuch für ‚Das Sozio-ökonomische Panel‘, Update 3/89, Berlin.
- Wagner, Gert und Michael Lutterbeck, 1985, Ergänzungen zur amtlichen Bevölkerungsstatistik — Strategien für die Datenbeschaffung und den Datenschutz im Sonderforschungsbereich 3, in: Materialien zur Bevölkerungswissenschaft, Nr. 42, Wiesbaden, S. 29-56.
- Wagner, Gert, 1990, Das Sozio-ökonomische Panel — Ein Instrument zur Dauerbeobachtung privater Haushalte, in: H. Rapin (Hrsg.): Der private Haushalt im Spiegel sozioempirischer Erhebungen, Frankfurt, New York (in Druck).
- Witte, James und Herbert Lahmann, 1988, Residential Mobility of One-Person Households. In: Vierteljahrshefte zur Wirtschaftsforschung, Heft 1-2/1988, S. 60-78.
- Wolter, Kirk, 1985, Introduction to Variance Estimation, New York.