

Rendtel, Ulrich

Working Paper — Digitized Version

Dokumentation des Sozio-oekonomischen Panels (SOEP):
Erhebungsdesign, Fallzahlen und erhebungsbedingte Ausfälle sowie
die Schätzung von Ausfallwahrscheinlichkeiten bis Welle 11 (1984 bis
1994)

DIW Discussion Papers, No. 110

Provided in Cooperation with:

German Institute for Economic Research (DIW Berlin)

Suggested Citation: Rendtel, Ulrich (1995) : Dokumentation des Sozio-oekonomischen Panels (SOEP): Erhebungsdesign, Fallzahlen und erhebungsbedingte Ausfälle sowie die Schätzung von Ausfallwahrscheinlichkeiten bis Welle 11 (1984 bis 1994), DIW Discussion Papers, No. 110, Deutsches Institut für Wirtschaftsforschung (DIW), Berlin

This Version is available at:

<https://hdl.handle.net/10419/95816>

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.



Diskussionspapiere
Discussion Papers

Diskussionspapier Nr. 110

**Dokumentation des Sozio-oekonomischen Panels (SOEP):
Erhebungsdesign, Fallzahlen und erhebungsbedingte
Ausfälle sowie die Schätzung von
Ausfallwahrscheinlichkeiten bis Welle 11 (1984 bis 1994)**

von
Ulrich Rendtel

Die in diesem Papier vertretenen Auffassungen liegen ausschließlich in der Verantwortung des Verfassers und nicht in der des Instituts.

Opinions expressed in this paper are those of the author and do not necessarily reflect the views of the Institute.

Deutsches Institut für Wirtschaftsforschung

Diskussionspapier Nr. 110

**Dokumentation des Sozio-oekonomischen Panels (SOEP):
Erhebungsdesign, Fallzahlen und erhebungsbedingte
Ausfälle sowie die Schätzung von
Ausfallwahrscheinlichkeiten bis Welle 11 (1984 bis 1994)**

von
Ulrich Rendtel

Berlin, im April 1995

Deutsches Institut für Wirtschaftsforschung, Berlin
Königin-Luise-Str. 5, 14191 Berlin
Telefon: 49-30 - 89 7 89 -0
Telefax: 49-30 - 89 7 89 -200

Dokumentation des Sozio-oekonomischen Panels (SOEP): Erhebungsdesign, Fallzahlen und erhebungsbedingte Ausfälle sowie die Schätzung von Ausfallwahrscheinlichkeiten bis Welle 11 (1984 bis 1994)

Zusammenfassung

Es wird ein Überblick über das Erhebungsdesign der ersten Welle des SOEP gegeben. Dieses ist für die drei SOEP-Teilstichproben A (West-Deutsche), B (Ausländer) und C (Ost-Deutsche) unterschiedlich. Weiterhin werden die Weiterverfolgsregeln des SOEP beschrieben.

Die Entwicklung der Fallzahlen wird getrennt für die Teilstichproben im Quer- und im Längsschnitt dargestellt. Ab Welle 7 (1990) werden die Wanderungen der Stichprobenmitglieder zwischen Ost- und Westdeutschland dokumentiert. Dabei wird auch die Trennung nach Privathaushalten und dem Anstaltsbereich berücksichtigt.

Die erhebungsbedingten Ausfälle werden auf zwei Ebenen beschrieben: Ausfälle aufgrund von Kontaktverlust und Ausfälle bei wiedererreichten Haushalten aufgrund von nicht gewährten Interviews. Auf diesen zwei Ebenen werden Ausfallraten nach einzelnen Haushaltsmerkmalen als auch die Schätzergebnisse bei Berücksichtigung mehrerer Merkmale dokumentiert. Die Schätzungen basieren auf Modellen der multiplen logistischen Regression. Sie sind die empirische Grundlage für die Längsschnittgewichtung des SOEP.

Die Dokumentation schließt mit einem Hinweis auf weiterführende Literatur.

Inhaltsverzeichnis

1	Das Erhebungsdesign der Startwelle des SOEP	3
1.1	Teilstichprobe A	3
1.2	Teilstichprobe B	4
1.3	Teilstichprobe C	4
2	Die Weiterverfolgungsregeln des SOEP	5
3	Entwicklung der Fallzahlen	6
3.1	Entwicklung der Anzahl der realisierten Interviews im Querschnitt	6
3.2	Entwicklung der feldbedingten Ausfälle im Längsschnitt	11
3.3	Gewinne der Stichprobe durch Geburten und Zuzüge	17
4	Ausfälle durch Kontaktverlust	18
4.1	Ausfallraten nach Mobilitätsverhalten	18
4.2	Definition der Modellvariablen für die Logit-Analyse	20
4.3	Geschätzte Wahrscheinlichkeiten für einen Ausfall durch Kontaktverlust	21
5	Ausfälle durch Antwortverweigerung	24
5.1	Ausfallraten nach verschiedenen Haushaltsmerkmalen	24
5.2	Definition der Modellvariablen	32
5.3	Geschätzte Wahrscheinlichkeiten für einen Ausfall durch Antwortverweigerung erreichter Haushalte	35
6	Literaturhinweise	41

1 Das Erhebungsdesign der Startwelle des SOEP

In diesem Abschnitt wird das Erhebungsdesign der Startwelle für die SOEP-Teilstichproben A (West-Deutsche), B (Ausländer) und C (Ost-Deutsche) dargestellt. Trotz vieler gemeinsamer Design-Elemente unterscheidet sich die Auswahl der Starthaushalte in den drei Teilstichproben. Die Weiterverfolgungsregeln der Mitglieder der Starthaushalte sind jedoch in allen Teilstichproben identisch.

Gemeinsam ist allen drei Teilstichproben die Zweistufigkeit des Auswahlverfahrens. Weiterhin wurde sowohl für die Auswahl der Primäreinheiten als auch der Sekundäreinheiten (Haushalte) bei allen drei Stichproben *größenproportionales systematisches Ziehen mit Intervall und zufälliger Startzahl* gewählt.

Die Primäreinheiten (Sample Points) sind in Teilstichprobe

A: Wahlweise

B: Kreise bzw. kreisfreie Städte

C: Gemeinden

1.1 Teilstichprobe A

Auswahl der Primäreinheiten

Die Auswahl der Primäreinheiten der Teilstichprobe A ist 2-phasig. Die erste Phase ist mit der Ziehung von 16 Netzen des 82-er ADM-Master Samples identisch:

- Anordnung aller Wahlkreise der (alten) Bundesrepublik nach regionalen Kriterien (Bundesland, Regierungsbezirk, Gemeindetyp, Wahlkreisnummer). Größe der Primäreinheiten ist die geschätzte Anzahl der Privathaushalte im Wahlkreis. Nach der regionalen Anordnung der Wahlkreise und der Bestimmung ihrer Haushaltgröße wird systematisches Ziehen mit Intervall und zufälliger Startzahl durchgeführt. Insgesamt wurden $16 \times 210 = 3360$ Wahlkreise gezogen¹.

In der zweiten Phase wird aus dem Master-Sample eine weitere Stichprobe gezogen:

- Aus den Wahlkreisen des Master-Samples wurde eine *geschichtete Stichprobe* von 563 Sample Points gezogen. Die Schichten (Regionalzellen) wurden durch die Merkmale Bundesland, Regierungsbezirk und Gemeindetyp definiert. Die Schichtgröße war proportional zur geschätzten Anzahl der Privathaushalte in der Schicht.

Innerhalb jeder der 148 Regionalzellen wurden die Wahlkreise ihrerseits angeordnet. Ihre Größe ist wiederum durch die geschätzte Anzahl der Haushalte des Wahlkreises gegeben. Innerhalb jeder Schicht wurde systematisches *größenproportionales Ziehen mit Intervall und zufälliger Startzahl* benutzt.

¹Diese Sample-Points entsprechen 16 Netzen der 82-er ADM-Stichprobe, vgl. Kirschner (1984).

Auswahl der Sekundäreinheiten

Begehung des Wahlkreises nach festen Begehungsregeln (Random Route Verfahren]. Von einer zufälligen Startadresse ausgehend wurden die Klingelknöpfe von Wohnungen längs der Random-Route abgezählt. Jede 7. Adresse wurde ausgewählt. Pro Sample-Point wurden 10 Adressen ausgewählt. Wurde bei weniger als 8 Haushalten in einem Sample Point ein Interview realisiert, so wurden noch 2 weitere Adressen der Random-Route Begehung eingesetzt. Das Verfahren kann als Systematisches Ziehen mit Intervall und zufälligem Start interpretiert werden.

1.2 Teilstichprobe B

Auswahl der Primäreinheiten

- Die Teilstichprobe B wurde geschichtet gezogen. Für Griechen, Italiener, Yugoslawen und Spanier wurde eine Schichtgröße (Kreise bzw kreisfreie Städte) von 40, für Türken eine Schichtgröße von 80 festgesetzt. Die Bestimmung der Schichtgrößen sollte eine Mindestgröße der Ausländer-Stichprobe für die genannten Nationalitäten gewährleisten.

- Innerhalb jeder Schicht wurden die Primäreinheiten regional angeordnet (Bundesländer, Regierungsbezirk) und es wurde systematisches Ziehen mit Intervall und zufälliger Startzahl durchgeführt. Die Größe der Primäreinheiten war durch die Anzahl der Ausländer der jeweiligen Nation gegeben².

Auswahl der Sekundäreinheiten

- Zunächst wurden pro Sample-Point die Adressen von 14 Personen ausgewählt. Die Ziehung basierte auf den Ausländerregistern der Primäreinheiten. Die Ziehung erfolgte durch systematisches Ziehen mit zufälliger Startzahl und Intervall anhand der Register. Auswahlbasis waren alle Personen der jeweiligen Nationalität, die das Befragungsalter von 16 Jahren erreicht hatten.

- Die Adressen der 14 Personen wurden unter regionalen Gesichtspunkten zu 2 Klumpen mit 7 Adressen geordnet. Einer der beiden Klumpen wurde zufällig gewählt.

1.3 Teilstichprobe C

Auswahl der Primäreinheiten

Die Auswahl der Primäreinheiten der Teilstichprobe C ist 2-phasig. Die erste Phase ist mit der Ziehung des Infratest-Master Samples für die DDR identisch:

- Auswahl von 753 Sample-Points über eine geschichtete Auswahl. Regionale Schichtung (Land-/Stadtkreise und Gemeindegrößenklassen). Schichtgröße proportional zur Anzahl der Einwohner. Innerhalb jeder Schicht systematisches

²Datenbasis war eine Sonderauswertung des Ausländer Zentralregisters, Stand 1982.

größenproportionales Ziehen mit Intervall und zufälliger Startzahl. Die Größe ist durch die Anzahl der Einwohner gegeben.

- Von diesen 753 Sample-Points wurden 360 Primäreinheiten ausgewählt. Die Schichtung war durch dieselben Regionalmerkmale wie beim ersten Schritt gegeben. Allerdings wurden jetzt die Schichtgrößen proportional zu der Anzahl der Haushalte bestimmt. Innerhalb jeder Schicht wurde wieder systematisches größenproportionales Ziehen mit Intervall und zufälliger Startzahl benutzt. Die Größe war dabei über die Anzahl der Haushalte gegeben.

Auswahl der Sekundäreinheiten

Begehung der Gemeinde nach denselben Begehungsregeln wie in Teilstichprobe A. Auswahl von 10 Adressen pro Sample-Points. Wurden in den 8 ersten Haushalten alle Interviews realisiert, so wurden die restlichen 2 Adressen nicht mehr eingesetzt.

2 Die Weiterverfolgungsregeln des SOEP

Die Weiterverfolgungsregeln des SOEP gelten für alle Teilstichproben des SOEP: Für alle Personen, die jemals ein Personen Interview gegeben haben, wird im Falle ihres Umzugs die neue Adresse ermittelt. Liegt die neue Adresse im Erhebungsgebiet, so wird versucht, von allen Personen des neuen Haushalts ein Interview zu erhalten. Das Erhebungsgebiet des SOEP ist 1990 auf die neuen Bundesländer erweitert worden.

In Haushalten ohne Umzug werden alle Personen, die das Befragungsalter von 16 Jahren überschritten haben, befragt. Dies schließt zugezogene Personen und Kinder, die das Befragungsalter erreichen, ein.

Die Weiterverfolgungsregeln des SOEP garantieren zwei Ziele:

- Neben den demographischen Verlusten durch Tod und Auswanderung gibt es keine Verluste für die Stichprobe durch eine absichtliche Nicht-Weiterverfolgung von Befragungspersonen.
- Alle Personen, die in einem befragten Haushalt leben, werden interviewt.

Personen, die die weitere Teilnahme nicht explizit ablehnen, werden weiterverfolgt. Erst wenn sie in der folgenden Panelwelle wiederum kein Interview geben, gelten sie als endgültiger Ausfall. Personen mit expliziter Verweigerung oder endgültigem Ausfall werden nicht weiter interviewt.

Für das SOEP sind keine "Auffrischungs"- Stichproben zum "Ausgleich" der erhebungsbedingten Ausfälle während des Panels vorgesehen. Die Behandlung dieser Ausfälle geschieht im SOEP durch die Bereitstellung geeigneter GewichtungsvARIABLEN, vgl. Rendtel (1991).

Die Zuwanderungen seit 1984 (Start des SOEP) werden 1994 im Rahmen einer Zuwanderer-Stichprobe berücksichtigt, vgl. Schupp/Wagner (1995) .

3 Entwicklung der Fallzahlen

Die Entwicklung der Fallzahlen wird hier unter folgenden Gesichtspunkten dokumentiert:

- Vergleich der Anzahl der realisierten Interviews im Querschnitt
- Entwicklung der feldbedingten Ausfälle im Längsschnitt
- Gewinne der Stichprobe durch Geburten und Zuzüge

3.1 Entwicklung der Anzahl der realisierten Interviews im Querschnitt

Die folgenden Abbildungen dokumentieren die Anzahl der realisierten Interviews in folgenden Darstellungen:

Abb. 1: Vergleich Personen — Haushalte (A+B) Welle 1 bis 11.

Abb. 2: Vergleich Stichproben A und B (Personen) Welle 1 bis 11.

Abb. 3: Vergleich Stichproben A, B und C (Personen) Welle 1 bis 5.

Abbildung 1: Vergleich der realisierten Interviews von Personen und Haushalten (Teilstichproben A+B), Welle 1 bis 11.

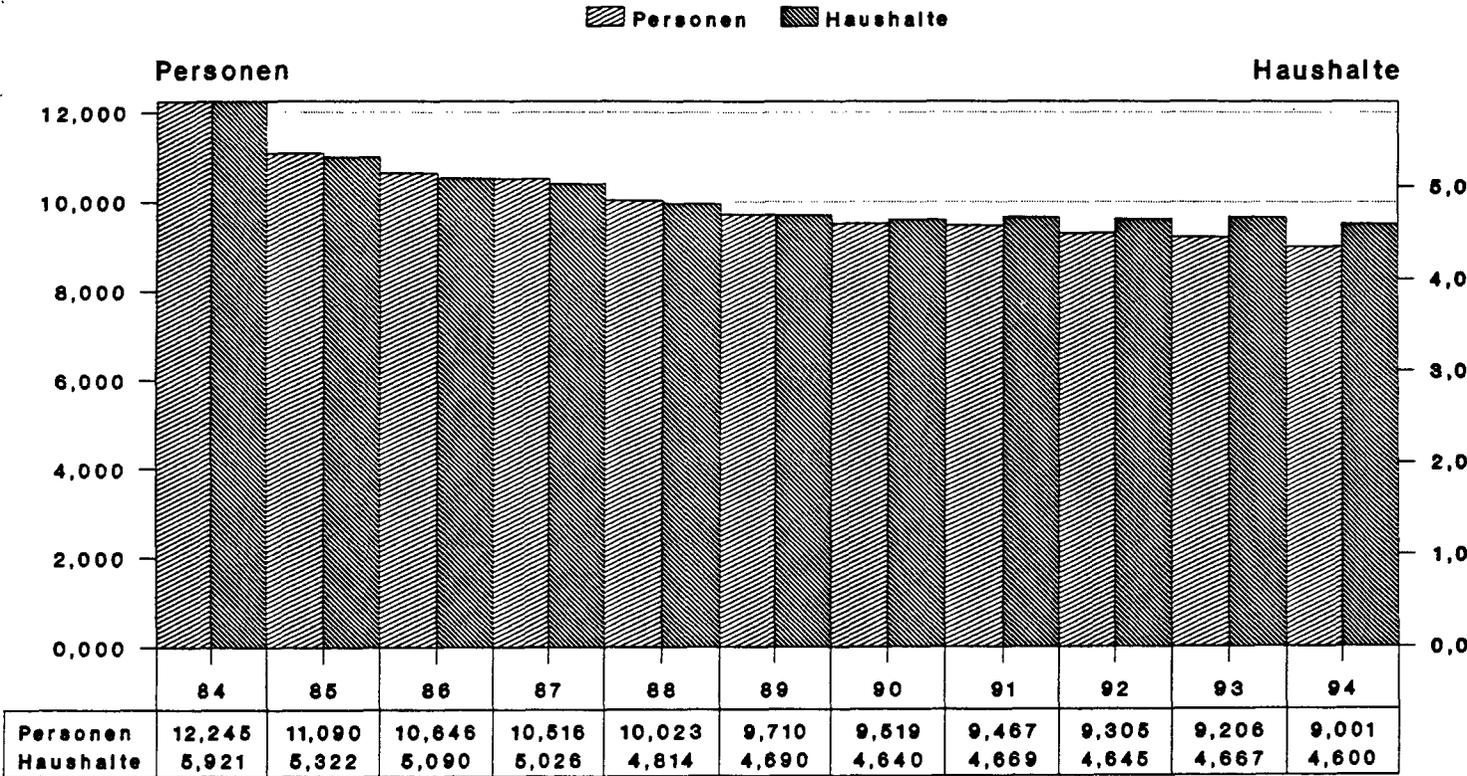


Abbildung 2: Vergleich der realisierten Interviews in den Teilstichproben A und B (Personen), Welle 1 bis 11.

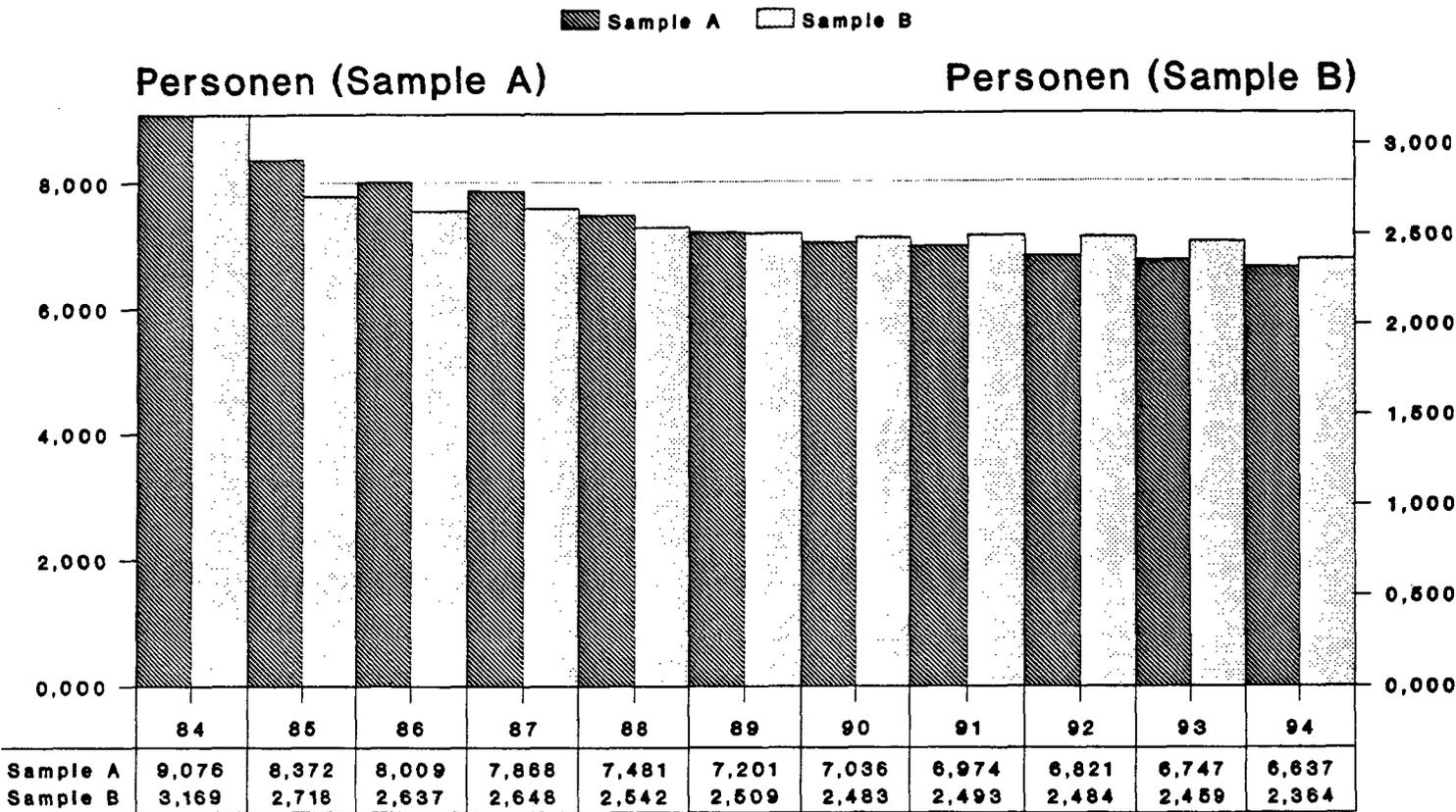
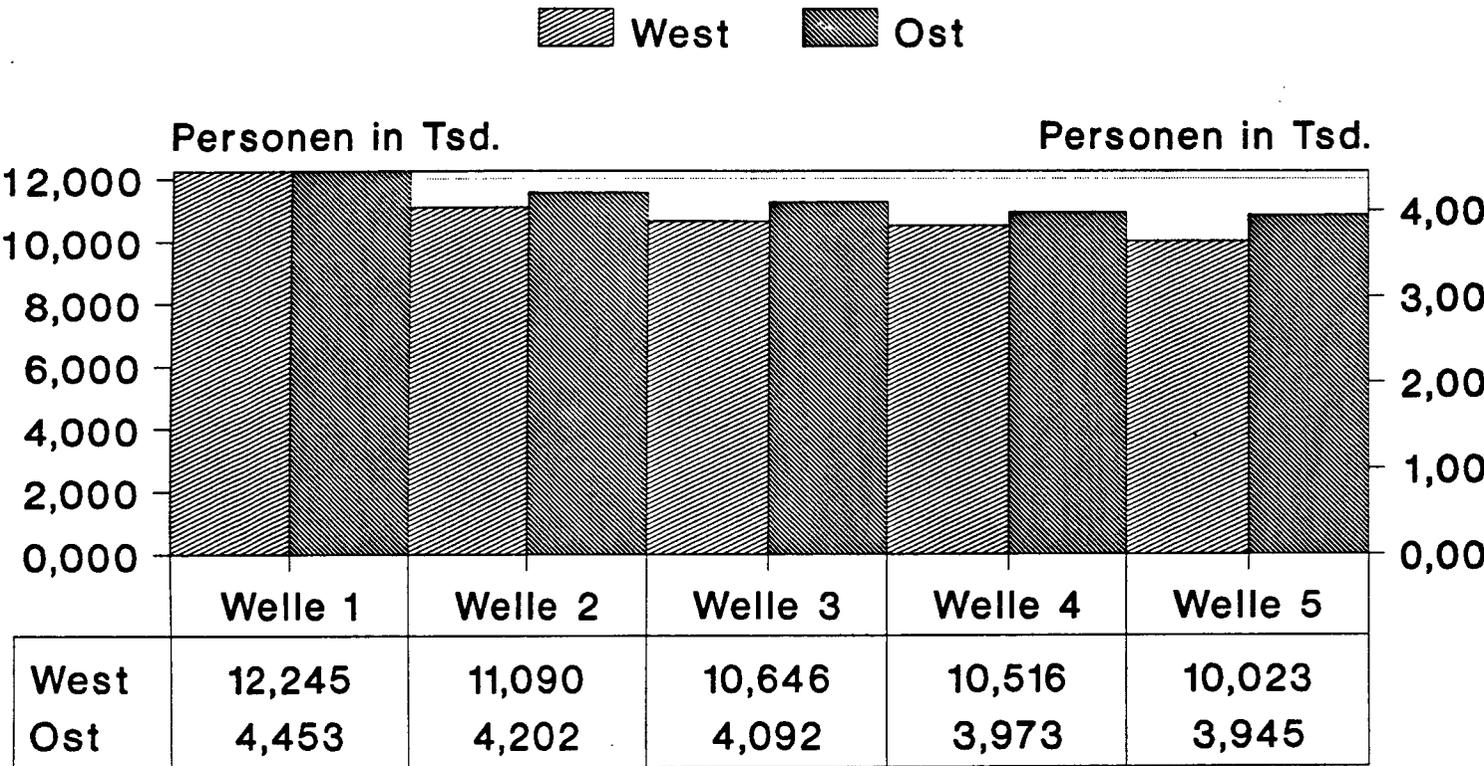


Abbildung 3: Vergleich der realisierten Interviews in den Teilstichproben A, B und C (Personen), Welle 1 bis 5.



Mit zunehmender Dauer des SOEP ist die Stichprobenzugehörigkeit ein immer ungenauerer Indikator für das aktuelle Erhebungsgebiet, weil Mitglieder der Stichprobe C (Ostdeutschland) nach Westdeutschland verziehen oder umgekehrt Mitglieder der Stichproben A und B (Westdeutschland) nach Ostdeutschland verziehen. Die Tabelle 1 dokumentiert daher das aktuelle Erhebungsgebiet der Panelhaushalte ab 1990 (Welle 7).

Tabelle 1: **Entwicklung der Fallzahlen nach Erhebungsgebiet und Anstaltsbereich, 1990 bis 1994.** n = Anzahl der Interviews. N = Hochgerechnete Anzahl in Tsd. Die Eckzahlen für Haushalte und die Personen in Privathaushalten wurden an den Mikrozensus angepaßt. Die Eckzahlen für 1994 sind vorläufig. Wegen unterschiedlicher definitorischer Konzepte stimmen die Zahlen für den Anstaltsbereich nicht mit dem Mikrozensus überein.

Erhebungsjahr		Erhebungsgebiet							
		West				Ost			
		Stichp. A+B		Stichp. C		Stichp. C		Stichp. A+B	
	Privat	Anstalt	Privat	Anstalt	Privat	Anstalt	Privat	Anstalt	
Haushalte									
1990	n	4592	48	-	-	2158	21	-	-
	N	28173	417	-	-	6703	90	-	-
1991	n	4620	49	22	-	1988	20	-	-
	N	28466	408	117	-	6672	108	-	-
1992	n	4598	46	58	3	1946	13	1	-
	N	28776	387	266	18	6654	71	2	-
1993	n	4609	53	78	5	1878	9	5	-
	N	29310	436	399	28	6615	50	40	-
1994	n	4545	47	93	5	1850	11	8	-
	N	29562	446	437	24	6698	77	111	-
Personen (inklusive Kinder)									
1990	n	12151	59	-	-	6014	30	-	-
	N	62380	471	-	-	16313	120	-	-
1991	n	12100	61	44	-	5613	26	-	-
	N	62971	455	236	-	15808	128	-	-
1992	n	11884	58	133	3	5331	18	2	-
	N	63441	434	559	18	15618	84	4	-
1993	n	11724	63	182	5	5078	11	7	-
	N	63934	464	836	24	15501	54	44	-
1994	n	11467	55	225	5	4938	13	11	-
	N	64387	452	1012	17	15338	82	161	-

3.2 Entwicklung der feldbedingten Ausfälle im Längsschnitt

Die folgenden Abbildungen dokumentieren die Entwicklung der feldbedingten Ausfälle unter folgenden Aspekten:

Abb. 4: Alle Stammpersonen der 1. Welle (A+B). Entwicklung bis Welle 11.

Abb. 5: Alle Stammpersonen der 1. Welle nur A. Entwicklung bis Welle 11.

Abb. 6: Alle Stammpersonen der 1. Welle nur B. Entwicklung bis Welle 11.

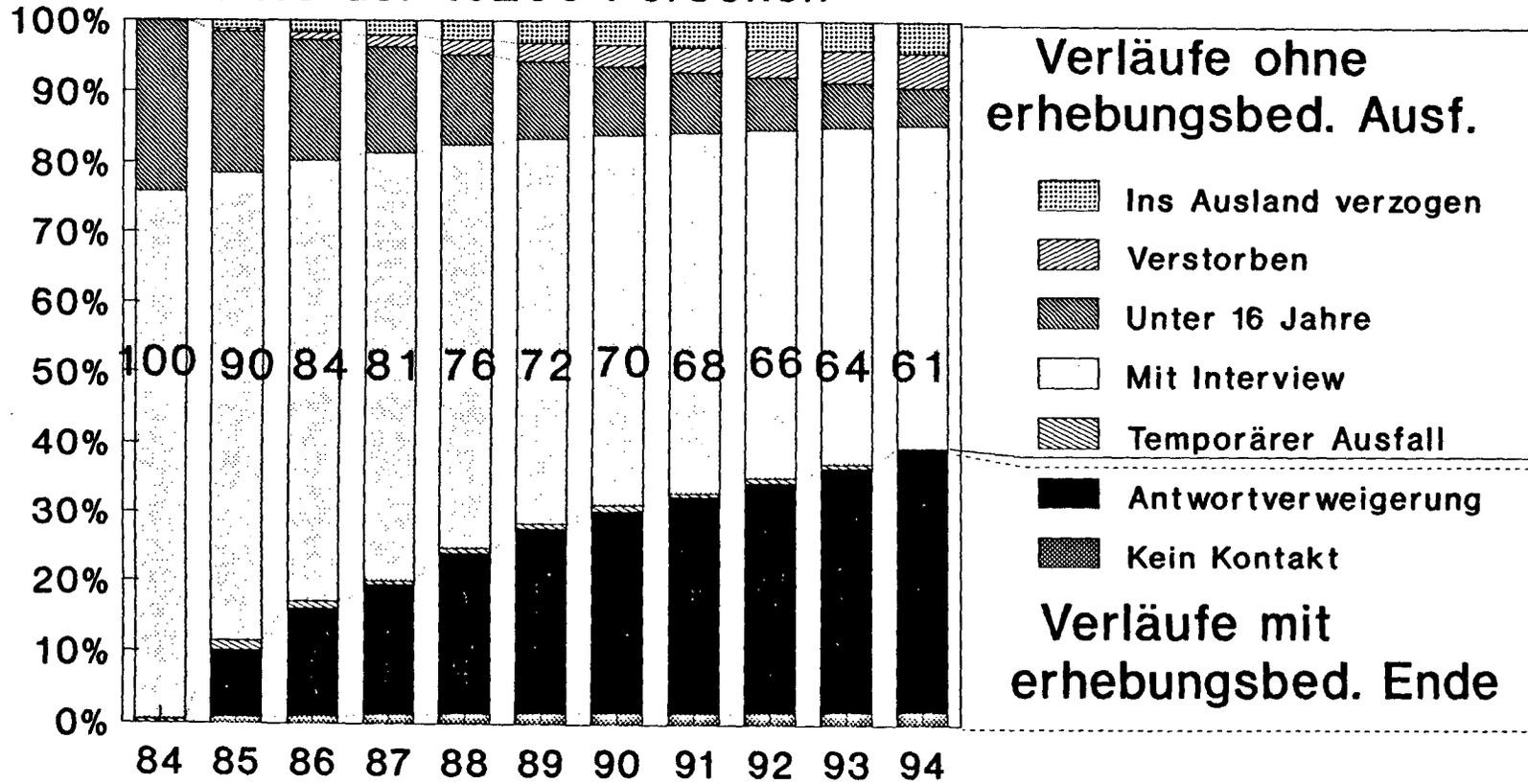
Abb. 7: Alle Stammpersonen der 1. Welle nur C. Entwicklung bis Welle 5.

Abb. 8: Alle Stammpersonen der 1. Welle. Vergleich der Entwicklung bis Welle 5.

Die in den Abbildungen hervorgehobenen Zahlen geben den Anteil der Verläufe an, die nicht erhebungsbedingt enden. Diese Zahlen sind damit Maßzahlen für den Umfang der Panelmortalität.

Abbildung 4: Alle Stammpersonen der 1. Welle (A+B). Entwicklung bis Welle II.

vH Verbleib der 16205 Personen



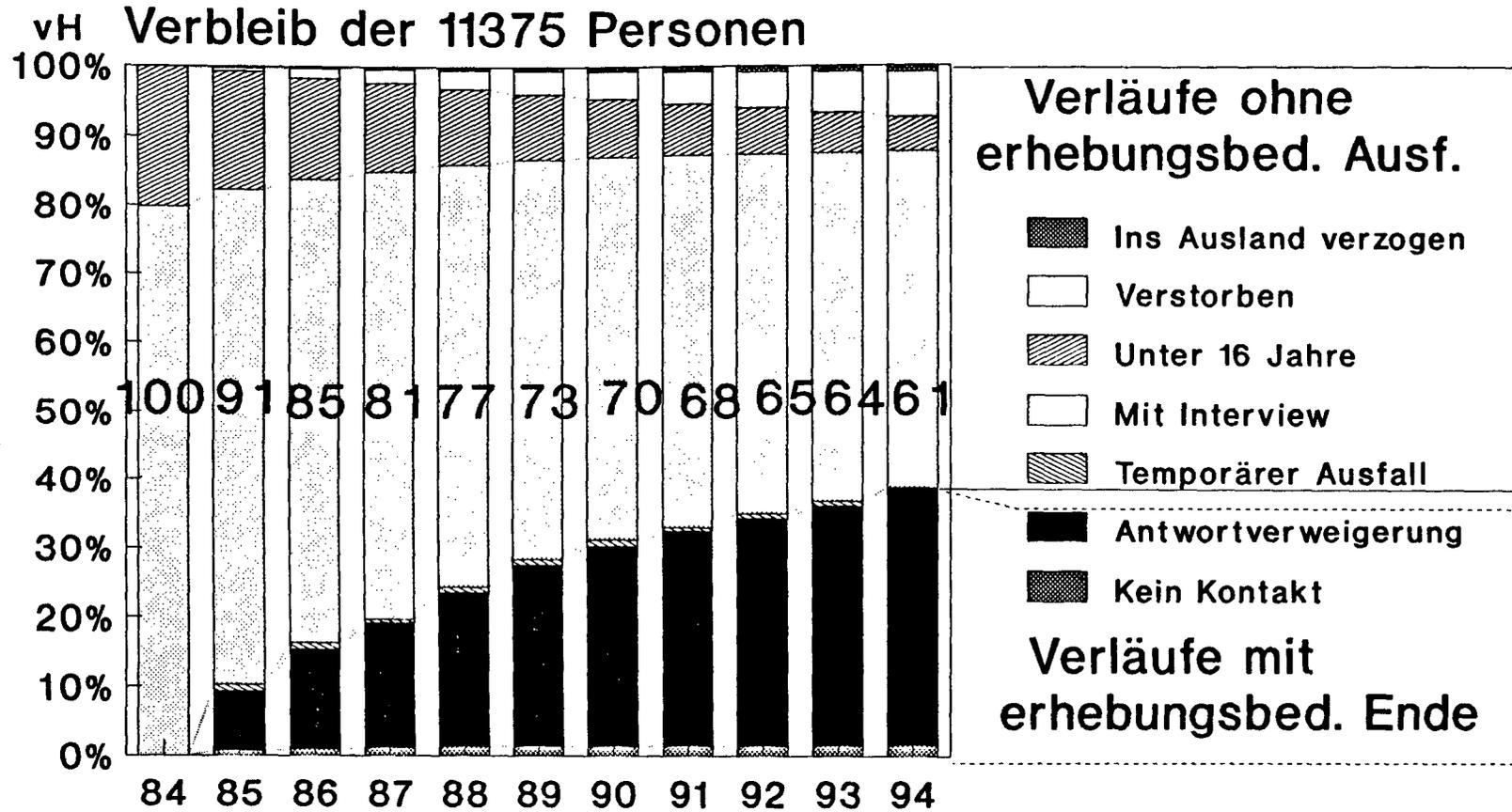
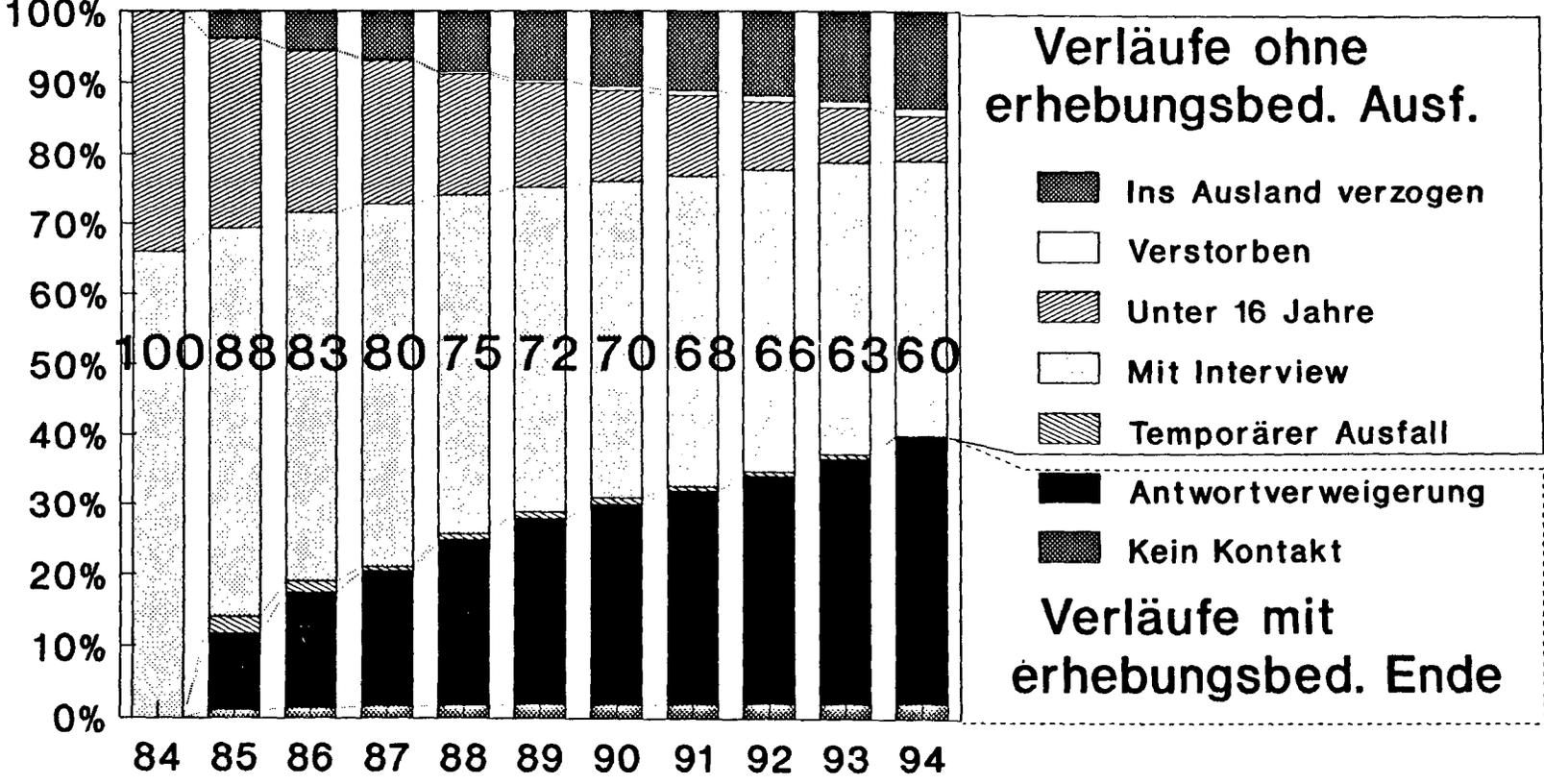


Abbildung 5: Alle Stammpersonen der 1. Welle Stichprobe A. Entwicklung bis Welle 11.

Abbildung 6: Alle Stammpersonen der 1. Welle Stichprobe B. Entwicklung bis Welle 11.

vH Verbleib der 4830 Personen



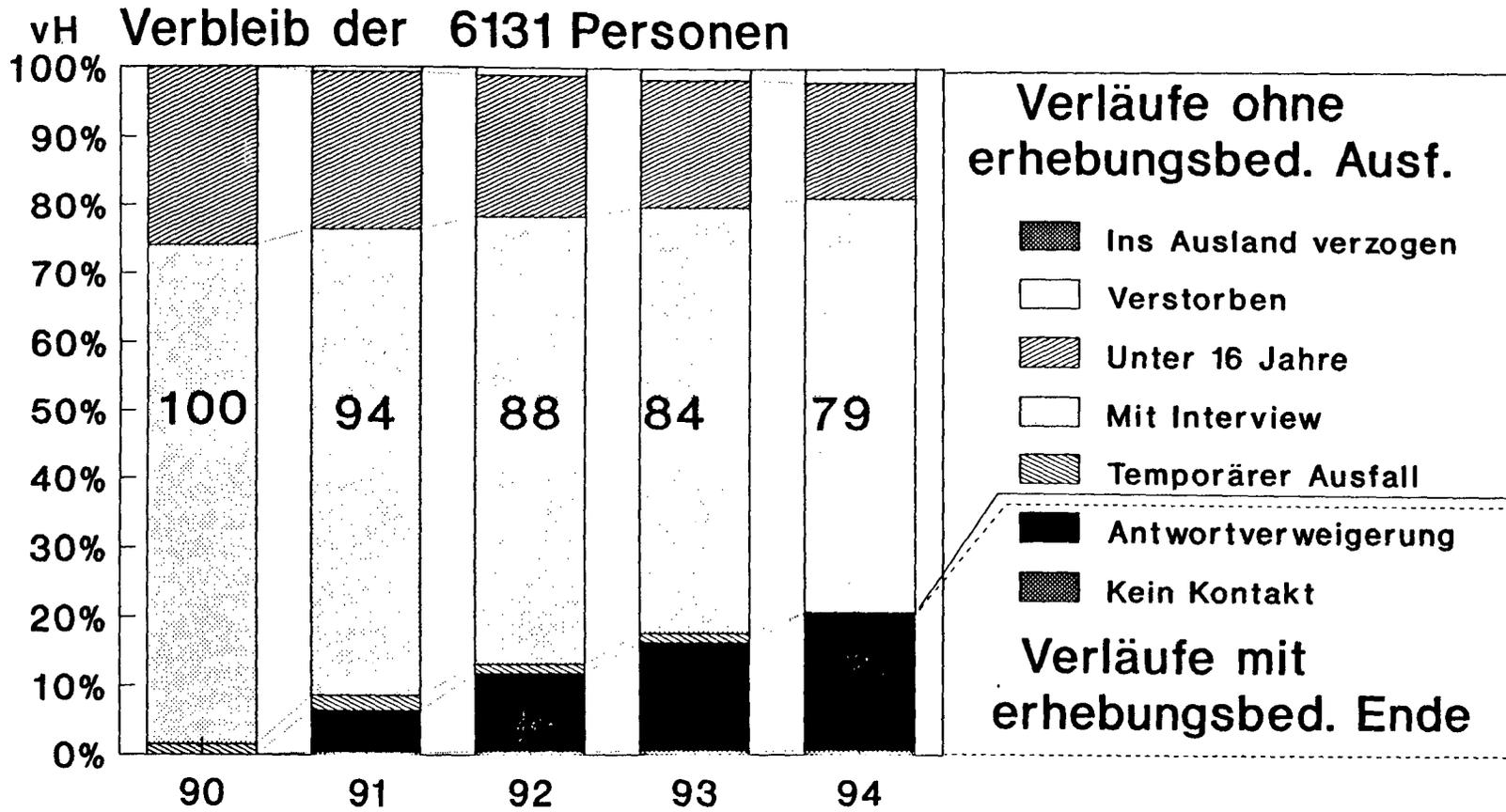


Abbildung 7: Alle Stammpersonen der 1. Welle (Ost) Stichprobe C.
Entwicklung bis Welle 5 (Ost).

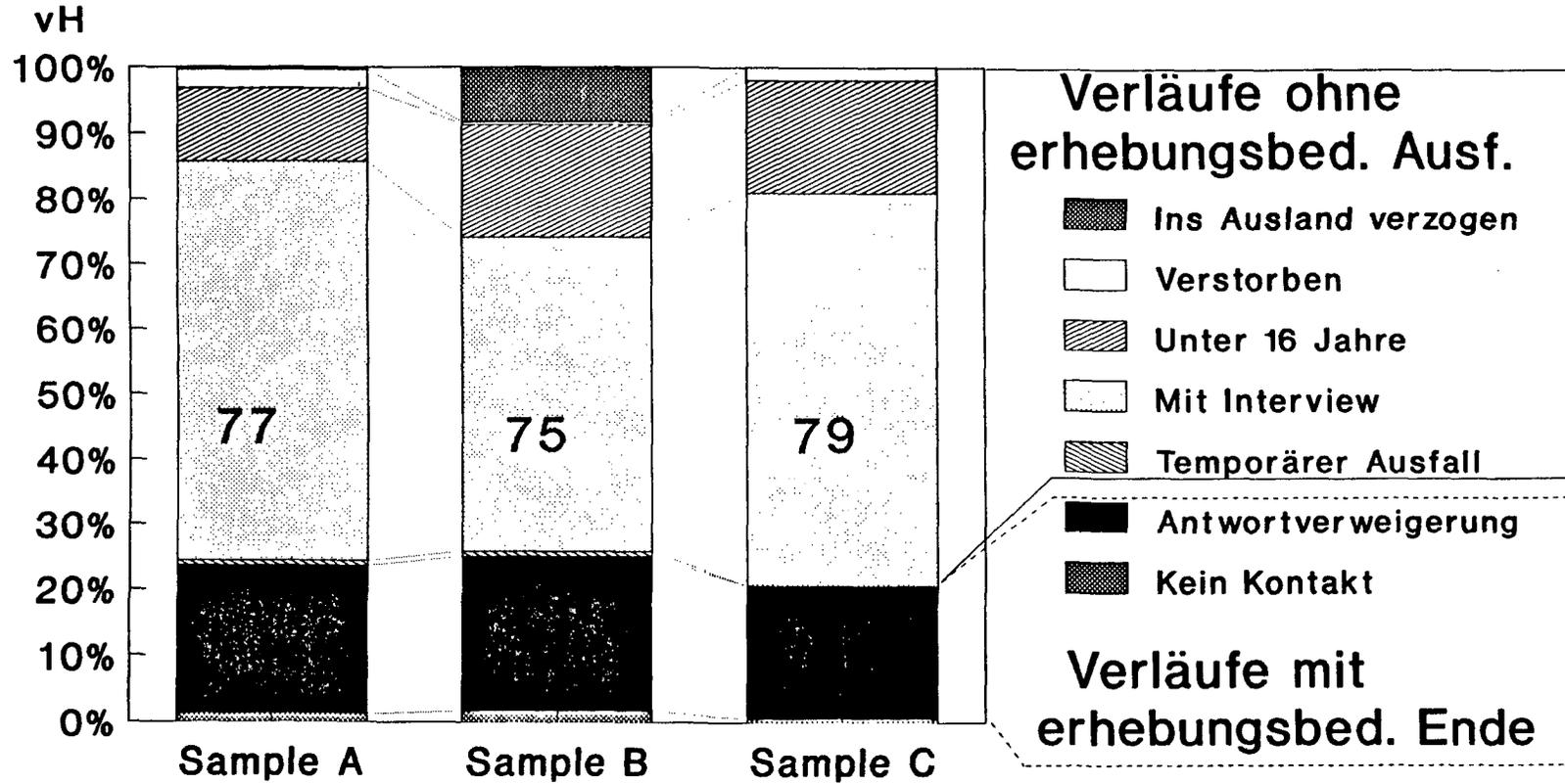
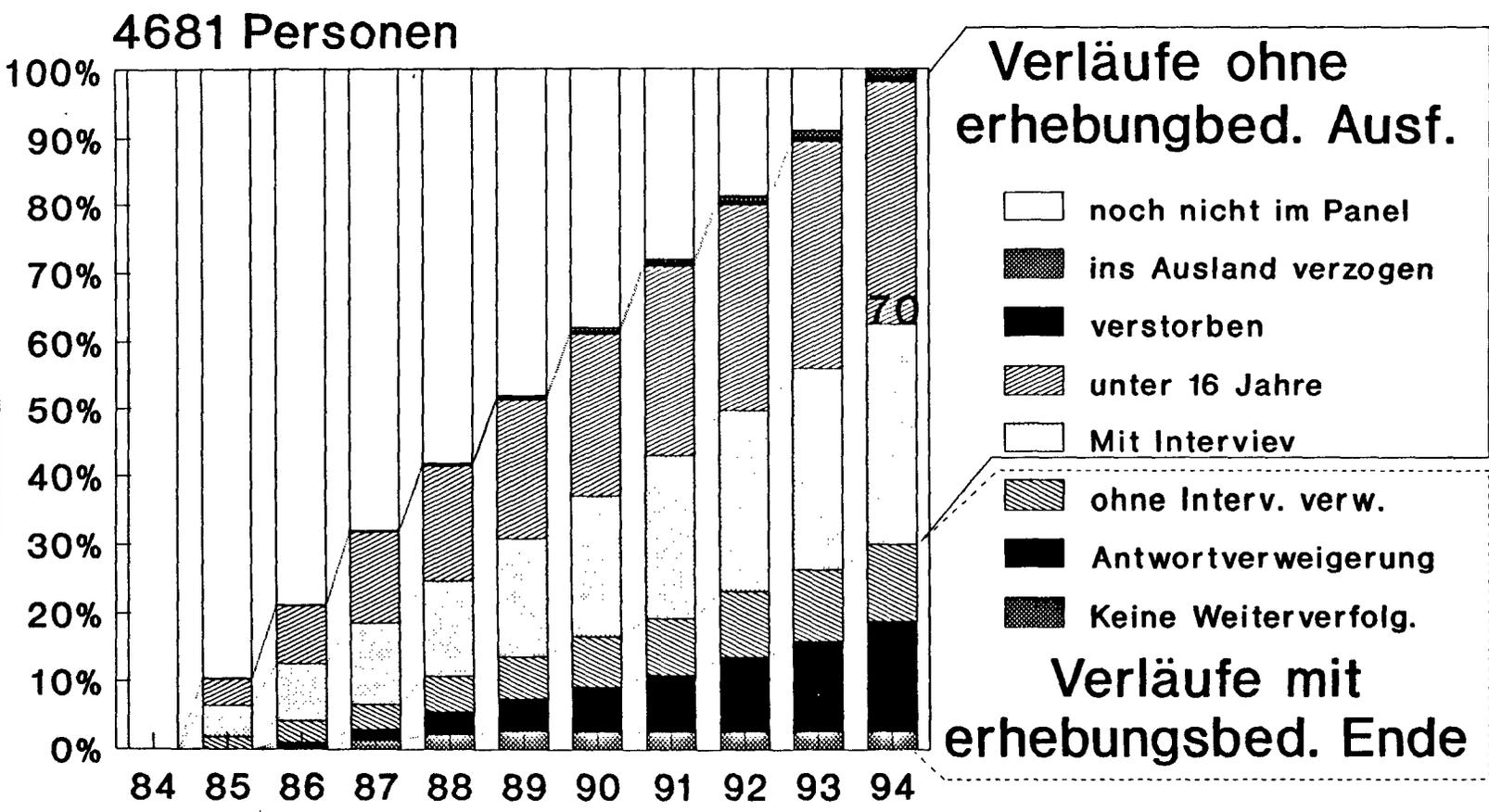


Abbildung 8: Alle Stammpersonen der 1. Welle. Vergleich der Entwicklung bis Welle 5.

3.3 Gewinne der Stichprobe durch Geburten und Zuzüge

Abbildung 9: Gewinne der Stichprobe (A+B) durch Geburten und Zuzüge.



4 Ausfälle durch Kontaktverlust

In jeder Panelwelle ist es notwendig, mit den Haushalten des Vorjahres wieder Kontakt aufzunehmen. Dabei wird festgestellt, ob:

- der Haushalt noch an der alten Adresse lebt.
- der Haushalt umgezogen ist.
- der Haushalt sich aufgespalten hat.
- der Haushalt verstorben ist.
- der Haushalt das Erhebungsgebiet verlassen hat.
- die Haushaltsmitglieder in einen bestehenden Panelhaushalt zurückgekehrt sind.

4.1 Ausfallraten nach Mobilitätsverhalten

Die folgende Tabelle dokumentiert den Erfolg der Feldarbeit bei der Kontaktaufnahme. Basis sind alle Haushalte der letzten Welle, die nicht verstorben sind oder ins Ausland gezogen sind, weiterhin alle abgespaltenen Haushalte. Ein Kontakt gilt als hergestellt, falls ein Interview erreicht wird bzw. die Verweigerung eines Interviews im Adressprotokoll vermerkt wird. Weiterhin gilt ein Kontakt als wiederhergestellt, falls der Rückzug der Haushaltsmitglieder in einen bestehenden Panelhaushalt festgestellt wird.

Tabelle 2: Ausfallraten aufgrund von Kontaktverlust im SOEP-West (Sample A und B). N = Anzahl der zu erreichenden Haushalte; % = Prozentsatz der Haushalte ohne Kontaktaufnahme.

Welle:	2	3	4	5	6	7	8	9	10	11
	Insgesamt									
N	6051	5814	5465	5342	5156	5044	5029	5006	5049	5008
%	1,9	1,4	1,0	0,9	0,9	0,9	0,5	0,4	0,9	0,8
	Haushalte ohne Adressenänderung									
N	5413	5039	4808	4683	4545	4472	4448	4447	4395	4359
%	0,8	0,4	0,1	0,1	0,2	0,0	0,04	0,0	0,02	0,1
	verzogene Mehrpersonen-Haushalte									
N	298	307	272	274	228	186	197	195	231	239
%	7,4	3,6	4,0	5,5	0,5	1,6	0,5	0,5	0,9	0,0
	verzogene Einpersonen-Haushalte									
N	119	180	142	143	126	122	94	90	105	146
%	21,0	14,4	7,7	5,6	4,7	5,7	1,1	0,0	7,6	6,2
	abgespaltene Haushalte									
N	221	295	242	242	246	263	290	273	317	264
%	11,7	8,4	10,4	7,4	11,8	12,9	7,6	7,3	10,7	9,9

Tabelle 3: Ausfallraten durch Kontaktverlust im SOEP-Ost von der 2. bis 5. Befragungswelle. N = Anzahl der zu erreichenden Haushalte; % = Prozentsatz der Haushalte ohne Kontaktaufnahme.

Merkmal	Welle 2		Welle 3		Welle 4		Welle 5	
	N	%	N	%	N	%	N	%
insgesamt	2246	1.5	2304	0.5	2227	0.9	2136	0.6
Haushalt ohne Umzug	2062	0.0	2043	0.05	2021	0.05	1904	0.0
umgezogener Mehrpers.-H.	81	11.1	106	0.0	82	3.7	92	2.2
umgezogener Einpers.-H.	21	14.3	43	9.3	14	0.0	39	2.6
abgespaltener Haushalt	82	25.6	112	6.3	110	13.6	104	8.6

4.2 Definition der Modellvariablen für die Logit-Analyse

Für die Schätzung der Wahrscheinlichkeit, daß ein Haushalt durch Kontaktverlust ausfällt, wurden die folgenden Merkmale benutzt:

Merkmal	Bezeichnung	Code	Ausprägungen
Umzug	UMZUG	1	Haushalt nicht umgezogen
		2	Umgezogener Mehrpersonen-Haushalt
		3	Umgezogener Einpersonen-Haushalt
		4	Abgespaltener Haushalt
Großstadt	GROSS	0	sonst
		1	Kernregion (nach Boustedt) mit mehr als 500 TSD Einwohnern (West) bzw. Gemeindegrößenkl. 500 TSD u. m. (Ost)
Haushaltsgröße	HHGR	1	1 Personen-Haushalt
		2	2 Personen-Haushalt
		3	3 Personen-Haushalt
		4	4 und mehr Personen-Haushalt
Einpersonen-HH.	EINP	0	sonst
		1	Einpersonenhaushalt
Wohnumfeld	WUM	1	Einzelhaus oder ländliches Umfeld
		2	Hochhaus oder Haus mit mehr als 10 Mietparteien
		3	sonst
Abgespaltener Haushalt	SPLIT	1	Verzogener Mehrpersonen-Haushalt
		2	Verzogener Einpersonen-Haushalt
		3	Abgespaltener Haushalt

4.3 Geschätzte Wahrscheinlichkeiten für einen Ausfall durch Kontaktverlust

Die im vorhergehenden Abschnitt definierten Merkmale wurden im Rahmen einer multiplen Logit-Analyse zur Schätzung von $P_K = \text{Pr}(\text{Kontakt}=\text{nein})$ benutzt. Für die Hochrechnung wurde ein Modell gewählt, das nur noch signifikante Variablen enthält.

Interpretation der Koeffizienten:

$$\ln \frac{P_{K,i}}{1 - P_{K,i}} = \text{const} + X'_i \beta$$

Positive Koeffizienten zeigen eine Vergrößerung des Ausfallrisikos im Vergleich zum Durchschnitt der Merkmalsausprägungen an.

In Tabelle 4 wird eine einfache symbolische Notation für die Modelle und ihre geschätzten Koeffizienten benutzt. Es bedeutet “+” das Hinzufügen eines weiteren Haupteffekts und “*” die Bildung eines Interaktionstermes. Merkmal 1 (Merkmal 2 = c) bedeutet die Bildung eines konditionalen Haupteffekts für das Merkmal 1, wenn das Merkmal 2 den Wert c annimmt. Die geschätzten Koeffizienten werden unterhalb der Modellgleichung wiedergegeben. Die Notation folgt der Regel: Merkmal (Code 1: Koeffizient 1/ Code 2: Koeffizient 2/ ...

Anhand der Werte in Tabelle 4 können die geschätzten Ausfallraten durch Kontaktverlust leicht berechnet werden. Zum Beispiel erhält man: Welle 2, Stichprobe A, für einen Mehrpersonenhaushalt, der aus einer Großstadt (GROSS = 1) umgezogen ist (UMZUG = 2), den Logit-Wert $-2.87 + 0.24 + 0.11 = -2.52$. Damit ergibt sich $\text{Pr}(\text{Kontakt} = \text{nein}) = \frac{e^{-2.52}}{1 + e^{-2.52}} = 0.074$.

Tabelle 4: Die Schätzung der Wahrscheinlichkeit des Ausfalls eines Haushalts durch Kontaktverlust im SOEP. Koeffizienten der logistischen Regression der Kontaktermittlung von SOEP-Haushalten in der jeweiligen Befragungswelle. Darstellung: Merkmal (Code 1: Koeffizient 1/ Code 2: Koeffizient 2/ ...).

Stichprobe A (Deutsche-West)	
Welle	Modellvariablen und Koeffizienten
2	Modell = Konst + GROSS + UMZUG Konstante (-2.87), GROSS (0: -0.24/1: 0.24), UMZUG (1: -2.52/2: 0.11/3: 1.53/4: 0.84)
3	Modell = Konst + GROSS + UMZUG Konstante (-3.62), GROSS (0: -0.36/1: 0.36), UMZUG (1: -1.79/2: -0.49/3: 1.48/4: 0.80)
4	Modell = Konst + UMZUG Konstante (-3.42), UMZUG (1: -3.01/2: 0.78/3: 0.98/4: 1.35)
5	Modell = Konst + UMZUG + EINP(UMZUG) Konstante (-3.76), UMZUG (1: -3.09/2,3: 1.34/4: 1.75 EINP (UMZUG =1) (0: -1.35/1: 1.35) EINP (UMZUG =2,3) (0: -0.28/1: 0.28) EINP (UMZUG =4) (0: -0.63/1: 0.63)
6	Modell = Konst + UMZUG + EINP(UMZUG) Konstante (-3.48), UMZUG (1: -2.33/2,3: 0.64/4: 1.69) EINP (UMZUG =1) (0: -0.75/1: 0.75) EINP (UMZUG =2,3) (0: -0.76/1: 0.76) EINP (UMZUG =4) (0: -0.26/1: 0.26)
7*	Modell = Konst + GROSS + SPLIT Konstante (-2.97), GROSS (0: -0.39/1: 0.39), SPLIT (1: -1.10/2: -0.07/3: 1.17)
8	Modell = Konst + UMZUG Konstante (-5.03) UMZUG (1:-2.79/2: -0.24/3: 0.50/4: 2.53)
9	Pr(Kontakt=nein)= 0 falls UMZUG=1,2,3 / =0.06 falls UMZUG=4
10	Modell = Konst + GROSS + UMZUG Konstante (-4.44), GROSS (0: -0.44/1: 0.44), UMZUG (1: -3.65/2: 0.10/3: 1.12/4: 2.42)
11	Modell = Konst + EINP. + UMZUG Konstante (-6.01), EINP. (0: -1.06/1: 1.06) UMZUG (1: -0.99/2: -5.13/3: 1.84/4: 4.28)
* In Welle 7 wurden alle nicht umgezogenen Haushalte wieder kontaktiert, so daß die Ausfallanalyse nur auf den umgezogenen Haushalten basiert.	

Fortsetzung von Tabelle 4

Stichprobe B (Ausländer)	
2	Modell = Konst + GROSS + UMZUG + HHGR Konstante (-2.28), GROSS (0: -0.50/1: 0.50), UMZUG (1: -1.66/2: 0.69/3: -0.07/4: 1.04) HHGR (1: 1.23/2: 0.26/3: -0.82/4: -0.67)
3	Modell = Konst + GROSS + UMZUG Konstante (-2.65), GROSS (0: -0.72/1: 0.72), UMZUG (1: -3.06/2: 0.16/3: 1.64/4: 1.26)
4	Konstante (-3.34), UMZUG (1: -3.60/2: -0.46/3: 2.19/4: 1.87)
5	wie Stichprobe A
6	wie Stichprobe A
7*	Modell = Konst + GROSS + SPLIT + WUM Konstante (-2.93), GROSS (0: 0.064/1: -0.64), SPLIT (1: -1.65/2: 0.58/3: 1.07), WUM (1: -0.73/2: 1.32/3: -0.59)
8	wie Stichprobe A
9	Pr(Kontakt=nein)= 0 falls UMZUG=1,2,3 / =0.10 falls UMZUG=4
10	Modell = Konst + GROSS + UMZUG Konstante (-7.98), GROSS (0: -0.81/1: 0.81), UMZUG (1: -7.63/2: -4.69/3: 6.50/4: -5.82)
11	Modell = Konst + EINP. + UMZUG Konstante (-5.39), EINP. (0: -1.5/1: 1.54), UMZUG (1: -1.19/2: -4.26/3: 2.07/4: 3.39)
* In Welle 7 wurden alle nicht umgezogenen Haushalte wieder kontaktiert, so daß die Ausfallanalyse nur auf den umgezogenen Haushalten basiert.	
Stichprobe C (Deutsche-Ost)	
Welle	Modellvariablen und Koeffizienten
2 Ost	Pr(Kontakt=nein)= UMZUG(1: 0.0/ 2: 0.11/3: 0.14/4: 0.25)
3 Ost	Pr(Kontakt=nein)= UMZUG(1,2: 0.0/3: 0.09/4: 0.07)
4 Ost	Pr(Kontakt=nein)= UMZUG(1: 0.0/2: 0.04/3: 0.0/4: 0.14)
5 Ost	Pr(Kontakt=nein)= UMZUG(1: 0.0/2: 0.02/3: 0.03/4: 0.09)

5 Ausfälle durch Antwortverweigerung

Für die Haushalte mit Kontaktaufnahme werden zwei mögliche Ergebnisse betrachtet:

- es wird ein Haushalts-Interview realisiert.
- es wird kein Haushalts-Interview realisiert.

Hierbei wird nicht nach der Art der Ausfallgründe differenziert.

5.1 Ausfallraten nach verschiedenen Haushaltsmerkmalen

Die folgenden Tabellen dokumentieren die Ausfallraten bei wieder kontaktierten Haushalten nach unterschiedlichen Haushaltsmerkmalen. Die Merkmale beziehen sie sich in der Regel auf die vorhergehende Welle. Lediglich die felddingten Variablen orientieren sich an der aktuellen Befragungswelle.

Die personenbezogenen Merkmale beziehen sich auf den Haushaltsvorstand. Bei abgespaltenen Haushalten beziehen sich die personenbezogenen Merkmale auf die aus dem Panelhaushalt ausgezogene Person (bei mehreren Personen auf die im Adressprotokoll zuerst genannte Person).

Tabelle 5: : **Teilnahmeverhalten von erreichten Haushalten nach sozio-demographischen Merkmalen des Haushaltsvorstands.** N = Anzahl der zu befragenden Haushalte, % = Prozentsatz der Haushalte ohne Interview.

		Welle									
		2	3	4	5	6	7	8	9	10	11
Insgesamt	N	5937	5732	5398	5285	5095	4982	4985	4977	4994	4960
	%	10.4	11.2	6.9	8.9	7.9	6.9	6.3	6.7	6.6	7.3
Ausfall im Vorjahr	N	-	259	197	154	169	154	183	145	164	146
	%	-	59.5	52.8	71.6	57.4	49.3	48.1	62.1	50.6	54.1
Haushalte ohne Antwortausfall im Vorjahr											
insgesamt	N	5937	5473	5201	5131	4926	4828	4802	4832	4830	4814
	%	10.4	8.9	5.1	7.0	6.2	5.5	4.7	5.0	5.1	5.8
Stichprobe											
A (Deutsche)	N	4611	4275	4058	3993	3834	3755	3716	3724	3718	3713
	%	10.2	8.7	5.2	7.1	6.2	5.3	4.6	4.9	5.1	5.3
B (Ausländer)	N	1326	1198	1143	1138	1092	1073	1086	1108	1112	1101
	%	10.9	9.6	5.0	6.9	6.4	6.1	5.3	5.5	6.2	7.7
Geschlecht											
männlich	N	4664	4226	3951	3840	3624	3486	3413	3372	3340	3286
	%	9.8	8.3	4.7	6.7	6.2	5.0	4.2	4.9	4.5	5.5
weiblich	N	1273	1247	1250	1291	1303	1342	1389	1460	1490	1528
	%	12.2	11.0	6.5	7.9	6.5	6.9	6.1	5.3	6.3	6.6
Alter											
75 u. älter	N	448	394	374	386	381	380	371	367	353	340
	%	18.3	13.7	6.7	6.5	6.8	5.0	3.0	5.2	6.2	5.0
65-74	N	562	513	487	480	465	462	477	503	527	550
	%	10.1	9.4	2.9	5.4	6.0	3.0	2.9	3.2	2.9	4.2
55-64	N	947	860	809	803	798	783	782	811	821	832
	%	9.8	8.7	4.2	4.7	5.9	5.5	3.8	3.4	3.4	4.7
35-54	N	2621	2401	2272	2226	2112	2017	1970	1899	1851	1797
	%	9.2	7.8	4.3	6.3	6.1	4.9	4.7	4.1	3.6	5.8
25-34	N	1116	1034	976	963	904	926	957	983	1020	1077
	%	8.9	8.0	6.1	9.9	6.1	7.7	6.1	6.2	7.7	6.3
16-24	N	243	271	283	273	266	260	245	269	258	218
	%	17.3	15.1	13.4	13.2	9.0	8.1	8.9	14.8	13.2	13.3

Fortsetzung von Tabelle 5:

		Welle									
		2	3	4	5	6	7	8	9	10	11
Haushalte ohne Antwortausfall im Vorjahr											
Familienstand											
verheiratet, zusammen	N	3893	3600	3366	3301	3144	3029	3015	3008	2990	2949
	%	9.6	8.0	4.4	6.2	6.2	4.6	3.4	4.9	4.3	5.2
verheiratet, getrennt	N	104	157	119	97	120	110	96	102	102	106
	%	7.7	12.1	5.9	13.4	6.7	8.2	12.5	6.9	8.8	9.4
ledig	N	836	811	802	783	764	764	782	797	824	846
	%	12.6	9.6	9.4	11.5	6.8	8.6	9.1	9.7	8.3	7.9
geschieden	N	349	345	328	347	327	351	356	369	353	364
	%	10.3	10.4	4.6	6.3	7.0	7.1	4.8	4.9	3.4	6.0
verwitwet	N	671	560	533	542	534	532	518	515	523	514
	%	12.6	11.9	3.9	5.1	4.7	4.5	4.4	3.9	4.8	4.9
Schulbildung											
Kein Abschluß	N	493	445	411	407	376	373	380	379	380	379
	%	10.9	11.0	6.3	8.8	6.7	5.6	5.8	5.5	5.0	6.3
Hauptschule	N	2952	2669	2493	2488	2405	2340	2314	2296	2272	2240
	%	11.8	9.8	4.7	6.2	6.3	5.2	4.6	4.6	4.8	5.6
Realschule	N	852	849	818	805	798	780	784	812	835	846
	%	6.8	8.8	5.6	8.3	8.0	6.3	4.9	4.7	4.4	5.7
Fachhoch- schule	N	223	205	205	201	183	180	184	188	199	212
	%	9.4	6.8	6.3	9.5	5.5	5.6	7.6	4.8	5.0	7.1
Abitur	N	601	588	582	569	542	552	533	557	564	580
	%	7.5	6.0	5.5	7.6	4.2	7.4	5.1	7.0	6.2	5.5
Stellung im Beruf											
nicht erwerbs- tätig	N	1527	1325	1290	1302	1276	1329	1339	1279	1314	1320
	%	13.9	10.1	5.5	6.9	6.4	5.2	4.3	4.9	4.4	4.5
arbeitslos	N	206	297	260	258	265	193	199	215	197	239
	%	9.7	10.7	7.7	8.5	4.9	2.6	6.0	6.5	6.6	8.4
hoch	N	585	578	522	530	519	511	496	518	531	557
	%	7.9	6.2	4.4	7.7	7.1	8.4	5.9	5.6	4.7	6.3
mittel	N	2248	2202	2053	1982	1911	1803	1857	1932	1989	1855
	%	8.8	8.1	5.3	6.6	6.2	5.1	4.7	5.1	5.5	5.7
niedrig	N	1364	1071	1076	1059	954	992	911	888	799	843
	%	11.2	10.0	4.3	7.4	6.2	5.9	4.6	4.3	5.0	7.2

Tabelle 6: **Teilnahmeverhalten von erreichten Haushalten nach feldbedingten Merkmalen.** N = Anzahl der zu befragenden Haushalte, % = Prozentsatz der Haushalte ohne Interview

		Welle									
		2	3	4	5	6	7	8	9	10	11
Haushalte ohne Antwortausfall im Vorjahr											
Typ des Haushalts											
nicht umgezogener Haushalt	N	5372	4810	4646	4567	4421	4353	4313	4379	4285	4248
	%	9.6	8.0	4.2	5.8	5.6	4.6	3.6	3.9	4.0	5.8
umgezogener Haushalt	N	370	425	373	370	322	277	274	275	309	362
	%	11.6	12.7	10.7	14.1	9.0	7.6	9.9	9.5	8.4	7.7
abgespaltener Haushalt	N	195	238	182	194	183	198	215	208	235	204
	%	29.2	21.4	17.6	23.7	16.9	23.2	21.4	22.1	20.4	17.2
Wechsel des Interviewers											
ja	N	2041	1203	816	715	826	742	717	751	340	385
	%	14.9	17.5	12.5	19.0	12.9	14.4	10.7	12.1	8.2	8.6
nein	N	3896	4265	4385	4416	4100	4086	4085	4081	3879	3824
	%	7.9	6.5	3.8	5.1	4.9	3.9	3.7	3.7	3.3	3.4
Sonderfälle	N									611	605
	%									14.6	19.3
Anzahl der durch den HV bereits gewährten Interviews											
vollständig seit Beginn	N	-	5419	5018	4826	4600	4384	4225	4060	3856	3693
	%	-	8.7	4.7	5.9	5.6	4.4	3.6	3.3	3.4	3.9
1 Interview fehlt	N	-	-	161	246	253	294	346	389	399	416
	%	-	-	16.7	23.1	14.6	13.9	13.8	10.8	10.3	11.1
2 Interviews fehlen	N	-	-	-	46	43	73	93	127	163	173
	%	-	-	-	43.5	16.2	23.2	12.9	18.9	14.7	17.3
3 Interviews fehlen	N	-	-	-	-	24	49	63	104	137	164
	%	-	-	-	-	12.5	14.3	9.5	10.6	8.0	8.5
Neue Befragungsperson im Haushalt											
ja	N	257	243	218	211	209	220	198	210	197	197
	%	9.0	11.1	6.4	6.6	4.8	6.4	6.6	2.4	3.6	4.0
nein	N	5680	5230	4983	4920	4717	4608	4604	4622	4633	4616
	%	10.4	8.8	5.1	7.1	6.3	5.5	4.7	5.1	5.2	5.9
Auszug einer Befragungsperson aus dem Haushalt											
ja	N								209	243	201
	%								7.2	6.3	5.0
Kein Telefon im Haushalt											
trifft zu	N									248	253
	%									9.7	7.9

Fortsetzung von Tabelle 6:

		Welle											
		2	3	4	5	6	7	8	9	10	11		
Haushalte ohne Antwortausfall im Vorjahr													
		Haushalt mit Paartrennung											
insgesamt	N										94	116	103
	%										24.5	20.7	9.7
alter Haushalt	N										47	60	52
	%										14.6	16.7	7.7
neuer Haushalt	N										47	56	51
	%										34.8	25.0	11.7
Subjektive Merkmale													
		Allgemeine Lebenszufriedenheit											
eher unzufrieden (≤ 4)	N										302	380	
	%										7.3	7.6	
eher zufrieden (≥ 5)	N										4528	4434	
	%										4.9	5.7	

Tabelle 7: **Teilnahmeverhalten von erreichten Haushalten nach Haushaltseinkommen und Anzahl der Vermögensarten im Haushalt.** N = Anzahl der zu befragenden Personen, % = Prozentsatz der Haushalte ohne Interview

		Welle									
		2	3	4	5	6	7	8	9	10	11
		Haushalte ohne Antwortausfall im Vorjahr									
		Angaben zum HH-Einkommen fehlen im Vorjahr									
	N	335	310	272	237	210	203	197	226	193	220
	%	17.9	18.4	12.9	16.9	11.4	15.2	12.7	10.6	7.8	15.0
		Haushaltseinkommen im DM									
bis 1000	N	456	368	293	270	241	213	182	165	157	151
	%	13.8	10.1	5.5	9.3	7.1	6.1	6.6	8.5	7.0	6.6
1000-2000	N	1816	1521	1383	1243	1140	995	870	802	721	665
	%	11.2	9.7	5.3	6.6	6.2	4.8	5.1	4.5	5.0	7.2
2000-3000	N	1713	1572	1469	1404	1354	1329	1260	1202	1129	1063
	%	8.3	7.6	4.2	6.5	6.6	4.6	4.0	5.1	4.4	4.9
3000-4000	N	992	996	1008	1087	1060	1073	1069	1085	1103	1039
	%	9.6	6.4	4.3	6.1	5.0	4.7	3.7	4.4	5.4	4.9
4000 und mehr	N	625	706	776	890	921	1015	1224	1352	1527	1676
	%	8.2	8.9	5.0	6.4	5.9	6.2	4.7	4.4	4.8	5.2
		Anzahl der Vermögensarten im Haushalt									
0	N	823	769	743	735	578	661	573	604	567	563
	%	12.5	12.3	6.9	11.3	8.3	8.8	6.6	9.9	6.5	10.7
1	N	1714	1561	1468	1429	1431	1262	1256	1191	1140	1197
	%	13.2	10.4	5.1	6.8	6.7	5.5	4.4	4.5	5.3	5.7
2	N	1709	1549	1427	1449	1444	1310	1350	1367	1412	1277
	%	8.3	8.2	5.6	5.8	5.9	5.2	5.0	4.1	5.5	5.6
3	N	1224	1161	1134	1122	1107	1152	1201	1180	1210	1250
	%	8.7	6.7	3.9	5.9	5.7	4.1	3.7	4.2	4.3	4.1
4	N	403	388	374	343	326	377	375	447	451	476
	%	7.7	5.9	3.7	7.0	4.6	5.3	4.5	4.5	3.8	5.9
5	N	64	45	55	53	40	66	47	43	50	51
	%	9.4	6.7	9.1	11.3	2.5	7.5	14.9	4.7	0.0	3.9
		Bezug vom Sozialhilfe									
ja	N								133	133	135
	%								6.8	8.3	6.7

Tabelle 8: Vergleich der Ausfallraten zwischen SOEP-West und SOEP-Ost bis zur 4. Befragungswelle % = Prozentsatz der Haushalte ohne Interview.

Merkmal	Welle 2		Welle 3		Welle 4		Welle 5		
	West %	Ost %	West %	Ost %	West %	Ost %	West %	Ost %	
Erreichte HH insg.:	10.4	8.3	11.2	11.8	6.9	10.8	8.9	7.7	
davon Ausfall im Vorj.	-	-	59.4	53.1	52.8	67.1	71.6	50.0	
Kein Ausfall im Vorj.	-	-	8.9	8.6	5.1	7.2	7.0	6.0	
Haushalte ohne Antwortausfall im Vorjahr									
Alter des Haushaltsvorstands:									
75+	17.5	18.1	13.7	11.3	6.5	11.7	6.5	8.5	
65-74	9.9	8.0	9.5	7.3	2.9	6.0	5.4	3.4	
55-64	9.7	7.0	9.0	7.0	4.2	5.4	4.7	4.3	
35-54	9.3	6.3	7.7	6.8	4.3	6.6	6.3	5.3	
25-34	8.9	9.2	7.9	9.4	6.1	7.8	9.9	7.7	
-25	19.9	13.4	14.8	23.4	13.4	13.1	13.2	11.3	
Geschlecht des Haushaltsvorstands:									
männlich	9.8	7.6	8.2	8.9	4.7	8.1	6.7	6.2	
weiblich	12.4	8.9	11.4	8.3	6.5	6.4	7.9	5.8	
Beruflicher Status des Haushaltsvorstands:									
Nicht erwerbstätig	12.9	10.8	10.1	8.2	5.5	7.3	6.9	6.6	
Arbeitslos	9.8	14.3	10.8	9.4	7.7	8.3	8.5	4.6	
Höchste Statusgruppe	7.9	5.5	6.4	5.8	4.4	6.0	7.7	3.9	
Niedrigste Statusgruppe	10.6	9.4	10.0	9.5	4.3	8.6	6.6	9.0	
Sonstige	9.1	7.8	8.1	9.2	5.3	6.8	7.4	5.9	
Höchster Schulabschluß									
West	Ost								
Abitur	Abitur	7.5	7.1	6.0	7.6	5.5	7.3	7.6	3.7
mittl. Reife	10. Kl.	6.8	7.0	8.8	8.6	5.6	7.7	8.3	6.6
Hauptsch.	8. Kl.	11.8	9.9	9.8	8.8	4.7	6.4	6.2	5.9
K. Abschluß		10.9	-	11.0	-	6.3	-	8.8	-
Netto Haushaltseinkommen:									
Keine Angabe		17.9	9.4	18.4	17.5	12.9	20.0	16.9	17.2
West	Ost								
-1000	-800	13.8	14.3	10.1	8.7	5.5	7.8	9.3	10.0
1000-2000	800-1200	11.2	8.7	9.7	8.8	5.3	8.6	6.6	6.2
2000-3000	1200-1800	8.3	8.1	7.6	9.8	4.2	8.0	6.5	5.9
3000-4000	1800-2500	9.6	6.2	6.4	7.5	4.3	6.9	6.1	7.1
4000+	2500+	8.2	6.5	8.9	8.1	5.0	6.0	6.4	4.7

Fortsetzung von Tabelle 8

Merkmal	Welle 2		Welle 3		Welle 4		Welle 5	
	West %	Ost %	West %	Ost %	West %	Ost %	West %	Ost %
	Familienstand des Haushaltsvorstands							
Verheiratet zusammen	-	-	-	-	4.4	6.5	6.2	4.6
Verheiratet getrennt	-	-	-	-	5.9	26.7	13.4	10.7
ledig	-	-	-	-	9.4	12.2	11.5	9.2
geschieden	-	-	-	-	4.6	3.9	6.3	7.6
verwitwet	-	-	-	-	3.9	5.9	5.1	8.2
	Typ des Haushalts							
Alter HH, nicht umgez.	9.6	7.7	7.9	7.3	4.2	6.5	5.8	5.1
Alter HH, umgezogen	11.6	8.9	12.7	16.6	10.7	13.2	14.1	7.4
neuer Haushalt	29.2	24.6	21.4	23.8	17.6	18.6	23.7	23.5
	Wechsel des Interviewers							
ja	14.9	8.4	17.5	11.6	12.5	11.4	19.0	7.8
nein	7.9	5.5	6.5	7.4	3.8	5.5	5.1	4.4
Sonderfälle	-	13.6	-	36.4	-	26.2	-	25.2
	Allgemeine Lebenszufriedenheit							
eher unzufrieden (≤ 4)	-	-	-	-	-	8.7	-	7.3
eher zufrieden (≥ 5)	-	-	-	-	-	7.0	-	5.7
	Verlust Arbeitsplatz							
sicher od. wahrscheinlich	-	9.1	-	8.9	-	5.4	-	7.6
sonst	-	7.8	-	8.5	-	7.6	-	5.6
Haushalt lebt in alter BRD	-	15.4	-	15.8	-	9.6	-	10.0
Bezug von Sozialhilfe	-	-	-	-	-	10.3	-	4.6
Kein Telephon im Haushalt	-	-	-	-	-	7.9	-	6.6

5.2 Definition der Modellvariablen

Die in dem vorhergehenden Abschnitt aufgeführten Merkmale wurden bis auf wenige Ausnahmen in eine simultane Analyse der Ausfälle durch Antwortverweigerung einbezogen. Für die Schätzung der entsprechenden Ausfallwahrscheinlichkeit wurde jedoch nur ein Modell benutzt, in dem alle Variablen signifikant sind. Hieraus kann geschlossen werden, daß die weggelassenen Variablen bzw. Differenzierungen zwischen Merkmalsausprägungen keinen systematischen Einfluß auf die Antwortgewährung erreicher Haushalte haben. Die Definition der Modell-Variablen findet man in der folgenden Liste:

Merkmals	Bezeichnung	Code	Ausprägungen
Alter des Haushaltsvorstands	ALTHV	1	älter als 75 Jahre
		2	65-74 Jahre
		3	55-64 Jahre
		4	35-54 Jahre
		5	25-34 Jahre
		6	jünger als 25 Jahre
Geschlecht des Haushaltsvorstands	SEX	0	männlich
		1	weiblich
Haushalts Typ	HTYP	1	alter Haushalt, nicht umgezogen
		2	alter Haushalt, umgezogen
		3	abgespaltener Haushalt
Wechsel des Inter- viewers	INTW	0	Kein Wechsel
		1	Wechsel gegenüber letzter Welle
		2	Keine reguläre Interviewernr.
Betreuung durch Interviewer	BETREUUNG		Anzahl der Wellen mit Betreuung durch den aktuellen Interviewer
Beginn	BEGINN	0	sonst
		1	HV seit Welle 1 im Panel
Auszug	AUSZUG	0	sonst
		1	Auszug einer Befragungsperson seit letzter Welle
Paar	PAAR	0	sonst
		1	Auszug des HV od. des Ehepartners (Lebenspartners) des HV
Typ	TYP	0	HH Typ =1,2 und Paar=1
		1	HH Typ =1 und Paar=0
		2	HH Typ =2 und Paar=0
		3	HH Typ =3 und Paar=0
		4	HH Typ =3 und Paar=1
Ost-Berlin	OSTB	0	sonst
		1	Haushalt lebt in Ostberlin
Familienstand	FAMSTD	1	verheiratet, zusammenlebend
		2	verheiratet, getrennt
		3	ledig
		4	geschieden
		5	verwitwet

Fortsetzung der Merkmalsliste

Merkmal	Bezeichnung	Code	Ausprägungen
Arbeitslos	ALOS	0	sonst
		1	HV ist arbeitslos gemeldet
Verlust Arbeitsplatz (subjektive Einschätzung)	VERLUST	0	sonst
		1	sicher od. wahrscheinlich
Status des Haushaltsvorstands	STATUSH	0	sonst
		1	hoher Erwerbsstatus des HV
Sozialhilfe	SOZH	0	sonst
		1	Haushalt bezieht Sozialhilfe
Haushalts- Einkommen West	EINKW	1	K.A. beim Einkommen
		2	bis 2000 DM
		3	2000-4000 DM
		4	über 4000 DM
Haushalts- Einkommen Ost	EINKO	1	K.A. beim Einkommen
		2	bis 800 DM
		3	800-1200 DM
		4	1200-1800 DM
		5	1800-2500 DM
		6	über 2500 DM
K.A.-Einkommen	KAEINK	0	sonst
		1	K.A. beim Einkommen
K.A.-Vermögensbilanz	KAVB	0	sonst
		1	Vermögensbilanz in Welle 5 nicht beantwortet
Anzahl Anlageformen	ANZAHL	1	Anzahl der Vermögens- anlageformen = 0
		2	Anzahl = 5
		3	sonst
Keine Anlageformen	ANZO	0	sonst
		1	Anzahl der Vermögens- anlageformen = 0
Betriebsvermögen	BETRIEB	0	sonst
		1	Besitz von Betriebsvermögen
Sparbuch als Wertanlage angegeben	SPAR	0	nein
		1	ja

5.3 Geschätzte Wahrscheinlichkeiten für einen Ausfall durch Antwortverweigerung erreichter Haushalte

Die im vorhergehenden Abschnitt definierten Merkmale wurden im Rahmen einer multiplen Logit-Analyse zur Schätzung von $P_R = \Pr(\text{Response}=\text{nein})$ benutzt. Für die Berechnung der SOEP GewichtungsvARIABLEN wurden nur Spezifikationen benutzt, wo alle Kovariaten einen signifikanten Einfluß haben.

Interpretation der Koeffizienten:

$$\ln \frac{P_{R,i}}{1 - P_{R,i}} = \text{const} + X_i' \beta$$

Positive Koeffizienten zeigen eine Vergrößerung des Ausfallrisikos im Vergleich zum Durchschnitt der Merkmalsausprägungen an.

In Tabelle 9 wird eine einfache symbolische Notation für die Modelle und ihre geschätzten Koeffizienten benutzt. Es bedeutet “+” das Hinzufügen eines weiteren Haupteffekts und “*” die Bildung eines Interaktionstermes. Merkmal 1 (Merkmal 2 = c) bedeutet die Bildung eines konditionalen Haupteffekts für das Merkmal 1, wenn das Merkmal 2 den Wert c annimmt. Die geschätzten Koeffizienten werden unter der Modellgleichung aufgeführt. Die Notation folgt der Regel: Merkmal (Code 1: Koeffizient 1/ Code 2: Koeffizient 2/ ...).

Die geschätzten Ausfallraten können leicht anhand der Koeffizienten von Tabelle 9 ermittelt werden. Beispielsweise erhalten wir in Welle 2 für Stichprobe A für einen Haushalt ohne Interviewerwechsel (INTW = 0), mit Alter des HV zwischen 35 und 74 Jahren (ALTHV = 2,3,4) und einem Haushaltseinkommen unter 2000 DM (EINKW = 2), der nicht umgezogen ist (HTYP = 1) den Logit-Wert $-1,53 - 0,25 + 0,03 - 0,68 + 0,12 = -2,31$. Damit erhält man $P(\text{Response} = \text{nein}) = \frac{e^{-2,31}}{1 + e^{-2,31}} = 0,09$.

Tabelle 9: Die Schätzung der Wahrscheinlichkeit eines Ausfalls durch Antwortverweigerung eines wiedererreichten Haushalts im SOEP. Koeffizienten der logistischen Regression der Antwortgewährung von wieder erreichten Haushalten in der jeweiligen Befragungswelle. Darstellung: Merkmal (Code 1: Koeffizient 1/ Code 2: Koeffizient 2/ ...)

Stichprobe A (Deutsche-West)	
Welle	Modellvariablen und Koeffizienten
2	Modell = KONST + INTW + ALTHV + HTYP + EINKW KONST (-1.53), INTW (0: -0.25/1: 0.25), ALTHV (1: 0.66/2,3,4: 0.03/5: -0.39/6: -0.30), HTYP (1: -0.68/2: -0.19/3: 0.87), EINKW (1: 0.61/2: 0.12/3: -0.35/4: -0.38)
3	Modell = KONST + INTW + ALTHV + INTW*ALTHV + HTYP + ALOS + KAEINK KONST (-1.22), INTW (0: -0.39/1: 0.39), ALTHV* (INTW =0) (1: -0.13/2: -0.11/3,4: -0.39/5: 0.26/6: 0.37), ALTHV* (INTW =1) (1: 0.13/2: 0.11/3,4: 0.39/5: -0.26/6: -0.37), ALTHV (1: 0.59/2: 0.16/3,4: -0.06/5: -0.53/6: -0.16) HTYP (1: -0.52/2: 0.10/3:0.42), ALOS (0: -0,21/1: 0.21), KAEINK (0: -0.39/1: 0.39)
4	Modell = KONST + ALTHV + INTW(ALTHV) + HTYP + KAEINK KONST (-1.83), INTW (ALTHV =1) (0: -0.44/1: 0.44), INTW (ALTHV =2) (0: -0.74/1: 0.74), INTW (ALTHV =3,4) (0: -0.59/1: 0.59), INTW (ALTHV =5) (0: -0.41/1: 0.41), INTW (ALTHV =6) (0: -0.32/1: 0.32), ALTHV (1: 0.21/2: -0.38/3,4: -0.24/5: 0.06/6: 0.35), HTYP (1: -0.45/2: 0.29/3: 0.19), KAEINK (0: -0.39/1: 0.39)
5	Modell = KONST + BETREUUNG + ALTHV(INTW=1) + HTYP + KAEINK + ANZO KONST (-1.60), BETREUUNG (1: 1.15/ 2: 0.41/3: 0.18/4: -0.71/5: -1.03), ALTHV (INTW =1) (1,2: 0.52/3,4,5: -0.11/6: -0.40), HTYP (1: -0.49/2: 0.11/3: 0.38) KAEINK (0: -0.45/1: 0.45), ANZO (0: -0.38/1: 0.38)

Fortsetzung von Tabelle 9:

Welle	Modellvariablen und Koeffizienten
6	<p>Modell = KONST + BETREUUNG + ALTHV(INTW=1) + HTYP + KAEINK + KAVB + BETRIEB</p> <p>KONST (-2.44), BETREUUNG (1: 0.75/2: 0.58/3: 0.21/4: -0.59/5: -0.43/6: -0.52), ALTHV (INTW =1) (1,2: 0.26/3,4,5: 0.05/6: -0.31), HTYP (1: -0.32/2: -0.04/3: 0.37), KAEINK (0: -0.26/1: 0.26), KAVB (0: -0.29/1: 0.29), BETRIEB (0: 0.41/1: -0.41)</p>
7	<p>Modell = KONST + HTYP + INTW(HTYP) + KAEINK + STATUSH</p> <p>KONST (-1.34), INTW (HTYP =1) (0: -0.75/1: 0.75), INTW (HTYP =2) (0: -0.56/1: 0.56), INTW (HTYP =3) (0: -0.12/1: 0.12), HTYP (1: -0.66/2: -0.24/3: 0.90), KAEINK (0: -0.58/1: 0.58), STATUSH (0: -0.30/1: 0.30)</p>
8	<p>Modell = KONST + INTW + HTYP + KAEINK + ANZAHL</p> <p>KONST (-1.15); INTW (0: -0.55/1: 0.55), HTYP (1: -0.83/2: -0.14/3: 0.97), KAEINK (0: -0.57/1: 0.57), ANZAHL (1: -0.08/2: 0.70/3: -0.62)</p>
9	<p>Modell = KONST + INTW(BEGINN) + BEGINN(ALTHV) + HTYP + AUSZUG(HTYP=1) + KAEINK + ANZO + SEX</p> <p>KONST (-1.31), INTW(BEGINN =0) (0: -0.17/1: 0.17), INTW(BEGINN =1) (0: -0.68/1: 0.68), BEGINN(ALTHV=1) (0: -0.09/1: 0.09), BEGINN(ALTHV=2) (0: 0.70/1: -0.70), BEGINN(ALTHV=3) (0: 1.20/1: -1.20), BEGINN(ALTHV=4) (0: 0.49/1: -0.49), BEGINN(ALTHV=5) (0: 0.48/1: -0.48), BEGINN(ALTHV=6) (0: 0.10/1: -0.10), HTYP (1: -0.53/2: 0.07/3: 0.46), AUSZUG(HTYP =1) (0: -0.47/1: 0.47) KAEINK (0: -0.25/1: 0.25) ANZO (0: -0.29/1: 0.29) SEX (0: 0.15/1: -0.15)</p>

Fortsetzung von Tabelle 9:

Welle	Modellvariablen und Koeffizienten
10	<p>Modell = KONST + HTYP + BEGINN(HTYP) + INTW(HTYP) + PAAR(HTYP=1) + ALTHV(HTYP=1) KONST (-1.89), HTYP (1: -0.12/2: -0.39/3: 0.51), INTW(HTYP=1) (0: -0.95/1: -0.08/2: 0.88), INTW(HTYP=2) (0: -0.24/1: -0.06/2: 0.30), INTW(HTYP=3) (0: 0.16/1: -0.47/2: 0.31), BEGINN(HTYP=1) (0: 0.43/1: -0.43), BEGINN(HTYP=2) (0: 0.21/1: -0.21), BEGINN(HTYP=3) (0: -0.07/1: 0.07), PAAR(HTYP=1) (0: -0.58/1: 0.58), ALTHV(HTYP=1) (1: 0.41/2: -0.26/3: -0.08/4: -0.50/5: 0.01/6: 0.42)</p>
11	<p>Modell = KONST + HTYP + BEGINN + INTW + KAEINK + TELEPHON(INTW=1) KONST (-1.68), HTYP (1: -0.39/2: -0.09/3: 0.48), BEGINN (0: 0.27/1: -0.27), INTW (0: -0.63/1: -0.10/2: 0.73) KAEINK (0: -0.35/1: 0.35) TELEPHON (INTW=1) (0: 0.49/1: -0.49)</p>

Fortsetzung von Tabelle 9:

Stichprobe B (Ausländer)	
Welle	Modellvariablen und Koeffizienten
2	Modell = KONST + INTW + HTYP KONST (-1.96), INTW (0: -0.55/1: 0.55), HTYP (1: -0.03/2: -0.58/3: 0.62)
3	Modell = KONST + SEX + HTYP KONST (-1.60), SEX (0: -0.31/1: 0.31), HTYP (1,2: -0.46/3: 0.46)
4	Modell = KONST + INTW(ALTHV) + HTYP + EINKW KONST (-1.69), INTW (ALTHV =1,2,3) (0: -0.47/1: 0.47), INTW (ALTHV =4) (0: -0.73/1: 0.73), INTW (ALTHV =5) (0: -0.60/1: 0.60), INTW (ALTHV =6) (0: -0.26/1: 0.26), HTYP (1: -0.34/2: 0.46/3: -0.12), EINKW (1: 0.75/2: 0.10/3: -0.85)
5	Modell = KONST + BETREUUNG + HTYP + KAEINK KONST (-1.87), BETREUUNG (1: 1.26/ 2: 0.14/3: -0.21/4: -0.70/5: -0.50), HTYP (1: -0.47/2: 0.89/3: -0.42), KAEINK (0: -0.43/1: 0.43)
6	Modell = KONST + BETREUUNG + HTYP + KAEINK KONST (-1.89), BETREUUNG (1: 0.83/ 2: 0.37/3: -0.31/4: -0.55/5: 0.04/ 6: -0.37), HTYP (1: -0.41/2: 0.22/3: 0.19), KAEINK (0: -0.54/1: 0.54)
7	Modell = KONST + HTYP + INTW(HTYP) + KAEINK KONST (-1.50), INTW (HTYP =1) (0: -0.55/1: 0.55), INTW (HTYP =2) (0: -0.98/1: 0.98), INTW (HTYP =3) (0: -1.06/1: 1.06), HTYP (0: -0.50/2: -0.88/3: 1.38), KAEINK (0: -0.66/1: 0.66)
8	Modell = KONST + INTW + HTYP KONST (-2.05), INTW (0: -0.48/1: 0.48), HTYP (1: -0.85/2: 0.22/3: 0.63)

Fortsetzung von Tabelle 9:

9	<p>Modell = KONST + INTW + BEGINN + TYP + ALTHV + KAEINK + ANZO + SOZH</p> <p>KONST (-1.79), INTW (0: -0.50/1: 0.50), BEGINN (0: 0.39/1: -0.39), TYP (0: 0.16/1: -0.59/2: -1.90/3: -0.03/4: 2.36), ALTHV (1,2,3: 0.28/4: -0.10/5: -0.65/6: 0.47), KAEINK (0: -0.66/1: 0.66), ANZO (0: -0.53/1: 0.53), SOZH (0: 0.73/1: -0.73)</p>
10	<p>Modell = KONST + HTYP + PAAR + ALTHV + INTW(ALTHV)</p> <p>KONST (-1.58), HTYP (1: -0.44/2: -0.11/3: 0.55), PAAR (0: -0.63/1: 0.63), ALTHV(1,2,3: -0.79/4: -0.04/5: 0.77/6: -0.05), INTW(ALTHV=4) (0: -1.11/1: -0.10/2: 1.21), INTW(ALTHV=5) (0: -0.79/1: -0.22/2: 1.01),</p>
11	<p>Modell = KONST + BEGINN + HTYP + INTW ANZO + FAMSTD</p> <p>KONST (-1.43), INTW (0: -0.69/1: 0.01/2: 0.70) BEGINN (0: 0.33/1: -0.33) HTYP (1/2: -0.48/3: 0.48) ANZO (0: -0.31/1: 0.31) FAMSTD (1: 0.25/2,3,4,5: -0.25)</p>

Fortsetzung von Tabelle 9

Stichprobe C (Deutsche-Ost)	
Welle	Modellvariablen und Koeffizienten
2 Ost	<p>Modell = KONST + HTYP + INTW + ALTHV + EINKO + VERLUST + OSTB</p> <p>KONST (-0.91), INTW (0: -0.47/1: -0.04/2: 0.51), ALTHV (1: 0.41/2,3,4,5,6: -0.41), HTYP (1,2: -0.84/3: 0.84), EINKO (1: 0.24/2: 0.44/3: 0.12/4: 0.00/5: -0.37/6: -0.44), VERLUST (0: -0.17/1: 0.17), OSTB (0: -0.29/1: 0.29)</p>
3 Ost	<p>Modell = KONST + HTYP + INTW(HTYP) + ALTHV + SPAR</p> <p>KONST (-1.36), HTYP (1: -0.39/2: 0.08/3: 0.31), INTW (HTYP =1) (0: -0.28/1,2: 0.28), INTW (HTYP =2) (0: 0.42/1,2: -0.42), INTW (HTYP =3) (0: -0.36/1,2: 0.36), ALTHV (1: 0.02/2,3,4: -0.38/5: -0.20/6: 0.56), SPAR (0: 0.35/1: -0.35)</p>
4 Ost	<p>Modell = KONST + HTYP + INTW + ALTHV + KAEINK + FAMSTD</p> <p>KONST (-0.62), HTYP (1: -0.47/2: 0.25/3: 0.12) INTW (0: -0.78/1: -0.04/2: 0.82), ALTHV (1: 0.47/2,3,4,5,6: -0.47), KAEINK (0: -0.54/1: 0.54), FAMSTD (1: -0.12/2: 1.13/3: 0.24/4: -0.73/5: -0.51)</p>
5 Ost	<p>Modell = KONST + HTYP + INTW + KAEINK + VANZAHL + VERLUST</p> <p>KONST (-0.82), HTYP (1: -0.45/2: -0.32/3: 0.77), INTW (0: -0.67/1: -0.18/2: 0.84), KAEINK (0: -0.49/ 1: 0.49), VANZAHL (0: 0.0/1: -0.31/2: -0.62/3: -0.93/4: -1.24/5: -1.51), VERLUST (0: -0.19/1: 0.19)</p>

6 Literaturhinweise

- Arbeitsgemeinschaft ADM-Stichproben und Bureau Wendt* 1994: Das ADM-Stichproben-System (Stand 1993). In: Gabler, S.; Hoffmeyer-Zlotnik, J.; Krebs, D. (Hg): Gewichtung in der Umfragepraxis. Westdeutscher Verlag, Opladen, S. 188-203.
- Kirschner, Hans-Peter* 1984: ALLBUS 1980: Stichprobenplan und Gewichtung. In: Mayer, K.-U.; Schmidt, P. (Hrsg): Allgemeine Bevölkerungsumfrage Sozialwissenschaften, Frankfurt/M., S. 114-182.
- Projektgruppe "Das Sozio-ökonomische Panel"* 1990: Das Sozio-ökonomische Panel für die Bundesrepublik Deutschland nach fünf Wellen. Vierteljahrshefte zur Wirtschaftsforschung, Heft 2, S. 141-151.
- Projektgruppe "Das Sozio-ökonomische Panel"* 1991: Das Sozio-ökonomische Panel (SOEP) im Jahre 1990/91. Vierteljahrshefte zur Wirtschaftsforschung, Heft 3/4, S. 146-155.
- Projektgruppe "Das Sozio-ökonomische Panel"* 1993: 10 Jahre Sozio-ökonomisches Panel (SOEP). Vierteljahrshefte zur Wirtschaftsforschung, Heft 1/2, S. 141-151.
- Projektgruppe "Das Sozio-ökonomische Panel"* 1995: Das Sozio-ökonomische Panel (SOEP) im Jahre 1994. Vierteljahrshefte zur Wirtschaftsforschung, Heft 1.
- Pischner, Rainer* 1994: Die Quer- und Längsschnittgewichtung des Sozio-ökonomischen Panels (SOEP). In: Gabler, S.; Hoffmeyer-Zlotnik, J.; Krebs, D. (Hg): Gewichtung in der Umfragepraxis. Westdeutscher Verlag, Opladen, S. 166-187.
- Rendtel, Ulrich* 1990: Teilnahmebereitschaft in Panelstudien: Zwischen Beeinflussung, Vertrauen und Sozialer Selektion. Kölner Zeitschrift für Soziologie und Sozialpsychologie, 42, S. 280-299.
- Rendtel, Ulrich* 1991: Die Schätzung von Populationswerten in Panelerhebungen. Allgemeines Statistisches Archiv, 75, S. 225-244.
- Rendtel, Ulrich* 1995: Panelausfälle und Panelrepräsentativität. Erscheint im Campus Verlag, Frankfurt/M.- New York.
- Schupp, Jürgen und Gert Wagner* 1995: Die Zuwandererstichprobe des Sozio-ökonomischen Panels (SOEP). Vierteljahrshefte zur Wirtschaftsforschung, Heft 1.

Wagner, Gert; Schupp, Jürgen und Ulrich Rendtel 1993: Das Sozio-ökonomische Panel - Methoden der Datenproduktion und -aufbereitung im Längsschnitt. In: Hauser, R.; Ott N.; Wagner, G. (Hrsg.): Mikroanalytische Grundlagen der Gesellschaftspolitik - Band 2: Erhebungsverfahren, Analysemethoden und Mikrosimulation. Akademie Verlag, Berlin, S. 70-112.

Rendtel, Ulrich 1991: Die Schätzung von Populationswerten in Panelerhebungen. Allgemeines Statistisches Archiv, 75,S. 225-244.