

Rendtel, Ulrich

Working Paper — Digitized Version

Über die Repräsentativität von Panelstichproben: Eine Analyse der feldbedingten Ausfälle im Sozio- Oekonomischen Panel (SOEP)

DIW Discussion Papers, No. 70

Provided in Cooperation with:

German Institute for Economic Research (DIW Berlin)

Suggested Citation: Rendtel, Ulrich (1993) : Über die Repräsentativität von Panelstichproben: Eine Analyse der feldbedingten Ausfälle im Sozio-Oekonomischen Panel (SOEP), DIW Discussion Papers, No. 70, Deutsches Institut für Wirtschaftsforschung (DIW), Berlin

This Version is available at:

<https://hdl.handle.net/10419/95805>

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. 70

**Über die Repräsentativität von Panelstichproben
Eine Analyse der feldbedingten Ausfälle im
Sozio-oekonomischen Panel (SOEP)**

von
Ulrich Rendtel*

*) Deutsches Institut für Wirtschaftsforschung (DIW), Berlin

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 views of the Institute.

Diskussionspapier Nr. 70

**Über die Repräsentativität von Panelstichproben
Eine Analyse der feldbedingten Ausfälle im
Sozio-oekonomischen Panel (SOEP)**

von
Ulrich Rendtel

*) Deutsches Institut für Wirtschaftsforschung (DIW), Berlin

Berlin, im April 1993

Über die Repräsentativität von Panelstichproben

Eine Analyse der felddingten Ausfälle im Sozio-ökonomischen Panel (SOEP)

Ulrich Rendtel

Deutsches Institut für Wirtschaftsforschung
Berlin

April 1993

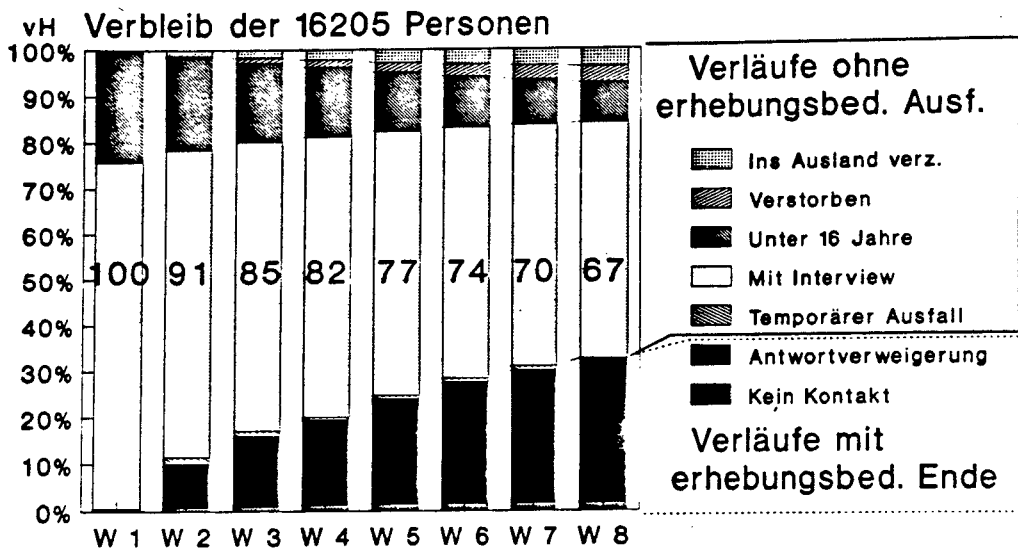
1 Einleitung

Ein wichtiges Argument, das gegen Panelerhebungen vorgebracht werden kann, lautet, daß die felddingten Ausfälle während des Panels nicht nur die Fallzahlen verringern, sondern auch "selektiv" sein können und so die "Repräsentativität der Stichprobe" zerstören. Hierbei werden Vermutungen geäußert, die an die klassischen Hypothesen über "Mittelstands-Bias" und "Bildungs-Bias" bei Querschnittssurveys (vgl. hierzu z.B. Hartmann/Schimpl-Neimanns 1992) angelehnt sind. Danach ist die Teilnahmebereitschaft bei Angehörigen der "Mittelschicht" in Umfragen größer als bei Angehörigen der "Unter-" bzw. "Oberschicht". Die Hypothese vom "Bildungs-Bias" besagt, daß die Antwortneigung mit dem Niveau des Schulabschlusses zunimmt. Daneben können jedoch auch Panel-spezifische Hypothesen über die Teilnahmebereitschaft aufgestellt werden. Diese Hypothesen sind ereignisorientiert. Beispielsweise könnte die Hypothese aufgestellt werden, daß der Eintritt einer Arbeitslosigkeit oder die Trennung einer Partnerschaft die weitere Teilnahmebereitschaft am Panel senkt. A-priori klingen derartige Hypothesen durchaus plausibel.

Auch wenn der Umfang der felddingten Ausfälle im Sozio-ökonomischen Panel (SOEP) von Welle zu Welle gering ist, kann doch nicht übersehen werden, daß der kumulative Effekt über alle Panelwellen beträchtlich sein kann. Abbildung 1 zeigt diese Entwicklung für das SOEP. Bis Welle 8 endeten ein Drittel aller Verläufe von Personen aus der ersten Welle erhebungsbedingt. Der kumulative Umfang dieser Ausfälle hat damit einen Umfang erreicht, der den Ausfällen bei

üblichen Querschnittbefragungen entspricht. Auch bei Start des SOEP (1. Welle) Betragen die Ausfälle etwa 30% aller zufällig ausgewählten Haushalte.

Entwicklung Stichprobenbestand des SOEP Basis : Alle Stammpersonen der 1. Welle



Datenbasis : Das Sozio-ökonomische Panel
Welle 1 bis Welle 8

Repräsentativitätsstudien sind für Panelerhebungen bisher kaum durchgeführt worden. Dies liegt im wesentlichen daran, daß bisher nur wenige Panelstudien mit hoher Fallzahl und langer Laufzeit existieren, die den Anspruch erheben, für die Bevölkerung im Erhebungsgebiet "repräsentativ" zu sein. Für Haushaltspansels liegen entsprechende Untersuchungen nur für die 1968 gestartete PSID vor (Beckett et al. (1983, 1988), Duncan/Hill (1989)).

Die vorliegende Arbeit setzt sich deswegen zunächst unter methodischen Gesichtspunkten mit dem Begriff der "Repräsentativität" von Panelstichproben auseinander. Hierbei steht der Begriff "Repräsentative Stichprobe" für ein nur vage umschriebenes Konzept der Proportionalität von Merkmalsverteilungen in Stichprobe und Grundgesamtheit (Kruskal/Mosteller 1979 a, b, c). Der Kern der eingangs beschriebenen Fragestellung richtet sich jedoch nicht auf diese Proportionalität, sondern auf die Zuverlässigkeit der Schätzergebnisse, die mit einem Panel gewonnen werden. Diese Arbeit beschäftigt sich dabei im wesentlichen mit Populationsschätzungen, d.h. der Schätzung von Merkmalsverteilungen in der Grundgesamtheit.

Die klassischen Populationsschätzer basieren auf einer Gewichtung der Stich-

probenergebnisse durch den Kehrwert der Auswahlwahrscheinlichkeit einer Einheit. Dieses Schätzverfahren kann für Panelerhebungen adaptiert werden, vgl. Rendtel (1991, 1993). Damit können über ein Panel erwartungstreue Populationsschätzungen geliefert werden, wenn die Auswahlwahrscheinlichkeiten für die Elemente der Panelstichprobe bekannt sind. Hierzu müssen die Wahrscheinlichkeiten einer Antwortgewährung in den einzelnen Panelwellen bekannt sein. Eine Analyse der Stichprobenausfälle während des Panels ist also eine notwendige Voraussetzung für erwartungstreue Populationsschätzungen.

Das Konzept der Proportionalität von Stichprobe und Grundgesamtheit besitzt zwar gewisse Beziehungen zu einer Ausfallanalyse. So kann beispielsweise das Verhältnis von Stichproben – zu Grundgesamtheitswerten als indirekte, auf Modellannahmen basierende Schätzung der Antwortwahrscheinlichkeiten aufgefaßt werden. In einem Panel steht jedoch vielfach der direkte Weg zur Verfügung: Die Schätzung von Ausfallprozessen auf Basis von Informationen über Respondenten und Nonrespondenten.

Weiterhin zeigt sich, daß in einem Panel das Repräsentativitätskonzept nicht mehr deckungsgleich mit dem Konzept erwartungstreuer Populationsschätzer ist. D.h., es kann der Fall eintreten, daß eine Panelstichprobe nicht mehr dem Konzept der Proportionalität von Stichprobe und Grundgesamtheit genügt und trotzdem eine zuverlässige Basis für erwartungstreue Schätzer liefert.

Die Studie von Duncan/Hill (1989) folgt dem Schema klassischer Repräsentativitätsstudien: Sie vergleicht die Stichprobenverteilung hinsichtlich Alter, Geschlecht, Hautfarbe, Schulabschluß, Familieneinkommen, Arbeitseinkommen und Familiengröße mit der entsprechenden Verteilung im amerikanischen Current Population Survey (CPS). Dieser Vergleich wird für die Startwelle und die 15. Erhebungswelle durchgeführt. Für beide Befragungswellen werden signifikante Unterschiede hinsichtlich Alter, Hautfarbe und Familieneinkommen ausgewiesen, jedoch richtet sich die Schlußweise mehr auf die Veränderung der Unterschiede bzw. auf die Verringerung der χ^2 -Werte eines Tests¹ auf Gleichheit der Verteilungen. Die Studie kommt zu dem Schluß, daß sich diese Abstandsmaße im Verlaufe des Panels nicht vergrößert haben. Es bleibt allerdings unklar, was aus dieser Tatsache für die Eigenschaften von Populationsschätzungen mit der PSID gefolgert werden kann.

Die Vorgehensweise von Beckett et al. (1983, 1988) ist hiervon grundsätzlich verschieden. Sie modellieren die Dauer der Teilnahme an der PSID durch ein Hazard-Raten Modell. Hierbei wird jedoch nicht unterschieden, ob der Ausfall aus dem Panel durch natürliche Lebensereignisse (z.B. Tod) oder durch feldbedingte Ereignisse hervorgerufen wurde.

Diese Studie rückt damit von dem Proportionalitätskonzept ab und stellt den Verbleib von Stichprobenmitgliedern im Panel in den Vordergrund. Aller-

¹Die Berechnung der Teststatistik wird nicht näher ausgewiesen. Man findet nur den Hinweis, daß die Teststatistik auf "Complex sampling variances" beruht.

dings sind für die Berechnung der Antwortwahrscheinlichkeiten die systematischen Ausfälle durch Tod und Fortzug ins Ausland irrelevant, da nur die feldbedingten Ausfälle bei der Analyse der Antwortwahrscheinlichkeiten berücksichtigt werden dürfen. Dies schränkt die Aussagekraft der Studie von Beckett et al. ein. Zudem wird die Untersuchung auf Personenebene durchgeführt und ignoriert dabei die zu erwartenden Abhängigkeiten auf Haushaltsebene. Als Resultat erhalten Beckett et al., daß Personen mit niedrigen und hohen Einkommen die PSID eher verlassen als Personen mit mittleren Einkommen. Der Bezug von Sozialhilfe in der letzten Interviewperiode erhöht das Ausfallrisiko. Die geschätzten Effekte sind jedoch klein im Vergleich mit den Effekten der Familienbeziehungen. Danach scheiden Personen ohne Familienbindung mit sehr hoher Wahrscheinlichkeit aus der PSID aus. Dieses Ergebnis reflektiert die Weiterverfolgungsregeln der PSID, nach denen Nicht-Familienmitglieder² bei einem Auszug aus dem Haushalt nicht weiterverfolgt werden³.

In der vorliegenden Arbeit wird zunächst das klassische Repräsentativitätskonzept für Querschnittsurveys kritisch überprüft. Daran anschließend werden Möglichkeiten aufgezeigt, die Auswahlwahrscheinlichkeiten auf den verschiedenen Befragungsstufen des Panels zu schätzen. Es zeigt sich dabei, daß nur im Falle von Veränderungsmerkmalen eine Notwendigkeit zu Vergleichen mit Populationsmerkmalen besteht. Daran anschließend werden Hypothesen formuliert, welche Einflußfaktoren das Ausfallverhalten im SOEP bestimmen. Hierbei wird nach den Stufen "Kontaktaufnahme" und "Antwortgewährung nach Kontaktaufnahme" unterschieden. Diese Hypothesen werden für die ersten 8 Wellen des SOEP überprüft.

Für der Veränderungsmerkmale "Geburt", "Tod" und "Ehescheidung" kann diese direkte Überprüfung nicht durchgeführt werden, da für die ausgefallenen Haushalte diese Merkmale nicht bekannt sind. Andererseits sind gehören diese Ereignisse zu den wenigen Veränderungsmerkmalen, über die verlässliche Informationen in der Grundgesamtheit vorliegen. Anhand eines Vergleichs der geschätzten Ereignisse mit den amtlichen Registerzahlen soll der Einfluß dieser Lebensereignisse auf das Teilnahmeverhalten von Haushalten im SOEP überprüft werden..

2 Kritik an "Repräsentativitätsprüfungen" für Querschnitte

Es ist bei Querschnitts-Erhebungen gängige Praxis, die "Repräsentativität" der gezogenen Stichprobe mittels diverser Vergleiche von ausgewählten Stichprobenverteilungen und Daten aus der amtlichen Statistik zu "beweisen", vgl. hierzu

²Genauer: Alle Personen, die im Verlaufe des Panels aus dem Inland in bestehende Panelhaushalte gezogen sind.

³Im SOEP wurde diese Regel ab der 5. Welle revidiert. Ab dieser Befragungswelle gilt die Regel, daß alle Personen mit mindestens einem Interview weiterverfolgt werden

etwa Blossfeld (1987) sowie Hartmann/Schimpl-Neimanns (1992). Diese Methode besitzt jedoch einige methodische Schwachstellen (Rendtel/Pötter 1992, 1993), die in diesem Abschnitt kurz angesprochen werden sollen.

Im Lichte dieser Kritik soll dann geklärt werden, wann dieses Verfahren überhaupt auf Panel-Erhebungen sinnvoll angewandt werden kann und wie die Ergebnisse eines derartigen Vergleichs interpretiert werden können.

Zunächst ist der Zielbegriff "Repräsentative Stichprobe" niemals exakt definiert worden (Basu 1971), so daß auch nicht klar ist, was eigentlich Gegenstand einer Repräsentativitätsprüfung ist. Kruskal/Mosteller (1979 a, b, c) geben einen kritischen Überblick, welche diffusen Zielvorstellungen über "repräsentative Stichproben" bestehen. Die am häufigsten vorkommende Zielvorstellung ist sicherlich das Bild von der repräsentativen Stichprobe als Miniatur der Grundgesamtheit: "Unter Repräsentativität verstehen wir, daß in einer Auswahl alle für die Grundgesamtheit typischen und charakteristischen Merkmale und Merkmalskombinationen getreu ihrer relativen Häufigkeit vertreten sein müssen und somit die Auswahl ein verkleinertes Abbild der Grundgesamtheit selbst für solche Merkmale ist, von deren Vorhandensein wir (noch) gar nichts wissen". (Böltken, 1976 S. 128).

Innerhalb dieses Ansatzes bedeutet ein "Repräsentativitätstest" die Überprüfung der Stichprobenverteilung auf Übereinstimmung mit den bekannten "typischen" und "charakteristischen" Merkmalen der Grundgesamtheit. Abgesehen von der Schwierigkeit der Festlegung, welche Merkmale für die Grundgesamtheit typisch und charakteristisch sind, ist dieses Miniatur-Konzept für die allermeisten Stichprobenkonzepte nicht anwendbar, wenn eine exakte Übereinstimmung gemeint ist. Lediglich eine Quotenstichprobe ist nach diesen Leitvorstellungen "repräsentativ". Stichproben, die über ein zufälliges Ziehungsverfahren gewonnen wurden, werden allein wegen der Zufälligkeit der Stichprobe nicht exakt mit der Verteilung in der Grundgesamtheit übereinstimmen. Bei geschichteten Stichproben mit ungleichen Ziehungswahrscheinlichkeiten kann die Stichprobenverteilung mit der Verteilung in der Grundgesamtheit nicht übereinstimmen. Es zeigt sich, daß die Definition einer "repräsentativen Stichprobe" im Sinne einer exakten Proportionalität zur Grundgesamtheit kein sinnvoller Qualitätsanspruch an eine Stichprobe ist.

Sie deckt sich auch nicht mit der Praxis von "Repräsentativitätsstudien". Diese akzeptieren gewisse Abweichungen von Stichprobe und Grundgesamtheit als zufällig und berücksichtigen unterschiedliche Ziehungswahrscheinlichkeiten durch die Verwendung von "Gewichtungen". Die gewichtete Stichprobenverteilung ist jedoch mit der Anwendung eines Populationsschätzers auf Basis der inversen Auswahlwahrscheinlichkeiten identisch. Auch wird das Repräsentativitäts-Argument meist im Zusammenhang mit möglichen selektiven Einflüssen bei der Stichprobenauswahl gebracht, so daß der Eindruck entsteht, "Repräsentativität" sei ein Synonym für fehlerfreie Durchführung des Erhebungsdesigns. Damit müßten sich aber die "zufälligen" Abweichungen der — eventuell gewich-

teten — Stichprobe von der Grundgesamtheit auf die Zufälligkeit des Ziehungs-experiments beziehen.

Repräsentativitätsstudien wie die von Blossfeld (1987) und Hartmann/Schimpl-Neimanns (1992) verwenden aber den χ^2 -Anpassungstest für die Feststellung, ob die Stichprobe "signifikant" von der Grundgesamtheit abweicht. Der χ^2 -Anpassungstest berücksichtigt jedoch in keiner Weise das Erhebungsdesign der Stichprobe. Das Erhebungsdesign beeinflusst aber ganz wesentlich die Varianz der Populationsschätzer. Der χ^2 -Anpassungstest rekurriert auf eine Situation unabhängig und identisch verteilter (i.i.d.) Merkmale und testet die Null-Hypothese, ob ihre Verteilung F mit einer vorgegebenen Verteilung F_0 übereinstimmt. Der χ^2 -Test macht damit Aussagen über die Verteilung der Merkmale und nicht Aussagen über die Verteilung der Populationsschätzer unter dem Ziehungsexperiment. Damit enden "Repräsentativitätsprüfungen" in dem Widerspruch, Aussagen über die Wahrscheinlichkeitsverteilung des Ziehungs-experiments mit Verteilungsaus-sagen über die Bildung von Merkmalen in der Population zu begründen.

Allerdings verwenden Repräsentativitätsstudien manchmal auch eine Sprechweise, als ob tatsächlich Aussagen über die Bildung von Merkmalen getestet werden sollen. Der Form nach wird die Nullhypothese getestet, ob die Merkmalsbildung F der Stichprobe mit einer vorgegebenen Verteilung F_0 übereinstimmt. Als Ergebnis eines derartigen Tests stellen Hartmann/Schimpl-Neimanns (1992) fest: "Es konnte festgestellt werden, daß die Sozialforschungsumfrage Allbus 1990 nicht aus derselben Grundgesamtheit wie der Mikrozensus stammt". Ganz abgesehen davon, daß ein derartiges Resultat wenig hilfreich dabei ist, was nun mit den Daten der betreffenden Stichprobe geschehen soll und sich sofort die Frage aufwirft, aus welcher Grundgesamtheit die Stichprobe denn gezogen wurde, ist die i.i.d.- Voraussetzung derartiger Resultate mehr als zweifelhaft. Gerade in einem Haushaltspanel sind viele Merkmale von Personen über den Haushaltskontext abhängig. Daß die marginale Verteilung für Merkmale wie Bildung und Stellung im Beruf für alle Personen der Grundgesamtheit identisch ist, ist alles andere als realistisch. Je heterogener diese Verteilung in der Bevölkerung ist, umso stärker schwanken die beobachteten Werte um die erwarteten Werte. Mit anderen Worten: Der χ^2 -Test mißt lediglich die Heterogenität der Bildungs-Verteilung in der Bevölkerung und nicht - wie behauptet - die Abweichung der Bildungsverteilung von Stichprobe und Grundgesamtheit.

Eine konsequente Fortsetzung des ursprünglichen Gedankenganges müßte jedoch die Verteilung eines Populationsschätzers \hat{P} unter dem Ziehungsexperiment betrachten. Hierbei ist die Bestimmung von $Var(\hat{P})$ für komplexe Erhebungsdesigns wie beispielsweise ein Panel — keine leichte Aufgabe. Nehmen wir einmal an, wir könnten $\sigma(\hat{P})$ sinnvoll schätzen. Dann ist $t = \frac{\hat{P}-P}{\hat{\sigma}(\hat{P})}$ eine sinnvolle Bewertung des Abstands von \hat{P} und P . Wir wollen zunächst annehmen, daß nur ein Vergleich von \hat{P} mit dem wahren Wert P durchgeführt wird und die Verteilung von t unter der Nullhypothese ("Korrekte Durchführung des Erhebungsdesigns") gut

durch eine Standardnormalverteilung approximativ wird. Ein Test auf Repräsentativität würde die Nullhypothese nicht ablehnen, solange $|t| \leq q_{\alpha/2}$ gilt, wobei $q_{\alpha/2}$ das $\alpha/2$ -Quantil der Standardnormalverteilung ist.

Kann man aus einem derartigen Resultat schließen, daß das Erhebungsdesign korrekt durchgeführt wurde? Hinter dieser Frage verbirgt sich die Frage nach dem Fehler 2. Art, also die Wahrscheinlichkeit H_0 irrtümlich zu akzeptieren. Ganz offensichtlich ist der Raum der Alternativen zu H_0 sehr groß und man kann sich leicht Abweichungen vom Erhebungsdesign vorstellen, die die Verteilung von \hat{P} nicht beeinflussen. Auf diesen Alternativen hat der "Test auf Repräsentativität" die Power α . An dieser Stelle zeigt sich der zweite wesentliche Mangel von "Repräsentativitäts-Tests": Es fehlt an der Formulierung expliziter Alternativen, die mit hoher Wahrscheinlichkeit aufgedeckt werden sollen. In der vorliegenden Form hat der "Repräsentativitäts Test" rein explorativen — d.h. Hypothesen generierenden — Charakter.

Schließlich werden derartige "Tests" aber für alle als wichtig erachteten Merkmale der Population durchgeführt. Dahinter mag der implizite Gedanke stecken, daß jeder Einzelvergleich bezüglich bestimmter Alternativen eine hohe Power hat, so daß die Gesamtprozedur eine hohe Power auf der Vereinigung der Alternative hat. Hierbei würde die Gesamtprozedur die Null-Hypothese ablehnen, wenn einer der Einzeltests die Null-Hypothese ablehnt. Eine offensichtliche Schwierigkeit dieser Herangehensweise ist die Bestimmung der gemeinsamen Verteilung der Prüfstatistiken unter H_0 . Wird für jeden Einzeltest das $\alpha/2$ -Quantil verwendet, so steigt mit wachsender Zahl der Einzeltests die Wahrscheinlichkeit, daß mindestens ein Test H_0 verwirft. Eine obere Grenze für den Fehler 1. Art ist durch $k(=$ Anzahl der durchgeführten Tests) $\times \alpha$ gegeben, vgl. Cox/Hinkley (1979, S. 78). Die von den meisten Repräsentativitätsstudien postulierten Fehlerniveaus berücksichtigen keinesfalls Effekte des multiplen Testens. Die Approximation des multiplen Niveaus durch $k\alpha$ ist konservativ, d.h. sie vergrößert den Fehler 2. Art. Insgesamt ist das α -Niveau eines derartigen "Repräsentativitätstests" sehr unsicher.

Somit ist dieser "Test" für den Anwender wenig hilfreich: Wird H_0 nicht abgelehnt ("Die Stichprobe ist repräsentativ"), so kann der Fehler 2. Art auf einigen Alternativen zu H_0 trotzdem beträchtlich sein. In der Konsequenz können die Schätzungen anderer Populationsmerkmale grob verzerrt sein. Diese Möglichkeit steht in deutlichem Widerspruch zu den Erwartungen der meisten Befürworter von "Repräsentativitätstests". Diese versprechen sich von einem positiven Ausgang der Tests, daß auch hinsichtlich aller anderen Merkmale die Stichprobenverteilung mit den entsprechenden Werten der Grundgesamtheit übereinstimmen ("Repräsentativitätsschluß").

Bei Ablehnung von H_0 ("Die Stichprobe ist nicht repräsentativ") muß in dieser Sichtweise auf die Auswertung des Datensatzes verzichtet werden, da alle Schätzungen "verzerrt" sind. Wie im folgenden gezeigt wird, ist diese Konsequenz zumindest für ein Panel keineswegs zwingend.

Insgesamt ist eine Repräsentativitätsprüfung wenig hilfreich bei der Beantwortung der in der Tat zentralen Frage, ob man aus einem gegebenen Datensatz Rückschlüsse auf die Population ziehen kann und wie dies bewerkstelligt werden kann.

3 Möglichkeiten zur Kontrolle von Selektionsprozessen in Panelerhebungen

Bevor man also Vergleiche von Populationswerten mit geschätzten Werten durchführt, sollte man überlegen, welchen Nutzen man überhaupt aus derartigen Vergleichen ziehen kann. Die Antwort hierauf ist, daß signifikante Abweichungen im Schätzwert eines Merkmals den Verdacht⁴ begründen, daß die erhebungsbedingten Ausfälle bezüglich dieses Merkmals nicht neutral sind. In einem Panel ist jedoch häufig das betreffende Merkmal auch für die Stichprobenausfälle bekannt. Somit läßt sich ein derartiger Verdacht leicht überprüfen, indem man die Ausfallvariable auf das betreffende Merkmal regressiert. Sollen mehrere Merkmale simultan auf ihren Einfluß auf die felddingten Ausfälle überprüft werden, führt dies auf eine multiple Logit- oder Probitanalyse der 0–1 Variablen "Felddingter Ausfall". Allerdings müssen alle Merkmale auch für die Nonrespondenten bekannt sein.

Die Schätzung von Ausfallwahrscheinlichkeiten im Rahmen einer Logit-Analyse kann zu einer Revision der bisher benutzten Populationsschätzer benutzt werden. Ist das betreffende Merkmal für die Nonrespondenten nicht bekannt, so kann man unter zusätzlichen Annahmen über den Ausfallprozeß, die Abweichungen von geschätzten und tatsächlichen Populationswerten zu einer indirekten Schätzung der Ausfallwahrscheinlichkeit benutzen, vgl. Rendtel/Pötter (1992).

Der negative Ausgang einer Repräsentativitäts-Prüfung bedeutet daher nicht, daß Populationsschätzungen mit der gezogenen Stichprobe unmöglich sind (da die Stichprobe "nicht repräsentativ" ist). Vielmehr offenbart ein derartiges Ergebnis, daß der Erhebungsprozeß anders abgelaufen ist, als dies bisher angenommen wurde. Folglich müssen die Schätzverfahren, die ja die bisherige Information über den Erhebungsprozeß widerspiegeln, geändert werden.

In einem Panel ergeben sich Vergleichsmöglichkeiten auf Querschnitt- und auf Längsschnittebene. Allerdings stellt die amtliche Statistik zunächst einmal Daten auf Querschnittebene bereit, während längsschnittbezogene Daten kaum zur direkten Verfügung stehen. Lediglich bei der Kumulation von Ereignissen gibt es Vergleichsmöglichkeiten. Insbesondere bei den demographischen Ereignissen Geburt, Tod und Scheidung existieren für die Population sehr genaue Angaben

⁴Es wird hier lediglich von Verdacht gesprochen, da immer die Möglichkeit besteht, daß infolge von unterschiedlichen Erhebungszeitpunkten, differierenden Merkmalsabgrenzungen, etc. das Stichprobenmerkmal etwas anderes mißt als das Merkmal der Referenzverteilung.

auf Basis der Register bei den Standesämtern. Hier ist das Ereignis "Scheidung", das meistens mit einer vorausgegangenem Haushaltsaufspaltung verbunden ist, besonders interessant für Vergleiche, da Haushaltsabspaltungen mit hoher Wahrscheinlichkeit aus dem Panel herausfallen.

Schon bei dem anderen wichtigen Ereignis, das ebenfalls mit einer Haushaltsabspaltung verbunden ist — dem Auszug von Kindern aus dem elterlichen Haushalt — existieren keine amtlichen Vergleichsdaten. Dies gilt auch für alle Merkmale, die sich auf die abgespaltenen (=erhebungstechnisch "neue") Haushalte beziehen. Es gibt keine sinnvolle amtliche Vergleichsgröße für die Merkmale der "neuen" Panelhaushalte. Symmetrisch hierzu gibt es keine sinnvolle Vergleichsgröße für die erhebungstechnisch "alten" Haushalte. Schließlich kann auch aus einem Vergleich von Merkmalen "alter" und "neuer" Haushalte kaum eine sinnvolle Schlußfolgerung für den Erhebungsprozeß gezogen werden. Beispielsweise kann die Alterstruktur in den "neuen" Haushalten ganz verschieden von der Altersstruktur in den "alten" Haushalten sein, weil überwiegend junge Personen den elterlichen Haushalt verlassen.

Das demographische Ereignis "Zuwanderung" belegt eine weitere Schwierigkeit des Vergleichs. Zum einen ist hier bekannt, daß der Panelansatz nur Zuzüge in bestehende Haushalte erfaßt. Ein Vergleich mit den ebenfalls nur geschätzten amtlichen Zuwandererzahlen ergibt also gar keine Hinweise, ob Haushalte mit Zuzügen aus dem Ausland verstärkt aus dem Panel ausgefallen sind. Umgekehrt liegen aber keine amtlichen Zahlen über Zuwanderungen vor, die direkt in bestehende Privathaushalte gezogen sind.

Die Möglichkeiten für Vergleiche mit Populationswerten reduzieren sich also — bis auf einige Ausnahmen — auf Querschnittsvergleiche. Die "klassischen" Vergleichsvariablen sind: Altersgruppen, Geschlecht, Familienstand, Schulbildung, Stellung im Beruf. Werden Vergleiche auf Haushaltsebene durchgeführt, so werden die entsprechenden Variablen für den Haushaltsvorstand gebildet. Zusätzlich wird meistens die Haushaltsgröße mit entsprechenden Populationswerten verglichen. Bis auf die Variable "Haushaltsgröße" sind alle Vergleichsvariablen über die Zeit relativ stabil.

Der Vergleich von Populations- und Schätzwerten im Querschnitt kumuliert die selektiven Effekte aller Erhebungsstufen des Panels bis zur Welle des jeweiligen Querschnitts. Eine signifikante Abweichung von Populations- und Schätzwert gibt also keine Information darüber, welche Stufe des Selektionsprozesses hinsichtlich des betrachteten Merkmals anders als bisher berücksichtigt werden muß.

Allerdings ist dieser Aufwand für ein Panel überflüssig, da man die Hypothesen über einen selektiven Effekt dieser Basismerkmale direkt überprüfen kann.

Eine Logitanalyse der Ausfälle kann nur auf Merkmalen basieren, die aus früheren Befragungsstufen des Panels bekannt sind. Bei zeitlich invarianten Merkmalen bedeutet dies keine Restriktion. Bei zeitlich variablen Merkmalen wie dem Familienstand kann ein Ausfall nur in Abhängigkeit von Y_{t-1} , dem Merkmal

vor der nächsten Auswahlstufe geschätzt werden. Wenn der Ausfall jedoch von $Y_t | Y_{t-1}$ abhängt, so wird diese Abhängigkeit nicht von einer Logitanalyse in Y_{t-1} erkannt. Konkret bedeutet dies im Fall des Familienstands, daß Ausfälle, die von einer Veränderung des Familienstands abhängen, nicht von einer Regression auf den Familienstand zur Zeit $(t - 1)$ aufgedeckt werden. Treten also Ausfälle von Welle $(t - 1)$ auf Welle t überproportional häufig bei Personen auf, die zwischen diesen beiden Befragungswellen geschieden werden, so wird die geschätzte Anzahl der geschiedenen Personen in Welle t unterschätzt, obwohl der Familienstand zur Zeit $(t - 1)$ als Kontrollvariable für die Ausfälle benutzt wurde.

Es läge also auf der Hand, die Schätzergebnisse hinsichtlich der Veränderungsgrößen mit bekannten Populationswerten zu vergleichen. Veränderungsgrößen sind in der Population aber nur für wenige Merkmale bekannt. Auch müssen für den Abgleich von Veränderungsgrößen wesentlich größere Ansprüche an die Zuverlässigkeit und die Vergleichbarkeit der Daten gestellt werden als bei Bestandsgrößen⁵.

Allerdings könnte man einwenden, daß ein Vergleich mit den Bestandsgrößen immer noch aufschlußreich ist. Vermutet man beispielsweise, daß Personen nach einer Ehescheidung mit einer höheren als bisher angenommenen Wahrscheinlichkeit aus dem Panel ausfallen, so müßte sich dies nach einer gewissen Dauer auch in verringerten Bestandszahlen an geschiedenen Personen niederschlagen. Wie aber sollen in diesem Fall die Inklusionswahrscheinlichkeiten modifiziert werden, um verbesserte Populationsschätzungen zu erhalten? Das Verfahren der Randanpassung, das eine Übereinstimmung der revidierten Populationsschätzer mit den Bestandsgrößen erzwingt, würde die bisherigen Inklusionswahrscheinlichkeiten mit einem Faktor multiplizieren, der sich in diesem Fall aus dem Verhältnis von geschätzten ("Ist"-Wert) geschiedenen Personen zu ihrem Bestand ("Soll"-Wert) in der Population ergibt. Diese "Korrektur" erzeugt aber mit Sicherheit verfälschte Schätzungen, da nun für alle geschiedenen Personen eine verringerte Inklusionswahrscheinlichkeit geschätzt wird. Dies ist jedoch für Personen, die vor Eintritt in das Panel geschieden wurden, falsch.

Insgesamt belegt dieser Abschnitt, daß Vergleiche von geschätzten und bekannten Bestandsgrößen für Panelerhebungen wenig hilfreich sind. Auch ist die Grundidee — die "proportionale Abbildung" der Bevölkerung durch die Stichprobe — nicht angemessen, da auch noch aus einer "verzerrten" Panel-Stichprobe unverzerrte, d.h. erwartungstreue, Populationsschätzungen gewonnen werden.

⁵Schon hinsichtlich der Bestandsgrößen "Schulabschluß" und "Haushaltsgröße" können Zweifel an der Zuverlässigkeit der Angaben aus dem Mikrozensus angemeldet werden, vgl. Rendtel/Pötter (1992).

4 Die Analyse von Ausfallprozessen im SOEP

Die Auswahl der klassischen Kontrollvariablen Alter, Geschlecht, Familienstand, Schulabschluß und Stellung im Beruf scheint eher durch Modellbasierte Überlegungen motiviert zu sein; etwa in dem Sinn, daß diese Variablen gute Prädiktoren für andere sozio-ökonomische Merkmale sind. Diese Begründung ist aber im Rahmen des Randomisierungsansatzes (vgl. Cochran 1977, Kapitel 1), bei dem die Merkmale der Befragungspersonen als Konstanten behandelt werden, irrelevant.

Im Rahmen des Randomisierungsansatzes sind nur diejenigen Variablen relevant, hinsichtlich der die Auswahlwahrscheinlichkeiten variieren. Es hat allerdings auch Versuche gegeben, von der Analyse der Interviewsituation ausgehend die Variable Schulabschluß als ursächlich für die Interviewbereitschaft herzuleiten. Beispielsweise stellt Esser (1988) die Hypothese auf, daß bei geringen, persönlichen Präferenzunterschieden hinsichtlich der Gewährung des Interviews schon geringe Motivationsunterschiede ausreichen, um zu unterschiedlichen Antwortentscheidungen zu kommen. Hierbei wird unterstellt, daß ein höheres Interesse am Befragungsgegenstand mit einer höheren Schulbildung einhergeht und auf diese Weise zu einer höheren Teilnahmebereitschaft führt, vgl. z.B. Hartmann/Schimpl-Neimanns (1992).

Bei einer Befragung sind zunächst 2 Stufen mit ganz unterschiedlichen Probenstrukturen zu unterscheiden: Die Erreichbarkeit der Haushalte und die Antwortbereitschaft der Personen, nachdem sie vom Interviewer erreicht wurden. Im folgenden wird versucht, Merkmale herzuleiten, deren Relevanz für das Teilnahmeverhalten sich aus der unmittelbaren Feldarbeit des SOEP ergibt.

4.1 Ausfälle aufgrund von Nichterreichbarkeit

Die Variable "Haushaltsgröße" ist im Zusammenhang mit der Erreichbarkeit der Haushalte zu sehen. Hier gelten Einpersonenhaushalte als schwerer erreichbar als Mehrpersonenhaushalte. In einem Panel tritt jedoch noch ein anderer Aspekt hinzu. Im Falle eines Umzugs muß die neue Adresse des Haushalts ausfindig gemacht werden. Je nach Gegebenheiten können dabei unterschiedliche Hilfen eingesetzt werden:

- Informationen von Nachbarn, der Post oder der Einwohnermeldeämter⁶ bei Umzug des gesamten Haushalts.
- Informationen der nicht verzogenen Haushaltsmitglieder bei Haushaltspaltungen.

⁶In Ländern ohne Einwohnermeldeämter, z. B. den USA, werden die Befragten gebeten, Adressen von Personen zu nennen, die im Falle eines Umzugs Auskunft über die neue Adresse der Befragten geben können.

Tabelle 1: Ausfallraten aufgrund von Nichterreichbarkeit. N = Anzahl der zu erreichenden Haushalte; % = Prozentsatz der Haushalte ohne Kontaktaufnahme

	Welle 2	Welle 3	Welle 4	Welle 5	Welle 6	Welle 7	Welle 8
	Insgesamt						
N	6051	5814	5465	5342	5156	5044	5029
%	1,9	1,4	1,0	0,9	0,9	0,9	0,5
	Haushalte ohne Adressenänderung						
N	5413	5039	4808	4683	4545	4472	4448
%	0,8	0,4	0,1	0,1	0,2	0,0	0,04
	verzogene Mehrpersonen-Haushalte						
N	298	307	272	274	228	186	197
%	7,4	3,6	4,0	5,5	0,5	1,6	0,5
	verzogene Einpersonen-Haushalte						
N	119	180	142	143	126	122	94
%	21,0	14,4	7,7	5,6	4,7	5,7	1,1
	abgespaltene Haushalte						
N	221	295	242	242	246	263	290
%	11,7	8,4	10,4	7,4	11,8	12,9	7,6

Weiterhin kann die Hypothese aufgestellt werden, daß in einem eher anonymen Wohnumfeld weniger Informationen von Nachbarn über die Antreffbarkeit von Haushaltsmitgliedern genutzt werden können. Im SOEP wurden daher noch ein Indikator für das direkte Wohnumfeld sowie der Großstadtindikator in die Analyse der Ausfälle durch Nichterreichbarkeit aufgenommen, vgl. Rendtel (1988).

Das Problem der Erreichbarkeit ist von unterschiedlicher Relevanz für die erste Welle eines Panels: Werden die Adressen über ein Register gezogen, so ergibt sich hier das Problem bei der veralteten Adressen die Haushalte an ihrer aktuellen Adresse zu erreichen. Im SOEP wurden die Haushalte der ersten Welle jedoch über eine Begehung der jeweiligen Sample-Points ermittelt, so daß die Ausfälle aufgrund von Nichterreichbarkeit gering ausfallen, vgl. Wagner/Schupp/Rendtel (1992).

Bei den nachfolgenden Panelwellen sind die Ausfälle aufgrund von Nichterreichbarkeit bei unveränderter Adresse vernachlässigbar gering, vgl. Tabelle 1. Allerdings zeigt sich, daß es bei verzogenen bzw. abgespaltenen Haushalten während der gesamten Laufzeit des Panels zu nennenswerten Ausfällen gekommen ist. So sind beispielsweise in Welle 2 immerhin 21% der verzogenen Einpersonenhaushalte ausgefallen. Deutlich erkennbar ist jedoch eine Tendenz zur Verbesserung der Feldarbeit.

4.2 Ausfälle aufgrund von Antwortverweigerungen

Wesentlich bedeutender als die Ausfälle durch Nichterreichbarkeit sind die Ausfälle aufgrund von Antwortverweigerungen. Die Ausfallraten aufgrund von Antwortverweigerungen sind ca. 6 mal so hoch wie bei Ausfällen durch Nichterreichbarkeit. Die Interviewsituation ist in einem Panel durch eine Reihe von spezifischen Variablen gekennzeichnet⁷. Wir wollen dabei von der Situation im SOEP⁸ ausgehen, bei der die Haushaltsmitglieder von einem Interviewer befragt werden⁹.

Zunächst einmal ist die Interviewsituation dadurch gekennzeichnet, daß alle Haushaltsmitglieder über 15 Jahre befragt werden. Hier ist es naheliegend, starke Abhängigkeiten zwischen den Teilnahmeentscheidungen der Haushaltsmitglieder zu vermuten.

Weiterhin wird es relevant sein, wie lange eine Person schon am Panel teilgenommen hat. Eng hiermit verbunden ist die Frage, ob der Interviewer seit der letzten Befragung gewechselt hat. Kommt es in einer größeren Zahl von Fällen zu einem Wechsel des Interviewers, ohne daß der Haushalt umgezogen ist, so kann zwischen zwei konkurrierenden Hypothesen entschieden werden. Die eine Hypothese besagt, daß sich eine hohe Antwortbereitschaft allein aus der bisherigen Teilnehmerdauer herleiten läßt. Die andere Hypothese besagt, daß die Antwortbereitschaft eher auf dem Vertrauensverhältnis zu der Person des Interviewers beruht, vgl. Rendtel (1990). Die Beantwortung dieser Frage hat erhebliche Konsequenzen für ein Panel, da hier eine Möglichkeit aufgezeigt wird, den Umfang der feldbedingten Ausfälle zu beeinflussen. Wenn nämlich die Antwortbereitschaft bei einem Wechsel des Interviewers sinkt, so müßte eine gute Feldarbeit darauf hinwirken, derartige Wechsel möglichst gering zu halten.

Weiterhin lassen sich aus dem früheren Antwortverhalten Rückschlüsse auf die Antwortbereitschaft in der aktuellen Panelwelle herleiten. Wird beispielsweise in der vorhergehenden Welle keine Angabe zum Haushaltsnettoeinkommen gemacht, so kann man dies als Indikator für eine unterdurchschnittliche Teilnahmebereitschaft werten, die in der nächsten Welle zum Abbruch der Zusammenarbeit führt¹⁰.

Da das SOEP ganz wesentlich auf den Themenschwerpunkt Einkommen ausgerichtet ist, kann man die Hypothese aufstellen, daß die Antwortbereitschaft mit dem Einkommen variiert. Allerdings läßt sich eine derartige Hypothese nur anhand der Einkommen aus der Welle ($t - 1$) überprüfen. Eine Hypothese, ob die Antwortbereitschaft von Einkommensänderungen abhängt, läßt sich mit den

⁷Vergleiche hierzu auch Rendtel (1988, 1990).

⁸In der PSID wurde nach der fünften Befragungswelle die telefonische Befragung der Haushalte eingeführt. Allerdings wird dort nur der Haushaltsvorstand befragt.

⁹Hierbei ist das persönliche Interview die hauptsächliche Befragungsform. Gelegentlich wird der Fragebogen in Gegenwart des Interviewers selbst ausgefüllt, vgl. Wagner/Schupp/Rendtel (1992).

¹⁰Im SOEP fehlt pro Welle in ca. 5% der Haushalte die Angabe zum Haushaltsnettoeinkommen.

Paneldaten nicht direkt überprüfen. Unter der Hypothese, daß negative Einkommensveränderungen die Antwortbereitschaft verringern, müßte die geschätzte Einkommensverteilung oberhalb der Einkommensverteilung in der Population liegen. Leider wird jedoch die entsprechende amtliche Vergleichserhebung, die Einkommens- und Verbraucherstichprobe (EVS), als Quotenstichprobe erhoben und ist von daher nicht mit dem SOEP vergleichbar,¹¹ so daß an dieser Stelle von einem derartigen Vergleich abgesehen wird.

Neben der Einkommensvariablen reflektiert die Nennung bestimmter Vermögensformen die ökonomische Situation eines Haushalts. Insgesamt können 5 Vermögensanlageformen angegeben werden¹². Die Summe dieser Nennungen kann als grober Indikator für die ökonomische Situation des Haushalts gesehen werden.

Da ein Haushaltspanel von den Befragten zunächst als Haushaltsbefragung wahrgenommen wird, sollte es nicht verwundern, daß bei einer wesentlichen Veränderung des Lebenszusammenhangs oder des Wohnorts der Personen die alte Erhebungseinheit als nicht mehr vorhanden betrachtet wird. In diesem Fall dürfte also die Antwortbereitschaft deutlich sinken, zumal in diesen Fällen häufig auch die Person des Interviewers wechselt. Allein aus diesen Gründen ist damit zu rechnen, daß die Antwortneigung in umgezogenen bzw. abgespaltenen Haushalten sinkt. Aus ähnlichen Gründen kann man erwarten, daß bei Zuzug einer neuen Befragungsperson die Antwortbereitschaft im Haushalt sinkt.

Diese Haushaltsmobilität ist jedoch mit einer Reihe von anderen Ereignissen konfundiert. So wurde bereits erwähnt, daß Haushaltsabspaltungen das Ergebnis des Auszugs von Kindern, der Auflösung einer Ehe aber auch die Folge anderer Lebensereignisse sein kann. Für den Umzug eines Haushalts gilt es eine Vielzahl von Gründen: Umzug in ein Altersheim, Beginn einer Ehe oder Lebensgemeinschaft, Wechsel des Arbeitsplatzes oder des Studienorts, Versetzung durch den Arbeitgeber etc. Wesentlich ist hierbei, daß bei den nicht teilnahmebereiten Haushalten (bzw. Personen) das den Umzug verursachende Ereignis nicht bekannt ist. Bekannt ist lediglich die Haushaltsmobilität. Natürlich lassen sich in Analogie zu dem Haushaltseinkommen eine Fülle von Vermutungen darüber anstellen, ob derartige Lebensereignisse die Antwortbereitschaft in einem Panel beeinflussen. Alle diese Hypothesen lassen sich nicht anhand der Paneldaten überprüfen. Die Überprüfung derartiger Hypothesen ist für die Bestimmung der Auswahlwahrscheinlichkeiten auch nicht notwendig, falls alle diese Ereignisse ungefähr denselben Effekt auf die Antwortbereitschaft haben.

Dieser Gedankengang sei am Beispiel der abgespaltenen Haushalte noch einmal verdeutlicht. Man kann davon ausgehen, daß das Gros der Haushaltsabspaltungen durch den Auszug von Kindern aus dem Elternhaus oder durch die Trennung einer Partnerschaft entsteht. Bei der Berechnung der Auswahlwahr-

¹¹Die Stichprobenmitglieder werden im Rahmen des Quotenplans über Zeitungsinserate gewonnen.

¹²Sparbuch, Bausparvertrag, Lebensversicherung, Wertpapiere, Betriebsvermögen.

scheinlichkeiten werden diese Ereignisse nicht unterschieden. In Abschnitt 5 wird u. a. die geschätzte Zahl der Ehescheidungen mit dem Populationswert verglichen. Eine signifikante Unterschätzung dieser Zahl läßt daher den Rückschluß zu, daß das Ausfallrisiko durch dieses Ereignis höher ist als das Ausfallrisiko von Kindern nach dem Verlassen des Elternhauses¹³.

Eine Ausfallanalyse kann auf Personen- wie auch auf Haushaltsebene durchgeführt werden. Allerdings ist im SOEP die Abhängigkeit der Teilnahmebereitschaft von Personen innerhalb eines Haushalts nicht zu übersehen. Hier war die Teilnahmebereitschaft aller Haushaltsmitglieder im Befragungsalter eine Voraussetzung für die Auswahl in der ersten Welle. Doch auch in den nachfolgenden Wellen reagiert die Mehrzahl aller Haushalte mit einem geschlossenen Ausfallverhalten. Durchschnittlich reagieren ca. 98% der Mehrpersonenhaushalte geschlossen.

Eine Möglichkeit, den Grad der Abhängigkeit der Einzelentscheidungen in Mehrpersonenhaushalten zu quantifizieren besteht darin, die konditionalen Antwortwahrscheinlichkeiten bei gegebenem Antwortverhalten einer Person zu analysieren. Hierfür wird im folgenden die Person des Haushaltsvorstands gewählt. Es bezeichne R_i den Responseindikator für Person i und $R_{h(i)}$ den Responseindikator für den Haushaltsvorstand¹⁴. Neben dem Antwortverhalten des Haushaltsvorstands können auch noch weitere Kovariaten X_i für die Antwortbereitschaft wichtig sein. Unter Unabhängigkeit der Antwortentscheidungen im Haushalt gilt:

$$Pr(R_i = r \mid R_{h(i)} = 1, X_i) = Pr(R_i = r \mid R_{h(i)} = 0, X_i) \quad (1)$$

Diese Bedingung ist damit identisch, daß in einer Logitanalyse der bedingten Wahrscheinlichkeiten der Effekt der Variablen $R_{h(i)}$ vernachlässigt werden kann.

Allerdings sind für die neu in das Panel eingetretenen Personen unter Umständen die Werte für die Kovariaten X_i unbekannt. Bei nicht teilnahmebereiten neuen Personen können Informationen über Alter und Geschlecht aus den Kontaktprotokollen entnommen werden. Es kann allerdings auch vorkommen, daß noch nicht einmal die Zusammensetzung bei nicht auskunftsbereiten Haushalten bekannt ist. In diesen Fällen ist nicht bekannt, wieviele Personen überhaupt die Teilnahme am Panel verweigert haben.

Bei der folgenden Analyse der bedingten Antwortbereitschaft werden daher nur die Kovariaten Stichprobenzugehörigkeit, Altersgruppe¹⁵, Geschlecht sowie Befragungsstatus benutzt. Beim Befragungsstatus wird danach differenziert, ob die Befragungsperson schon im Vorjahr interviewt wurde, als Kind in einem Panelhaushalt das Befragungsalter erreicht oder als neue Befragungsperson in einem Panelhaushalt gezogen ist. Personen, für die diese Merkmale nicht bekannt sind, werden von der Analyse ausgeschlossen.

¹³Dies ist jedoch nicht der einzig mögliche Rückschluß, vgl. die Diskussion in Abschnitt 5.

¹⁴Genauer: Diejenige Person, die in der letzten Welle den Haushaltsfragebogen beantwortet hat. Bei abgespaltenen Haushalten: Diejenige Person, die bisher am Panel teilgenommen hat.

¹⁵Ausprägungen: 16-24, 25-34, 35-54, 55-64, 65-74, 65+.

Tabelle 2: p-Werte und Koeffizienten eines Logitmodells für personenbezogene Ausfälle bei gegebenem Antwortverhalten des Haushalts-Vorstands

Variable	p-Wert in Welle						
	2	3	4	5	6	7	8
Stichprobe A/B	0.02	0.76	0.16	0.97	0.27	0.44	0.06
Geschlecht	<0.001	0.002	0.23	<0.001	<0.001	0.002	0.02
Altersgruppe	0.46	0.001	0.005	0.69	0.34	0.035	0.007
Befragungsstatus	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
Verhalten des HV	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
Beobachtungen	6588	6305	5965	5824	5580	5394	5209
	Koeffizienten in Welle						
Konstante	0.7	-0.1	-0.3	0.2	0.6	-0.3	-0.3
Geschlecht							
Mann	0.4	0.3	0.1	0.5	0.5	0.3	0.2
Frau	-0.4	-0.3	-0.1	-0.5	-0.5	-0.3	-0.2
Befragungsstatus							
alte Befragungsperson	-0.9	-0.3	-0.6	-0.6	-0.8	-0.6	-0.7
Kind erreicht 16 Jahre	-0.1	-1.6	-1.1	-1.4	-1.0	-0.9	-0.9
neue Befragungsperson	1.0	1.9	1.7	2.0	1.8	1.5	1.6
Verhalten des HV							
HV nicht teilnahmebereit	3.7	3.7	3.5	3.4	3.6	2.9	2.7
HV teilnahmebereit	-3.7	-3.7	-3.5	-3.4	-3.6	-2.9	-2.7

Tabelle 2 zeigt im oberen Teil die Signifikanzniveaus (p-Werte) der genannten Kovariaten in den Wellen 2 bis 8. Hierbei gibt der p-Wert das α -Niveau an, bei dem die Nullhypothese gerade noch abgelehnt wird. Ein niedriger p-Wert zeigt also einen deutlichen Einfluß der betreffenden Variablen auf das Antwortverhalten an. Dieser Test überprüft die Nullhypothese, ob alle freien Koeffizienten zu einer Kovariaten den Wert 0 haben. Hat eine Variable M Ausprägungen, so sind $M - 1$ Koeffizienten frei schätzbar¹⁶, d.h. die Nullhypothese lautet:

$$\beta = (\beta_1, \dots, \beta_{M-1})' = 0$$

Die benutzte Teststatistik ist:

$$T = \hat{\beta}' \hat{V}^{-1}(\hat{\beta}) \hat{\beta}$$

Für die Verteilung von T unter H_0 wird die χ^2 -Approximation benutzt. Es zeigt sich lediglich bei den Variablen "Geschlecht", "Befragungsstatus" und "Verhalten des Haushaltsvorstands" ein über alle ¹⁷ Befragungswellen signifikanter Einfluß auf die Antwortbereitschaft.

Die Koeffizienten im unteren Teil von Tabelle 2 geben den Einfluß eines Merkmals auf das logarithmierte Verhältnis von Ausfällen zu den Antworten wieder:

$$\ln \frac{Pr(R_i = 0 \mid R_{h(i)}, X_i)}{Pr(R_i = 1 \mid R_{h(i)}, X_i)} = \beta_0 + \beta_{R_{h(i)}} + \beta_{X_i} \quad (2)$$

Ist β positiv, so vergrößert sich das Verhältnis von Verweigerungen zu Antworten. Man erhält damit das plausible Ergebnis, daß neue Befragungspersonen weniger auskunftsbereit sind als alte Befragungspersonen bzw. deren Kinder, die das Befragungsalter erreichen. Weiterhin sind Männer, wenn sie nicht Haushaltsvorstand sind, weniger auskunftsbereit als Frauen.

Allerdings können diese Unterschiede in der Antwortbereitschaft ignoriert werden angesichts des Effekts, den das Antwortverhalten des Haushaltsvorstands auf die Antwortbereitschaft der übrigen Personen hat. Bei sonst gleichen Kovariaten vergrößert sich das Chancenverhältnis von Ausfällen zu Antworten um einen Faktor von $e^{2 \times 3.5} = 1096$, wenn der Haushaltsvorstand statt zu antworten die Teilnahme am Panel verweigert.

Als Resultat dieser Analyse erhält man, daß die Teilnahmeentscheidung im wesentlichen eine Entscheidung auf Haushaltsebene ist. Dies hat den Vorteil, daß bei der Schätzung der Ausfallwahrscheinlichkeiten bekannte Haushaltscharakteristika aus der vorhergehenden Befragungswelle benutzt werden können¹⁸. Viele der über die Feldarbeit hergeleiteten Merkmale (Interviewerwechsel, Umzug

¹⁶Das hier benutzte SAS-Programm-Modul CATMOD verwendet die Standardparametrisierung, d.h. $\beta_M = \sum_{m=1}^{M-1} \beta_m$.

¹⁷Mit Ausnahme der Welle 4 für das Merkmal Geschlecht.

¹⁸Dies war auf Personenebene bei zugezogenen Befragungspersonen nicht möglich.

des Haushalts, Haushaltsabspaltung, Angaben zum Haushaltsnetto) sind ohnehin haushaltsspezifisch. Lediglich die bisherige Teilnahmedauer am Panel ist innerhalb eines Haushalts unter Umständen nicht einheitlich. Hier bieten sich verschiedene Operationalisierungen an: Teilnahmedauer des Haushaltsvorstands oder Mittelwertbildung über alle Haushaltsmitglieder. Eine ähnliche Problematik ergibt sich bei den sozio-demographischen Variablen, Alter, Geschlecht, Schulbildung und Stellung im Beruf. Da hier eine Mittelung über die Haushaltsmitglieder sinnlos ist, wurden die Merkmale des Haushaltsvorstand für die Ausfallanalyse gewählt.

Tabelle 3 vergleicht Umfang (N) und Ausfallraten (%) erreichter Haushalte hinsichtlich der klassischen Basisvariablen, sowie der Teilstichprobenabgrenzung in Deutsche (A) und Ausländer (B). Insgesamt zeichnet sich eine abnehmende Tendenz bei den Ausfallraten ab, die nur von einem Anstieg in der fünften Befragungswelle unterbrochen wird. Zu diesem unregelmäßigen Verlauf dürften zwei Ereignisse beigetragen haben: Zum einen eine massive öffentliche Kampagne für die Volkszählung während der Feldarbeit von Welle 4, zum anderen der schwierige Befragungsschwerpunkt der 5. Welle, der die Vermögenssituation der Haushalte zum Gegenstand hatte.

Die Ausfallraten schwanken zwischen den Stichproben nur unbedeutend. Dasselbe gilt hinsichtlich des Geschlechts des Haushaltsvorstands. Beim Merkmal Alter zeigt sich in der Gruppe der über 75-jährigen eine erhöhte Ausfallrate. Allerdings ist dieses Muster nicht über die Zeit stabil. In Welle 8 ist die Ausfallrate bei dieser Altersgruppe zusammen mit den 65-bis 75-jährigen niedriger als bei den übrigen Altersgruppen. Lediglich bei der jüngsten Altersgruppe zeigen sich durchgängig überdurchschnittliche Ausfallraten. Diese sind aber – wie die multiple Analyse zeigt – eher der Tatsache zuzuschreiben, daß die meisten dieser Haushalte "neue" Haushalte sind.

In abgeschwächter Form gilt dieses Argument auch für die Ausprägung "ledig" bei der Variablen Familienstand, bei der überdurchschnittliche Ausfallraten auftreten. Die Unterschiede hinsichtlich der übrigen Ausprägungen des Familienstandes sind gering und zeigen kein über die Wellen stabiles Muster.

Bei der Variablen "Schulabschluß" gibt es kein stabiles Gefälle zwischen den Ausfallraten hinsichtlich des Niveaus der Abschlüsse. Die Ergebnisse von Tabelle 3 liefern keine Basis für die Gültigkeit der Hypothese vom "Bildungsbias" im SOEP: Insgesamt sind die Unterschiede zwischen den Ausfallraten eher gering. In den Wellen 4,5, und 7 ist die Ausfallraten bei den Abiturienten höher als bei den Personen mit Hauptschulabschluß.

Die Variable "Stellung im Beruf" reflektiert den "Mittelstandsbias". Dieser postuliert für Personen auf niedrigem sozialen Niveau eine geringe Teilnahmebereitschaft aufgrund geringen Interesses an der Befragung. Personen in hoher sozialer Stellung wird eine geringe Teilnahmebereitschaft aufgrund mangelnder Zeit unterstellt. Die Variable "Stellung im Beruf" weist un- und ausgelernten Arbeitern sowie einfachen Angestellten die Position "niedrig" zu. Die Position

Tabelle 3: : **Teilnahmeverhalten von erreichten Haushalten nach sozio-demographischen Merkmalen des Haushaltsvorstands.** N = Anzahl der zu befragenden Haushalte, % = Prozentsatz der Haushalte ohne Interviewer

		Welle						
		2	3	4	5	6	7	8
Insgesamt	N	5937	5473	5201	5131	4926	4828	4802
	%	10.4	8.9	5.1	7.0	6.2	5.5	4.7
		Stichprobe						
A (Deutsche)	N	4611	4275	4058	3993	3834	3755	3716
	%	10.2	8.7	5.2	7.1	6.2	5.3	4.6
B (Ausländer)	N	1326	1198	1143	1138	1092	1073	1086
	%	10.9	9.6	5.0	6.9	6.4	6.1	5.3
		Geschlecht						
männlich	N	4664	4226	3951	3840	3624	3486	3413
	%	9.8	8.3	4.7	6.7	6.2	5.0	4.2
weiblich	N	1273	1247	1250	1291	1303	1342	1389
	%	12.2	11.0	6.5	7.9	6.5	6.9	6.1
		Alter						
75 u. älter	N	448	394	374	386	381	380	371
	%	18.3	13.7	6.7	6.5	6.8	5.0	3.0
65-74	N	562	513	487	480	465	462	477
	%	10.1	9.4	2.9	5.4	6.0	3.0	2.9
55-64	N	947	860	809	803	798	783	782
	%	9.8	8.7	4.2	4.7	5.9	5.5	3.8
35-54	N	2621	2401	2272	2226	2112	2017	1970
	%	9.2	7.8	4.3	6.3	6.1	4.9	4.7
25-34	N	1116	1034	976	963	904	926	957
	%	8.9	8.0	6.1	9.9	6.1	7.7	6.1
16-24	N	243	271	283	273	266	260	245
	%	17.3	15.1	13.4	13.2	9.0	8.1	8.9
		Familienstand						
verheiratet, zusammen	N	3893	3600	3366	3301	3144	3029	3015
	%	9.6	8.0	4.4	6.2	6.2	4.6	3.4
verheiratet, getrennt	N	104	157	119	97	120	110	96
	%	7.7	12.1	5.9	13.4	6.7	8.2	12.5
ledig	N	836	811	802	783	764	764	782
	%	12.6	9.6	9.4	11.5	6.8	8.6	9.1
geschieden	N	349	345	328	347	327	351	356
	%	10.3	10.4	4.6	6.3	7.0	7.1	4.8
verwitwet	N	671	560	533	542	534	532	518
	%	12.6	11.9	3.9	5.1	4.7	4.5	4.4

Tabelle 3: Fortsetzung

		Schulbildung							
Kein Abschluß	N	493	445	411	407	376	373	380	
	%	10.9	11.0	6.3	8.8	6.7	5.6	5.8	
Hauptschule	N	2952	2669	2493	2488	2405	2340	2314	
	%	11.8	9.8	4.7	6.2	6.3	5.2	4.6	
Realschule	N	852	849	818	805	798	780	784	
	%	6.8	8.8	5.6	8.3	8.0	6.3	4.9	
Fachhochschule	N	223	205	205	201	183	180	184	
	%	9.4	6.8	6.3	9.5	5.5	5.6	7.6	
Abitur	N	601	588	582	569	542	552	533	
	%	7.5	6.0	5.5	7.6	4.2	7.4	5.1	
		Stellung im Beruf							
nicht erwerbs- tätig	N	1527	1325	1290	1302	1276	1329	1339	
	%	13.9	10.1	5.5	6.9	6.4	5.2	4.3	
arbeitslos	N	206	297	260	258	265	193	199	
	%	9.7	10.7	7.7	8.5	4.9	2.6	6.0	
hoch	N	585	578	522	530	519	511	496	
	%	7.9	6.2	4.4	7.7	7.1	8.4	5.9	
mittel	N	2248	2202	2053	1982	1911	1803	1857	
	%	8.8	8.1	5.3	6.6	6.2	5.1	4.7	
niedrig	N	1364	1071	1076	1059	954	992	911	
	%	11.2	10.0	4.3	7.4	6.2	5.9	4.6	

“hoch” wird für Freie Berufe, Selbständige mit mehr als 10 Mitarbeitern, Angestellte mit hochqualifizierter Tätigkeit oder Leitungsfunktion sowie für Beamte im höheren Dienst verwendet. Wegen der besonderen Bedeutung der Arbeitslosen für Analysen mit dem SOEP werden die arbeitslos gemeldeten Personen extra ausgewiesen. Wiederum findet man in den Ausfallraten von Tabelle 3 keine Hinweise auf einen “Mittelstands-Bias” im SOEP. Lediglich in Welle 5 und 7 stehen die Ausfallraten in der erwarteten Relation. In Welle 4 haben die Mittelständler höhere Ausfallraten als die Haushalte mit hohem und mit niedrigem Status. Hinsichtlich der Arbeitslosigkeit ergeben sich ebenfalls keine stabilen Relationen. In Welle 6 und 7 haben die Arbeitslosen die insgesamt niedrigsten Ausfallraten, in den Wellen 3, 4, 5 und 8 die höchsten.

Kommen wir nun zu der Abhängigkeit der Ausfallraten von feldbedingten Merkmalen. Diese sind in Tabelle 4 aufgeführt. Während bei den sozio-demographischen Basisvariablen die Ausfallraten nur gering differieren, zeigen sich bei den feldbedingten Merkmalen deutliche Unterschiede in den Ausfallraten.

Ca. 7% der Haushalte ziehen pro Welle um. Tabelle 4 gibt die Ausfallraten unter den umgezogenen Haushalten an, die an ihrer neuen Adresse erreicht wurden. Es zeigt sich, daß die umgezogenen Haushalte an ihrem neuen Wohnort einem deutlich höheren Ausfallrisiko ausgesetzt sind als die nicht umgezogenen alten Haushalte. Die Diskrepanz in den Ausfallraten steigt mit zunehmender Dauer des Panels. Bedenkt man die nicht unerheblichen Verluste beim Wiedererreichen umgezogener Haushalte (vgl. Tabelle 1), so ergibt sich insgesamt ein hohes feldbedingtes Ausfallrisiko für mobile Haushalte.

Noch schlechter ist die Situation bei den abgespaltenen Haushalten. Hier schwankt die Ausfallrate zwischen 16.9 und 29.2 Prozent, ohne daß sich ein Trend zur Verringerung der Ausfallrate im Laufe des Panels erkennen läßt. Zu diesen hohen Ausfallraten addieren sich nochmals ca. 10 Prozentpunkte aufgrund der Verluste beim Erreichen dieser Haushalte.

Bei dem Merkmal “Neue Befragungsperson im Haushalt” stimmen die beobachteten Ausfallraten keinesfalls mit den Erwartungen überein. Aufgrund der geringeren Teilnahmebereitschaft neuer Befragungspersonen (vgl. Tabelle 2) kann erwartet werden, daß sich diese geringere persönliche Teilnahmebereitschaft auch auf der Haushaltsebene niederschlägt. Tabelle 4 weist aber in Welle 2, 5 und 6 eine geringere Ausfallrate für Haushalte mit neuen Befragungspersonen als für Haushalte ohne neue Befragungspersonen aus. Dieses Ergebnis wird noch fragwürdiger, weil ca. 45% dieser Haushalte abgespaltene Haushalte sind. Offensichtlich ist die Klärung der Haushaltszusammensetzung in vielen Fällen ein guter Indikator für die Antwortbereitschaft des Haushalts¹⁹. Danach muß es den

¹⁹Bei der Bildung dieser Variablen, wurden alle Haushalte, in denen keine neuen Befragungsperson im Kontaktprotokoll genannt wurde, zusammengefaßt. Alternativ hätte man die Haushalte, in denen die Haushaltsgröße als fehlend (missing) eingetragen ist, als Extrakategorie ausweisen können. Die Tatsache, daß die Ausgaben zur Haushaltsgröße fehlen, ist aber meist gleichbedeutend mit dem Ausfall des Haushalts.

Tabelle 4: **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
		Typ des Haushalts						
nicht umgezogener Haushalt	N	5372	4810	4646	4567	4421	4353	4313
	%	9.6	8.0	4.2	5.8	5.6	4.6	3.6
umgezogener Haushalt	N	370	425	373	370	322	277	274
	%	11.6	12.7	10.7	14.1	9.0	7.6	9.9
abgespaltener Haushalt	N	195	238	182	194	183	198	215
	%	29.2	21.4	17.6	23.7	16.9	23.2	21.4
		Neue Befragungsperson im Haushalt						
ja	N	257	243	218	211	209	220	198
	%	9.0	11.1	6.4	6.6	4.8	6.4	6.6
nein	N	5680	5230	4983	4920	4717	4608	4604
	%	10.4	8.8	5.1	7.1	6.3	5.5	4.7
		Wechsel des Interviewers						
ja	N	2041	1203	816	715	826	742	717
	%	14.9	17.5	12.5	19.0	12.9	14.4	10.7
nein	N	3896	4265	4385	4416	4100	4086	4085
	%	7.9	6.5	3.8	5.1	4.9	3.9	3.7
		Anzahl der bereits gewährten Interviews						
vollständig	N	-	5419	5018	4826	4600	4384	4225
	%	-	8.7	4.7	5.9	5.6	4.4	3.6
-1	N	-	-	161	246	253	294	346
	%	-	-	16.7	23.1	14.6	13.9	13.8
-2	N	-	-	-	46	43	73	93
	%	-	-	-	43.5	16.2	23.2	12.9
-3	N	-	-	-	-	24	49	63
	%	-	-	-	-	12.5	14.3	9.5
		Antwortverhalten von verwandten Haushalten						
Keiner vorhanden	N	5538	4731	4252	3962	3606	3301	3064
	%	9.9	8.4	4.2	6.3	6.1	4.5	4.0
Interview verweigert	N	89	123	119	152	135	152	152
	%	14.6	21.1	19.3	15.8	12.6	10.2	9.9
Interview gewährt	N	310	619	830	1017	1185	1370	1586
	%	17.4	10.2	7.9	8.6	6.1	7.3	5.7
		Angaben zum HH-Einkommen fehlen im Vorjahr						
	N	335	310	272	237	210	203	197
	%	17.9	18.4	12.9	16.9	11.4	15.2	12.7

Tabelle 5: Anteil der Haushalte mit Wechsel des Interviewers. Basis: Erreichte Haushalte in der jeweiligen Befragungswelle

Haushaltstyp	Welle						
	2	3	4	5	6	7	8
alte Adresse umgezogener Haushalt	32.2	19.8	14.1	12.1	15.3	13.8	13.5
abgespaltener Haushalt	40.8	30.8	21.2	23.5	25.4	28.5	21.5
abgespaltener Haushalt	79.1	50.8	45.1	39.1	36.1	31.3	34.8

Interviewern in vielen nicht antwortbereiten Haushalten mit Personenzuzügen nicht gelungen sein, überhaupt die Zugehörigkeit neuer Personen zum Haushalt im Kontaktprotokoll zu klären.

Ein Wechsel des Interviewers vergrößert die Ausfallrate um das 3- bis 4-fache. Zwar ist ein Wechsel des Interviewers bei Umzügen über eine große Distanz nicht zu vermeiden, jedoch zeigt Tabelle 5, daß in der Mehrzahl der Umzüge und Haushaltsabspaltungen der Haushalt von dem Interviewer aus der vorhergehenden Befragungswelle betreut wird. Das Gros der Interviewerwechsel (ca. 80%) fand im SOEP bei nicht umgezogenen "alten" Haushalten statt. Die sehr hohe Anzahl von Interviewerwechseln (34% aller Haushalte in Welle 2, 22% aller Haushalte in Welle 3) resultierte aus einer Halbierung des Interviewerstabes, der mit der Durchführung des SOEP beschäftigt war. Diese Verkleinerung des Interviewerstabes wurde mit der Auswahl der motiviertesten und erfolgreichsten Interviewer begründet. Es zeigt sich allerdings, daß eine derartige Selektion des Interviewerstabes während eines Panels nennenswerte nachteilige Effekte auf die Ausfallrate in einem Panel hat. Die Tatsache, daß sich der Anteil der Haushalte mit Interviewerwechsel auf 15% eingespielt hat, weist auf eine permanente Fluktuation im Interviewerstab hin²⁰.

Die Stabilität des Interviewerstammes ist ein spezielles Problem bei der Durchführung von Panelstudien. Üblicherweise ist die Stabilität des Interviewstabs bei den kommerziellen Befragungsinstituten eher gering²¹.

Der Vollständigkeit halber sei noch auf eine weitere Möglichkeit hingewiesen, die einen Wechsel des Interviewers nach sich zieht. In einigen Fällen wird ein neuer Interviewer zu dem betreffenden Haushalt geschickt, wenn der Interviewer der vorherigen Befragungswelle nicht erfolgreich ist. In diesem Fall ist ein Interviewerwechsel also ein Indikator auf einen bevorstehenden Ausfall des Haushalts.

²⁰Hierbei werden größtenteils keine neuen Interviewer eingesetzt, sondern Interviewer aus dem bestehenden Mitarbeiterstamm übernehmen weitere Sample Points.

²¹Beispielsweise gibt Reuband (1984) an, daß nur 50% des Interviewerstammes der meisten Institute länger als ein Jahr dort beschäftigt ist.

Allerdings wird nur bei einem positivem Ausgang, d.h. bei einem Interview, der Interviewerwechsel im Kontaktbericht protokolliert, während bei einem negativen Ausgang der Interviewerwechsel nicht dokumentiert wird. Folglich unterschätzt die Variable "Interviewerwechsel" den negativen Einfluß der Interviewermobilität auf die Antwortbereitschaft.

Bei der Anzahl der bereits gewährten Interviews fällt zunächst auf, daß der Anteil der Haushaltsvorstände, die seit Beginn des Panels kontinuierlich geantwortet haben nur sehr langsam abnimmt. In Welle 8 liegt ihr Anteil noch bei 88% aller Haushalte. Bei den Haushalten, deren Haushaltsvorstand weniger Interviews gegeben haben, steigt die Ausfallrate sprunghaft an. Allerdings zeigt sich in den Wellen 6, 7 und 8 keine deutliche Tendenz, daß die Ausfallrate mit der Anzahl der fehlenden Interviews steigen würde. Dies ist aber — bei aller Vorsicht wegen der geringen Fallzahlen — nicht mit den Werten von Tabelle 4 belegbar.

Beim Antwortverhalten in erhebungstechnisch "verwandten"²² Haushalten zeigt sich bis auf die 2. Welle ein deutlicher Einfluß der Teilnahmebereitschaft über die Haushaltsgrenzen hinweg. Die Ausfallrate der Haushalte verdoppelt sich, wenn der verwandte Haushalt die Mitarbeit in der aktuellen Panelwelle verweigert. Dieser Befund ist dann plausibel, wenn es zu Absprachen zwischen den Haushalten hinsichtlich der Teilnahme am Panel kommt. Da ein Großteil der Haushalte bei Abspaltungen jedoch noch von demselben Interviewer betreut wird, kann die räumliche Distanz zwischen diesen verwandten Haushalten in vielen Fällen nicht groß sein, so daß eine direkte Absprache zwischen diesen Haushalten nicht unplausibel erscheint. Allerdings kann diese Abhängigkeit im Antwortverhalten auch über die Geschicklichkeit des Interviewers vermittelt sein. Weiterhin können Lebensereignisse, die beide Haushalte betreffen, eine Abhängigkeit in der Antwortbereitschaft erzeugen. Diese Situation ist beispielsweise gegeben, wenn sich Partnerschaften auflösen und die früheren Partner in unterschiedliche Haushalte ziehen.

Fehlende Angaben zum Haushaltseinkommen im Vorjahr werden hier zu den felddingten Merkmalen gezählt. In der Tat bestätigen die hohen Ausfallraten um ca. 15%, daß der Itemnonresponse bei der Basisvariablen Haushaltseinkommen ein guter Indikator für eine erhöhte Panelmortalität in der folgenden Befragungswelle ist.

Da die Einkommens- und Vermögensvariablen eine besondere Bedeutung für das SOEP haben, werden in Tabelle 6 die Ausfallraten von erreichten Haushalten hinsichtlich dieser beiden Variablen aufgelistet. Beim Haushaltseinkommen wird eine äquidistante Einteilung in Einkommensklassen gebildet. Bis auf die Haushalte mit Niedrigeinkommen, die etwas höhere Ausfallraten haben, sind die

²²Für jeden Haushalt wurde geprüft, ob es noch einen weiteren Haushalt gibt, so daß beide Haushalte von demselben Stammhaushalt abstammen. Existieren mehrere derartige Haushalte im Bruttobestand, so das Teilnahmeverhalten des jeweils ältesten Haushaltes ausgewählt.

Unterschiede zwischen den Ausfallraten der höheren Einkommensklassen nur sehr gering.

Ca. 13% der Haushalte kreuzen von den 5 möglichen Wertanlageformen keine einzige an. Dies ist in dieser Größenordnung ein unerwartetes Bild, da bereits der Besitz eines Sparbuchs unabhängig von der Größe des Guthabens erfragt wird. Allerdings stimmt die Anzahl der Nennungen "Nein, keine dieser Wertanlagen im Haushalt" ziemlich genau mit der Anzahl der "0" -Vermögensnennungen überein, so daß eher vermuten werden kann, daß bei dieser Antwort Sparbücher mit geringen Guthaben nicht als Wertanlage interpretiert werden. Lediglich in Welle 5 wurde nach der Höhe der Guthaben gefragt.

Mit Ausnahme der Haushalte, die alle 5 Vermögensarten angekreuzt haben, scheint in den Wellen 2, 3 und 4 mit zunehmender Anzahl der Vermögensarten im Haushalt die Ausfallrate zu sinken. Die Ausfallraten für Haushalte mit 5 Vermögensarten sind aufgrund der geringen Fallzahlen nicht sehr stabil.

In der nachfolgenden Analyse werden nun alle Merkmale simultan auf ihren Einfluß auf die Ausfallraten analysiert. Für jede Befragungswelle wurde ein Logit-Modell geschätzt, bei dem die Teilnahmeentscheidung die abhängige Variable ist und die in den Tabellen 3, 4 und 6 genannten Variablen als Haupteffekte eingehen. Da nicht auszuschließen ist, daß sich der Einfluß der Variablen im Verlaufe des Panels verändert, wird das Modell für jede Welle separat geschätzt.

Tabelle 7 gibt die p-Werte des Tests an, der die Nullhypothese überprüft, ob alle freien Koeffizienten zu einer Variablen den Wert 0 haben. Bei der Bewertung der Signifikanzen ist zu beachten, daß sowohl Effekte des multiplen Tests als auch die Richtung der geschätzten Effekte berücksichtigt werden müssen (vgl. hierzu etwa die Diskussion bei Cox/Hinkley, 1979, S.77 ff). Da für alle 7 Zeitpunkte dieselbe Hypothese ("Kein Einfluß der Variablen auf die Antwortbereitschaft") getestet wird, ist der Sachverhalt des multiplen Testens gegeben. Benutzt man die Regel, daß H_0 abgelehnt wird, wenn nur ein Test H_0 ablehnt, so ist das Gesamt-Fehlerniveau nicht durch den p-Wert dieses Tests sondern approximativ durch den siebenfachen p-Wert gegeben.

Allerdings schwankt das multiple Testniveau sehr stark in Abhängigkeit von der Formulierung der Null-Hypothese, wobei die interpretatorischen Unterschiede zwischen den unterschiedlichen Formulierungen nur gering sind. Lautet die Null-Hypothese beispielsweise "Kein Einfluß der sozio-demographischen Basisvariablen Stichprobe, Geschlecht, Alter, Familienstand, Schulbildung und Stellung im Beruf", so kann man das dieser Null-Hypothese entsprechende multiple Testniveau dadurch approximieren, indem man dem geringsten p-Wert für die Einzelvariablen mit der Anzahl der Variablen und der Anzahl der Wellen multipliziert. Mit den Werten von Tabelle 7 erhöht man beispielsweise die Approximation $0.042 = 6 \times 7 \times 0.001$.

Neben der formalen Beurteilung durch einen p-Wert muß allerdings auch die Richtung der geschätzten Effekte in Hinblick auf die Erwartungen bei der Formalisierung der Hypothesen berücksichtigt werden.

Tabelle 6: **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
		Haushaltseinkommen						
-1000	N	456	368	293	270	241	213	182
	%	13.8	10.1	5.5	9.3	7.1	6.1	6.6
1000-2000	N	1816	1521	1383	1243	1140	995	870
	%	11.2	9.7	5.3	6.6	6.2	4.8	5.1
2000-3000	N	1713	1572	1469	1404	1354	1329	1260
	%	8.3	7.6	4.2	6.5	6.6	4.6	4.0
3000-4000	N	992	996	1008	1087	1060	1073	1069
	%	9.6	6.4	4.3	6.1	5.0	4.7	3.7
4000+	N	625	706	776	890	921	1015	1224
	%	8.2	8.9	5.0	6.4	5.9	6.2	4.7
		Anzahl der Vermögensarten im Haushalt						
0	N	823	769	743	735	578	661	573
	%	12.5	12.3	6.9	11.3	8.3	8.8	6.6
1	N	1714	1561	1468	1429	1431	1262	1256
	%	13.2	10.4	5.1	6.8	6.7	5.5	4.4
2	N	1709	1549	1427	1449	1444	1310	1350
	%	8.3	8.2	5.6	5.8	5.9	5.2	5.0
3	N	1224	1161	1134	1122	1107	1152	1201
	%	8.7	6.7	3.9	5.9	5.7	4.1	3.7
4	N	403	388	374	343	326	377	375
	%	7.7	5.9	3.7	7.0	4.6	5.3	4.5
5	N	64	45	55	53	40	66	47
	%	9.4	6.7	9.1	11.3	2.5	7.5	14.9

Tabelle 7: p-Werte für den Einfluß unterschiedlicher Merkmale auf das Teilnahmeverhalten von SOEP-Haushalten

Variable	p-Wert in						
	Welle 2	Welle 3	Welle 4	Welle 5	Welle 6	Welle 7	Welle 8
Klassische Prüfmerkmale							
Stichprobe A/B	0.5	0.1	0.05	0.2	0.3	0.03	0.2
Geschlecht	0.8	0.6	0.5	0.2	0.7	0.2	0.8
Alter	0.01	0.1	0.1	0.2	0.4	0.01	0.05
Familienstand	0.7	0.3	0.4	0.4	0.2	0.4	0.001
Schulbildung	0.001	0.1	0.3	0.3	0.03	0.2	0.4
Stellung im Beruf	0.5	0.4	0.3	0.6	0.08	0.004	0.8
Feldbedingte Prüfmerkmale							
Haushaltstyp Wechsel des Interviewers	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
Neue Person im Haushalt	<0.001	0.1	0.003	<0.001	0.003	0.001	0.001
Teilnahmedauer am SOEP	-	0.13	0.002	<0.001	<0.001	<0.001	0.003
Antwort in verw. Haushalten	0.03	<0.001	<0.001	0.02	0.001	0.3	0.02
Antwort in verw. Haush.: Altern. 1	-	0.7	<0.001	0.002	0.06	0.1	0.03
Antwort in verw. Haush.: Altern. 2	0.7	0.006	0.1	0.2	0.9	0.2	0.2
Einkommensvariablen							
Haush.Einkommen	<0.001	<0.001	0.002	0.001	0.09	0.001	<0.001
Beträge Einkommen	0.10	0.3	0.9	0.9	0.4	0.7	0.8
Trend i.d. Anzahl der Vermögensanl.Formen	0.01	0.01	0.2	0.2	0.04	0.3	0.5
Anzahl = 0	0.2	0.9	0.4	0.07	0.9	0.2	0.2
Anzahl der Vermögensanl.Formen	0.02	0.03	0.2	0.005	0.2	0.08	0.03

Auch wird man bei gleichen p-Werten den Einfluß einer Variablen bei über die Zeit gleichgerichteten Koeffizienten anders beurteilen als bei ungleich gerichteten Koeffizienten. Zusätzlich wird eine Bewertung des Einflusses von Kovariaten auf die Antwortbereitschaft dadurch erschwert, daß es durchaus möglich ist, daß eine Variable nur während bestimmter Phasen des Panels einen Effekt hat. Schließlich liegt auch die Anzahl der Beobachtungen mit ca. 5000 relativ hoch, so daß auch noch relativ kleine Effekte als "signifikant" bewertet werden.

Bei der Diskussion der Größe der Koeffizienten sollte man beachten, daß diese nur angeben, wie stark sich der Logarithmus des Verhältnissen von Ausfällen zu Antworten verändert. Ist das durchschnittliche Ausfallniveau gering, z.B. 5%, so bedeutet die Vergrößerung von Ausfällen zu Antworten um den Faktor $e^{0.5} = 1.6$ lediglich einen Anstieg der Ausfallraten auf $7.8\% \approx 1.6 \times 5\%$. Aus diesem Grund implizieren dieselben Koeffizienten bei Beginn (Welle 2, 3) und bei Ende (Welle 7, 8) der betrachteten Perioden des SOEP doch unterschiedliche Wirkungen hinsichtlich des Absolutniveaus der Ausfallraten. Während in Welle 2 für Durchschnittshaushalte ein Anstieg des Logits um den Faktor 1.6 die Zunahme der Ausfallrate von 10% auf 16% impliziert, bedeutet dies in Welle 8 lediglich eine Zunahme von 4.7% auf 7.5%.

Hinsichtlich der klassischen Basisvariablen zeigt sich in Tabelle 7 bei den Merkmalen "Stichprobe" und "Geschlecht" in keiner Welle ein deutlicher Effekt auf die Ausfallrate. Bei der Variablen "Familienstand" gibt es jeweils nur eine Welle, in der die Variable einen deutlichen Effekt auf die Ausfallneigung hat. Betrachtet man die geschätzten Koeffizienten in Tabelle 8, so ist es allerdings wenig plausibel, warum ausgerechnet in Welle 8 das Verhältnis von Ausfällen zu Antworten um den Faktor $e^{(0.8+0.4)} = 3.3$ steigt, wenn man von verheiratet-Zusammenlebenden zu verheiratet-Getrenntlebenden übergeht. In Welle 4 und 6 bestehen dagegen keine Unterschiede zwischen diesen Gruppen im geschätzten Verhältnis von Ausfällen zu Antworten.

Ähnliche Überlegungen gelten für die Variable "Stellung im Beruf", die nur in Welle 7 signifikant ist. Der auffälligste Unterschied im Ausfall-/Antwortverhältnis besteht hier zwischen den Arbeitslosen und den Personen mit einer hohen Stellung im Beruf. Dieses Verhältnis steigt um den Faktor $e^{(0.7+1.2)} = 6.6$, wenn man in Welle 7 von den Arbeitslosen zu den Personen mit hoher Position übergeht. Doch schon eine Welle später liegt der Faktor nur noch bei 1.5; in den Wellen 4 und 5 fast bei 1. Auch hinsichtlich des erwarteten "Mittelstands-Bias" findet man nur in Welle 7 eine Unterstützung der Hypothese, daß die "Mittelständler" niedrigere Ausfallraten als Personen in hohen oder niedrigen Positionen haben.

Die Variable "Schulbildung" liefert für die 2. und 6. Welle signifikante Beiträge für das Antwortverhalten. Hier sind in der 6. Welle die Unterschiede zwischen Personen mit Realschulabschluß und Abitur am größten. Das Verhältnis von Ausfällen zu Antworten sinkt von den Realschülern zu den Abiturienten um den Faktor $e^{(-0.5-0.4)} = 0.4$. Doch in den Wellen 2, 4, 7 und 8 ist dieses Verhältnis nahezu identisch. Auch hinsichtlich der ursprünglichen Hypothesen

Tabelle 8: **Geschätzte Koeffizienten für den Einfluß sozio-demographischer Variablen auf die Ausfallrate von SOEP-Haushalten**

	Koeffizienten in Welle						
	2	3	4	5	6	7	8
Alter							
75 u. älter	0.5	0.4	0.3	0.2	0.4	0.3	-0.2
65-74	-0.1	0.1	-0.6	-0.1	0.2	-0.2	-0.1
55-64	0.0	-0.1	-0.1	-0.2	0.0	0.5	0.3
35-54	-0.1	-0.2	0.0	0.0	0.0	0.2	0.5
25-34	-0.3	-0.2	0.1	0.2	-0.1	0.2	0.1
16-24	0.0	0.0	0.3	-0.1	-0.5	-1.0	-0.6
Familienstand							
verheiratet, zusammen	0.2	-0.1	0.0	-0.1	0.1	-0.2	-0.4
verheiratet, getrennt	-0.2	0.3	0.0	0.6	0.1	0.1	0.8
ledig	0.2	-0.3	0.3	0.1	-0.2	0.2	0.4
geschieden	0.1	0.0	0.0	0.0	0.1	0.1	-0.2
verwitwet	0.0	0.6	-0.3	-0.2	-0.6	-0.3	0.1
Ehepartner im Ausland	-0.3	-0.4	0.0	-0.4	0.5	0.0	-0.6
Schulbildung							
Kein Abschluß	0.1	0.1	0.4	0.3	0.0	-0.3	0.2
Hauptschule	0.3	0.3	-0.3	-0.2	0.2	0.3	0.1
Realschule	-0.3	0.2	-0.2	0.1	0.4	0.3	0.0
Fachhochschule	0.1	0.0	-0.2	0.0	0.1	0.1	0.3
Abitur	-0.3	-0.1	-0.3	-0.2	-0.5	0.2	-0.1
sonstiger Abschluß	0.1	-0.3	0.6	0.0	-0.2	-0.6	-0.5
Stellung im Beruf							
nicht erwerbstätig	0.1	-0.2	0.3	-0.1	-0.2	0.1	0.1
arbeitslos	-0.3	0.2	0.1	0.0	-0.4	-1.2	-0.2
hoher Status	0.1	-0.1	-0.1	0.1	0.5	0.7	0.2
mittlerer Status	0.0	0.0	0.0	-0.1	0.1	0.1	0.0
niedriger Status	0.1	0.1	-0.3	0.1	0.0	0.3	-0.1

vom Bildungsbias zeigt sich nicht das erwartete Bild. Sieht man von den Personen mit "sonstigen Schulabschlüssen" ab, so ist in Welle 7 die Ausfallrate für Personen ohne Abschluß am geringsten. In Welle 6 ist sie für Realschüler am größten. In keiner Welle zeigt sich eine Monotonie der Ausfallraten nach dem Bildungsniveau.

Als letzte der sozio-demographischen Basisvariablen liefert die Variable "Alter" in 3 Wellen signifikante Beiträge. Da der kleinste p-Wert bei 0.01 liegt, erreicht das multiple Testniveau nur den Wert $0.01 \times 7 = 0.07$. Allerdings offenbart die eine Diskussion der geschätzten Koeffizienten, daß dieses Merkmal durchaus, geeignet ist, einige Aspekte von Ausfällen zu erklären. Zunächst zeigt sich, daß die Gruppe der über 75-jährigen bis Welle 7 ein erhöhtes Ausfallrisiko hat. Die Größe dieses Risikos nimmt in der Tendenz ab. In Welle 8 ist das Ausfallrisiko dieser Gruppe unterdurchschnittlich. Dieser Effekt kann daher rühren, daß mittlerweile viele der Personen aus der Altersklasse 65-74 die Altersgrenze von 75 Jahren überschritten haben. In der Altersgruppe 65-74 ist die Ausfallrate über alle Panelwellen eher unterdurchschnittlich, so daß diese niedrige Ausfallneigung auf die nächste Altersgruppe "vererbt" wurde.

Weiterhin sinkt die Ausfallrate für die Altersgruppe 16-24 seit Welle 5. Auch dieses Ergebnis ist plausibel, da bei dieser Personengruppe ein wachsender Anteil von "Kindern" aus Panelhaushalten ist, die schon über längere Zeit Kontakt mit dem Panel haben²³. Nicht zuletzt gibt die Variable "Alter" ziemlich deutlich den Effekt der massiven Werbung für die Volkszählung während der Feldarbeit für die 4. Welle wieder. Diese Kampagne hatte offensichtlich in der Gruppe der 65 bis 74-jährigen einen Antwort-steigernden Effekt, während die Resonanz bei den 16 bis 24-jährigen eher negativ war.

Kommen wir nun zu den feldbedingten Variablen. Die Signifikanz dieser Merkmale für das Antwortverhalten steht außer Frage (vgl. Tabelle 7). Auch weisen die geschätzten Koeffizienten in Tabelle 9 gleichmäßig über alle Wellen auf dieselbe Richtung der Effekte hin. Lediglich die Größe der Effekte schwankt etwas im Verlaufe des Panels.

Zunächst weisen die Schwankungen der Konstanten in Tabelle 9 darauf hin, daß mehrere Variablen des Modells überwiegend auf einer Ausprägung konzentriert sind und daher Kollinearitätsprobleme auftauchen. Dies gilt insbesondere für die Variable Teilnahmedauer, die in das Modell als metrischen Variable mit einem Koeffizienten aufgenommen wurde. Wegen der Kollinearität mit der Konstanten wurde diese Variable in der 2. Welle nicht in das Modell aufgenommen.

Beim Haushaltstyp liegt das geschätzte Verhältnis von Ausfällen zu Antworten bei den umgezogenen und abgespaltenen Haushalten deutlich über dem entsprechenden Verhältnis bei den nicht umgezogenen Haushalten. Der Vergrößerungsfaktor gegenüber den alten, nicht umgezogenen Haushalten liegt bei den

²³Dieser Effekt wird nicht über die Variable Teilnahmedauer gemessen, weil diese Variable nur die Anzahl der gegebenen Interviewers, nicht jedoch den Eintritt ins Panel mißt.

Tabelle 9: Geschätzte Koeffizienten für den Einfluß von felddingten Variablen auf die Ausfallrate

	Koeffizienten in Welle						
	2	3	4	5	6	7	8
Konstante	-1.9	-0.3	-0.5	1.8	1.0	-0.4	-0.7
	Typ des Haushalts						
Alter Haushalt	-1.1	-0.6	-0.5	-0.9	-0.8	-0.7	-1.0
Umgezogener Haushalt	-0.7	-0.1	0.2	-0.1	-0.3	-0.4	0.0
Abgespaltener Haushalt	1.8	0.7	0.3	1.0	1.1	1.1	1.0
	Wechsel des Interviewers						
ja	0.3	0.5	0.6	0.7	0.5	0.6	0.4
nein	-0.3	-0.5	-0.6	-0.7	-0.5	-0.6	-0.4
	Neue Befragungsperson im Haushalt						
ja	-0.5	-0.2	-0.5	-1.0	-0.6	-0.5	-0.6
nein	0.5	0.2	0.5	1.0	0.6	0.5	0.6
	Teilnahmedauer						
pro Welle	-	-0.5	-0.6	-1.0	-0.6	-0.5	-0.2
	Antwort im verwandten Haushalt.						
Keiner vorhanden	0.4	0.0	-0.5	0.0	-0.1	-0.1	-0.1
Interview verweigert	0.3	0.6	0.9	0.4	0.6	0.3	0.5
Interview gewährt	-0.7	-0.6	-0.4	-0.4	-0.5	-0.2	-0.4
	Antwort im verwandten Haushalt						
	Alternative 1						
Keiner vorhanden	-	0.0	-0.6	-0.2	-0.1	-0.2	0.1
Interview verweigert	-	0.1	1.0	0.6	0.4	0.4	0.4
Interview gewährt	-	-0.1	-0.4	-0.4	-0.3	-0.2	-0.5
	Antwort im verwandten Haushalt						
	Alternative 2						
Interview verweigert	0.1	0.7	0.4	0.3	0.0	0.3	0.3
Interview gewährt	-0.1	-0.7	-0.4	-0.3	0.0	-0.3	-0.3

umgezogenen Haushalten bei ca. $e^{0.5} = 1.6$, und bei den abgespaltenen Haushalten sogar bei ca. $e^{1.7} = 5.5$. Lediglich in der 2. und 4. Welle weicht bei den abgespaltenen Haushalten die Vergrößerung des Ausfallrisiko von diesen Werten nennenswert ab. In der zweiten Welle vergrößert sich das Ausfallrisiko um den Faktor $e^{2.9} = 18.1$, während es in der 4. Welle nur um den Faktor $e^{0.8} = 2.2$ ansteigt. Beide Abweichungen erscheinen plausibel: Der hohe Faktor in der 2. Welle aufgrund der bisher kurzen Laufzeit des Panels, der niedrige Faktor in der 4. Welle aufgrund der externen Sonderbedingungen (Werbung für Volkszählung) während der 4. Welle, die die Panel-spezifischen Teilnahmegründe nivellieren.

Der Wechsel des Interviewers beeinflusst das Ausfall-/Antwort-Verhältnis in allen Wellen deutlich negativ. Dabei zeigt sich eine nicht un plausible Tendenz zur Verstärkung dieses Effekt bis zur 5. Welle (Anstieg von $e^{0.6} = 1.82$ auf $e^{1.4} = 4.1$). Da die meisten Interviewerwechsel durch Ausfall eines Interviewers entstehen, der die Befragten seit Beginn des Panels betreut, ist der Vertrauensverlust umso größer, je länger der Interviewer die Befragten vor dem Wechsel befragt hat. Zusätzlich baute die 5. Befragungswelle mit dem Schwerpunkt "Vermögen" besonders stark auf dem Vertrauen der Befragten zum Interviewer auf. Dies bestätigten auch zahlreiche Feldberichte der Interviewer²⁴. Allein aus diesem Grund ist der maximale Effekt eines Interviewerwechsels in Welle 5 verständlich. Vor dem Hintergrund voraussehbarer weiterer Interviewerwechsel ist die Abflachung des Negativ-Effektes eines Wechsels in den Wellen 6 bis 8 erfreulich.

Die Tatsache, daß neue Befragungspersonen im Kontaktprotokoll aufgeführt werden, weist in allen Befragungswellen auf eine deutlich erhöhte Antwortbereitschaft dieser Haushalte hin. Dieser Befund, der aus den univariaten Vergleichen nicht mit dieser Deutlichkeit ablesbar ist, zeigt, daß die Kontaktprotokolle der ausgefallenen Haushalte nur hinsichtlich der "alten" Befragungspersonen vollständig sind. Allerdings hat dieses Merkmal kaum noch eine Aussagekraft im Hinblick auf die Erklärung von Ausfällen.

Beim Merkmal "Teilnahmedauer" zeigen die negativen Koeffizienten an, daß mit wachsender Anzahl von Interviews, die der Haushaltsvorstand gegeben hat, die Ausfallraten sinken. Allerdings ist die Anzahl der Haushalte in denen der Haushaltsvorstand weniger als die maximal mögliche Anzahl von Interviews gegeben hat, in den ersten Panelwellen noch sehr gering, vgl. Tabelle 4. Die Variable mißt also eher den Effekt, ob der Haushaltsvorstand schon in der ersten Welle an der Befragung teilgenommen hat oder nicht. Erst in der 8. Welle differenzieren sich die Haushalte etwas stärker nach der Anzahl der Interviews. In dieser Welle nimmt die Größe des Koeffizienten für den Effekt je Welle jedoch deutlich ab (von ca. -0.5 auf -0.2). Dies kann als Hinweis darauf verstanden werden, daß

²⁴Zahlreiche Interviewer berichteten, daß dieser Schwerpunkt von den Befragten als "Zumutung" empfunden wurde und die Befragten nur aus Rücksichtnahme gegenüber dem Interviewer weiterhin an der Befragung teilnahmen.

die Zunahme der Antwortbereitschaft schwächer als linear mit der Anzahl der Interviews steigt. Diese Interpretation stellt in guter Übereinstimmung mit den univariaten Ergebnissen in Tabelle 4.

Beim "Antwortverhalten in verwandten Haushalten" zeigt sich in allen Wellen ein deutlich negativer Effekt auf die Teilnahmebereitschaft, wenn in einem verwandten Haushalt die Teilnahme am Panel verweigert wird. Die in Tabelle 7 wiedergegebenen Signifikanzniveaus beziehen sich auf den Test, ob das Antwortverhalten eines verwandten Haushalts vernachlässigt werden kann. Die Größe dieses Effekts ist beachtlich: Das Ausfall- /Antwort-Verhältnis vergrößert sich ungefähr um den Faktor $e^1 = 2.7$, wenn der verwandte Haushalt statt zu antworten die Teilnahme am Panel verweigert. Im Gegensatz zu der univariaten Darstellung in Tabelle 4 sind in der multiplen Analyse auch in der 2. Welle die Effekte einer Antwortverweigerung eines verwandten Haushalts negativ.

Wir wollen im folgenden prüfen, wie diese Abhängigkeit im Antwortverhalten zwischen den Haushalten erzeugt wird. Wenn diese Abhängigkeit nur über gemeinsame Lebensereignisse vermittelt wird, so betrifft dies im wesentlichen Ereignisse, die mit der Abspaltung von Haushalten verbunden sind, z.B. die Trennung von Paarbeziehungen. Würde die Abhängigkeit nur über die gemeinsamen Lebensereignisse induziert, so müßte das Antwortverhalten in verwandten Haushalten insignifikant werden, wenn man die abgespaltenen Haushalte und deren Stammhaushalte von der Analyse ausschließt. Die Ergebnisse einer derartigen Analyse sind in Tabelle 7 unter "Alternative 1" aufgelistet. Zunächst einmal treten verwandte Haushalte nach diesem modifizierten Konzept frühestens ab Welle 3 auf. Auch ist in dieser Welle die Zahl der verwandten Haushalte noch sehr klein. Ab Welle 4 jedoch ist das Antwortverhalten in verwandten Haushalten auch in diesem modifizierten Konzept signifikant für die Teilnahmebereitschaft. Die geschätzte Größe des Effektes stimmt gut mit den ursprünglichen Ergebnissen überein.

Bezieht man umgekehrt nur die neuen Haushalte in die Analyse ein und betrachtet die Abhängigkeit der Ausfälle vom Antwortverhalten des Stammhaushalts, so hat auch dieses einen signifikanten Einfluß auf die Ausfallrate²⁵. Die geschätzte Größe des Effektes ($e^1 = 2.7$) entspricht ziemlich genau dem Effekt bei den übrigen Haushalten unter der Alternative 1. Man muß daher die Hypothese, daß die Interaktion im Teilnahmeverhalten verwandter Haushalte nur über gemeinsame Lebensereignisse vermittelt wird, ablehnen.

Wir wollen nun prüfen, ob die Abhängigkeit im Antwortverhalten nur durch einen gemeinsamen Interviewer induziert wird. Tabelle 10 gibt einen Überblick über das Antwortverhalten in verwandten Haushalten. Hierbei wird danach unterschieden, ob die Haushalte von einem gemeinsamen Interviewer betreut wer-

²⁵Wegen der geringen Fallzahlen bei den abgespaltenen Haushalten wurde ein "gepooltes" Modell geschätzt, d.h. es wurde ein Modell mit zeitabhängigen Konstanten für alle Haushalte geschätzt.

Tabelle 10: **Einfluß des Antwortverhaltens in verwandten Haushalten auf die Ausfallraten.** Vereinigung der Fälle über die Wellen 2 bis 8

Antwortverhalten in verwandten Haushalten	Anzahl	Ausfallrate
	Haushalte mit gemeinsamem Interviewer	
Kein Interview	362	14.9
Mit Interview	4237	5.8
	Haushalte ohne gemeinsamen Interviewer	
Kein Interview	565	14.2
Mit Interview	2680	10.8

den. Es zeigt sich, daß in etwas weniger als der Hälfte der Fälle die verwandten Haushalte nicht durch einen gemeinsamen Interviewer betreut wurden. Wenn die Beeinflussung nur über den Interviewer hervorgerufen wird, wäre zu erwarten, daß sich das Teilnahmeverhalten in Haushalten ohne gemeinsamen Interviewer nicht beeinflußt. In Tabelle 10 ist aber bei diesen Haushalten die Ausfallrate deutlich größer (14.2%), wenn der verwandte Haushalt nicht auskunftsbereit ist, als wenn dieser ein Interview gegeben hat (10.8%). Zwar ist die Differenz zwischen den Ausfallraten bei den Haushalten mit gemeinsamen Interviewern deutlich größer (14.8% bei Antwortverweigerung zu 5.8% bei Antwortgewährung) als bei den Haushalten ohne gemeinsamen Interviewer, doch ist diese Abnahme des Einflusses schon allein aufgrund der größeren regionalen Distanz bei den Haushalten ohne gemeinsamen Interviewer zu erwarten.

Um diesen Befund auch im Zusammenhang mit den anderen Kovariaten auf seine Signifikanz zu prüfen, wurde die multiple Ausfallanalyse für die Population der verwandten Haushalte ohne gemeinsamen Interviewer wiederholt. Die Ergebnisse dieser Analyse sind in Tabelle 7 und 9 unter "Alternative 2" aufgelistet. Aufgrund der geringen Fallzahlen in den einzelnen Wellen wurden ein gemeinsames Modell für alle Beobachtungen von Welle 2 bis Welle 8 geschätzt, wobei die einzelnen Erhebungswellen über eine Konstante berücksichtigt wurde. Man erhält als Resultat, daß der Einfluß die Antwortverhalten im verwandten Haushalt einen deutlichen Einfluß hat (p -Wert < 0.001). Der geschätzte Effekt der Vergrößerung des Ausfall-/Antwort-Verhältnisses durch eine Antwortverweigerung im verwandten Haushalt ist $e^{2 \times 0.31} = 1.8$. Dieser Faktor ist etwas geringer als bei Einbeziehung der verwandten Haushalte mit gemeinsamen Interviewer ($e^{2 \times 0.5}$). Dies läßt darauf schließen, daß die Beeinflussung bei Haushalten mit gemeinsamen Interviewer stärker ist. In der Tat erhält man bei einer separaten Analyse dieser Haushalte den Faktor $e^{2 \times 0.6}$. Somit muß auch die Hypothese abgelehnt werden, daß die Interaktion im Teilnahmeverhalten verwandter Haushalte nur über den gemeinsamen Interviewer vermittelt wird.

Tabelle 11: Geschätzte Koeffizienten für den Einfluß von Einkommens- und Vermögensvariablen auf die Ausfallrate

	Welle						
	2	3	4	5	6	7	8
	Haushaltsnetto						
Keine Angaben	0.6	0.6	0.8	0.7	0.5	0.8	0.9
bis 1000	0.2	-0.1	-0.3	-0.1	0.1	0.0	0.0
1000 bis 2000	-0.1	-0.1	-0.1	-0.1	0.0	-0.2	0.0
2000 bis 3000	-0.3	-0.2	-0.2	-0.1	0.0	-0.2	-0.3
3000 bis 4000	-0.1	-0.3	-0.1	-0.1	-0.3	-0.3	-0.3
4000 und mehr	-0.3	0.1	-0.1	-0.3	-0.3	0.0	-0.3
	Anzahl der Anlageformen im Haushalt						
pro Nennung	-0.14	-0.16	-0.11	-0.08	-0.15	-0.08	0.06
Anzahl = 0	0.1	0.4	0.0	0.4	0.5	0.4	0.0
Anzahl = 1	0.3	0.2	0.1	0.0	0.3	0.0	-0.3
Anzahl = 2	-0.2	0.0	0.1	-0.2	0.2	-0.1	-0.1
Anzahl = 3	-0.1	-0.2	-0.3	-0.3	0.1	-0.4	-0.4
Anzahl = 4	-0.2	-0.3	-0.4	-0.3	-0.3	-0.1	-0.2
Anzahl = 5	0.1	0.1	0.5	0.4	-0.8	0.2	1.0

Als letzter Block werden in Tabelle 7 die Signifikanz der Einkommens- und Vermögensvariablen wiedergegeben. Die entsprechenden Schätzwerte sind in Tabelle 11 aufgelistet. Zunächst scheint das Haushaltseinkommen in allen Befragungswellen einen deutlichen Einfluß auf die Teilnahmebereitschaft zu haben. Eine Inspektion der zugehörigen Schätzwerte zeigt aber, daß lediglich die fehlende Nennung des Haushaltseinkommens in der vorherigen Welle eine deutliche Wirkung auf Ausfallrate hat. Die Koeffizienten für die Einkommensbeträge zeigen nur geringe Unterschiede, die auch nicht über die Zeit stabil sind. In der Tat sind die Unterschiede zwischen den Koeffizienten für die Einkommensbeträge zu keinem Zeitpunkt signifikant. Die p-Werte des entsprechenden Tests werden in Tabelle 7 in der Zeile "Beiträge-Einkommen" ausgewiesen. Die Analyse bestätigt, daß der Item- Nonresponse beim Haushaltseinkommen ein guter Indikator für ein erhöhtes Ausfallrisiko ist. Das Verhältnis von Ausfällen zu Antworten erhöht sich hierdurch um ca. $e^{0.8} = 2.2$.

Bei der Anzahl der Wertanlageformen liegt der kleinste p-Wert in Tabelle 7 bei 0.01. Approximiert man das multiple Testniveau mit 7×0.01 , so liegt dieser Wert in einem Grenzbereich üblicher Signifikanzniveaus. Allerdings muß hier beachtet werden, daß die Schätzung auf einen linearen Trend mit nur einem Koeffizienten restringiert war. In der Tendenz sind Haushalte umso auskunftsbereiter, je mehr unterschiedliche Formen der Wertanlage in einem Haushalt vorhanden

sind. Allerdings wird in Welle 8 dieser Trend genau umgekehrt geschätzt, so daß doch Zweifel an einem deutlichen Vermögenseffekt bei der Antwortbereitschaft bestehen bleiben. In das Modell wurde zusätzlich eine Variable aufgenommen, die überprüfen soll, ob die Nennung keiner einzigen Anlageform für den Haushalt wie eine "Keine Angabe" Nennung beim Haushaltseinkommen interpretiert werden kann. Dies müßte zu einer zusätzlichen Erhöhung der Ausfallneigung über den linearen Trend hinaus führen. Allerdings ist diese Variable (Bezeichnung "Anzahl = 0" in Tabelle 7) in keiner Welle signifikant. Dieses Resultat ist auch plausibel, da die Anzahl der Nennungen mit keiner Vermögensart gut mit der Anzahl der Nennungen "Nein, keine dieser Wertanlagen" übereinstimmt. Es handelt sich in den allermeisten Fällen also nicht um einen Item-Nonresponse.

In einem letzten Analyseschritt soll die Annahmen fallen gelassen werden, daß die Anzahl der Anlageformen linear in die Ausfallneigung eingeht. Hier zeigt schon die univariate Verteilung in Tabelle 6, daß die Nennung aller 5 Anlageformen, die immer Betriebsvermögen einschließt, sich nicht in einen linearen Trend einpaßt. Diese Sonderrolle wird auch durch die multiple Analyse (Zeile "Anzahl Anlageformen" in Tabelle 7 bzw. Zeile "Anzahl = 5" in Tabelle 11) bestätigt, wo diese Haushalte bis auf Welle 6 eine unterdurchschnittlichen Antwortbereitschaft haben. Diese Ausnahme ist im Zusammenhang mit dem Befragungsschwerpunkte Vermögen in der 5. Welle zu sehen. Die Beantwortung dieser Fragen bedeutete für diese Haushalte eine hohe Belastung. Daher zeichnen sich Haushalte, die trotz dieser hohen Belastung an der 5. Welle teilgenommen haben, durch eine hohe Antwortbereitschaft aus. Konsequenterweise sind diese Haushalte auch in der Nachfolgewelle überdurchschnittlich antwortbereit, was durch den negativen Koeffizienten(-0.8) in Tabelle 11 angezeigt wird.

Allerdings ist die Anzahl der Haushalte mit 5 Wertanlageformen relativ gering (ca. 50 Haushalte) und keineswegs mit der Anzahl der Haushalte mit Betriebsvermögen identisch, deren Zahl in der Stichprobe etwa 4-5mal so hoch ist. Hinsichtlich des Teilnahmeverhaltens ist die Ausfallrate der Haushalte mit Betriebsvermögen verschieden von den "Anzahl = 5"-Haushalten²⁶. Bezieht man die einzelnen Anlageformen in die multiple Analyse mit ein, so erweisen sich diese als nicht signifikant (Ausnahme: Welle 6, Betriebsvermögen). Es kommt also mehr auf die Kumulation der Anlageform als auf die Anlageformen selbst an. Durchgängig sind Haushalte mit 3-4 Anlageformen etwas auskunftsbereiter als die übrigen Haushalte. Man kann dieses Ergebnis im Sinne eines "Mittelstandbias" interpretieren. Allerdings sind die absoluten Unterschiede in der Antwortbereitschaft (gemessen in Prozentpunkten) in der Antwortbereitschaft doch ziemlich gering (vgl. Tabelle 6). Gemessen am Bestand der Stichprobenmitglieder des SOEP hat sich daher die Verteilung der Haushalte nach Anzahl der Anlageformen nur geringfügig verändert (Vgl. Tabelle 12). So hat der Anteil

²⁶In Welle 8 beträgt die Ausfallrate der 215 Haushalte mit Betriebsvermögen beispielsweise 7.4% gegenüber 14.9% bei den 47 "Anzahl = 5"-Haushalten.

Tabelle 12: Vergleich der Verteilung der SOEP-Haushalte nach der Anzahl der Anlageformen

Anzahl	Anteil in	
	Welle 2	Welle 8
0	13.8	11.9
1	28.8	26.1
2	28.8	28.1
3	20.6	25.0
4	6.8	7.8
5	1.1	1.0

der Haushalte ohne Wertanlage nur um 1.9 Prozentpunkte abgenommen und der Anteil der Haushalte mit allen Anlageformen ist nahezu konstant geblieben²⁷. Dieses Beispiel zeigt, daß die ausschließliche Betrachtung der Koeffizienten des Logitmodells ohne Bezug auf das absolute Niveau der geschätzten Ausfallraten leicht zu Fallschlüssen verleitet.

Da die Anzahl der Anlageformen im Haushalt jedoch relativ stabil über die Zeit ist, bedeutet der beobachtete Trend im Teilnahmeverhalten jedoch keine Einschränkung für die Gültigkeit von Populationsschätzungen, solange die Ausfallwahrscheinlichkeiten entsprechend dem hier geschätzten Modell in die Hochrechnungen eingehen.

5 Vergleiche von Schätzwerten und Registerangaben für demographische Ereignisse

Geburt, Tod und die Trennung von Ehepaaren sind Ereignisse, die häufig mit der Neugründung von Haushalten, Umzügen bzw. der Aufspaltung von Haushalten in Verbindung gebracht werden können. Die Ergebnisse der vorhergehenden Abschnitte haben gezeigt, daß Haushaltsmobilität mit einem deutlich erhöhten Ausfallrisiko einhergeht. In diesem Abschnitt soll mit dem Mittel des Vergleichs von Stichprobe und Population geprüft werden, ob dieses erhöhte Ausfallrisiko homogen bezüglich dieser Lebensereignisse ist. Unter der Hypothese, daß nicht das spezielle Lebensereignis sondern lediglich die Haushaltsmobilität einen Einfluß auf das Teilnahmeverhalten hat, ist die Schätzung der Anzahl dieser Lebensereignisse erwartungstreu, da die Variable "Haushaltstyp" bei der Berechnung der Auswahlwahrscheinlichkeiten berücksichtigt wurde.

²⁷Auch hat mittlerweile die Popularität der Anlageform "Wertpapier" zugenommen, so daß allein aus diesem Grund die Anzahl der Anlageformen in Haushalten gestiegen sein sollte.

Tabelle 13: Vergleich der Schätzung demographischer Ereignisse über das SOEP mit entsprechenden Registerangaben. Angaben in TSD

Jahr	Geburten		Todesfälle		Ehescheidungen			Register
	SOEP	Register	SOEP	Register	SOEP	M	F	
1984	480	584	423	696	85	106	95	130
1985	657	586	654	704	82	75	79	128
1986	693	626	534	702	100	101	100	122
1987	712	642	778	687	92	97	95	130
1988	666	677	694	688	121	77	99	129
1989	736	682	674	698	121	69	95	127
1990	902	727	732	713	129	90	110	123
85-90	4367	3940	4066	4202	645	509	577	760
$\hat{P} - P$	427	-	-136	-	-115	-251	-183	-
$\hat{\sigma}(\hat{P})$	251	-	134	-	70	62	55	-
p-Wert	0.09	-	0.3	-	0.1	<0.001	0.001	-

Tabelle 13 vergleicht die Schätzung der Anzahl der Ereignisse "Geburt", "Tod" und "Scheidung" für die Jahre 1984 bis 1990 mit den entsprechenden Angaben aus den Standesamt-Registern (Statistisches Bundesamt 1989,1992). Zunächst zeigt sich, daß die Schätzwerte für das Startjahr 1984 bei allen Merkmalen unter dem entsprechenden Registerwert liegen. Dies hat seine Ursache darin, daß die Feldarbeiten für die erste Welle des SOEP teilweise erst im Sommer 1984 abgeschlossen wurden und die Angaben über Geburten und Todesfälle aus den Kontaktprotokollen gewonnen wurden, die sich auf Ereignisse seit der letzten Befragung beziehen. Wegen dieses ungenauen Zeitbezugs wurden die Ereignisse von 1985 bis 1990 kumuliert und in der Zeile "85-90" mit den entsprechenden Registeraggregaten verglichen.

Bei den Geburten werden für den Zeitraum von 1985 bis 1990 die Registerzahlen um 427 Tausend überschätzt. In 2 (1985 und 1986) der 6 Jahre wird die Anzahl der Geburten unterschätzt.

Bei den Todesfällen wird die Anzahl um 136 Tausend gegenüber den Registerangaben unterschätzt. Erwartungsgemäß liegen die höchsten Unterschätzungen im Anfang der Periode (1985, 1986), da sich hier noch die Folgen der Ausblendung des Anstaltsbereichs bei Beginn des SOEP zeigen. Da vermutet werden kann, daß ein größerer Anteil der Verstorbenen zuvor in Altersheimen gelebt hat, muß die Nichterfassung dieses Bereichs zu einer Unterschätzung der Todesfälle für die Anfangszeit des Panels führen.

Bei den Ehescheidungen werden 3 Schätzungen benutzt. Einmal die Ehe-

scheidung eines Mannes (\hat{M}), zum anderen die Ehescheidung einer Frau (\hat{F}). Der Schätzer $\widehat{M + F}$ basiert auf der Gesamtzahl der geschiedenen Personen. Er ergibt sich als Mittelwert der Schätzungen \hat{M} und \hat{F} . Alle drei Schätzungen sind plausible Verfahren zur Ermittlung der Anzahl der Ehescheidungen. Die beiden Schätzungen sowie der Mittelwert dieser Schätzungen ($\hat{M} + \hat{F}$) liefern unterschiedliche Werte für die Anzahl der Ehescheidungen, wobei in 4 von 7 Jahren der Schätzer über \hat{M} höhere Werte als der Schätzer über \hat{F} liefert ²⁸. Allerdings unterschätzen beide Varianten die Gesamtzahl der Scheidungen im Zeitraum 1985 bis 1990.

Wir wollen nun testen, ob die kumulativen Abweichungen von den Registerwerten allein durch die Zufälligkeit der Stichprobenziehung entstanden sein können. Unter der Nullhypothese stimmt also P mit $E(\hat{P})$ überein. Ist daher $t = (\hat{P} - P) / \hat{\sigma}(\hat{P})$ vom Betrage groß, so wird man die Nullhypothese verwerfen. Für die Verteilung von t wird die Standardnormalverteilung benutzt. Der in Tabelle 13 ausgewiesene p -Wert gibt an, zu welchem Signifikanzniveau die Nullhypothese gerade noch abgelehnt wird.

Für die Schätzung von $\sigma(\hat{P})$ wird bei Rendtel (1993, Kapitel 3) eingeführte Random-Group Schätzer benutzt. Es zeigt sich, daß hinsichtlich der Geburten und Todesfälle die Nullhypothese nach üblichen Signifikanzniveaus nicht verworfen werden kann.

Bei den Ehescheidungen wird dieselbe Hypothese zweimal getestet (über die Männer und über die Frauen). Während der Test über die Männer die Nullhypothese nicht unbedingt ablehnt, muß die Nullhypothese bei den Frauen klar abgelehnt werden. Auch hinsichtlich des multiplen Niveaus ist die Nullhypothese abzulehnen. Benutzt man die gemeinsame Schätzung über Männer und Frauen, so zeigt sich, daß auch diese die Anzahl der Ehescheidungen signifikant unterschätzt. Pro Welle wird die Zahl der Ehescheidungen um ca. 30 000 unterschätzt. Dies entspricht bei einem durchschnittlichen Hochrechnungsfaktor von 5000 lediglich 6 Ehescheidungen in der Stichprobe.

Die Ablehnung der Nullhypothese ("Der Erwartungswert von \hat{P} stimmt mit P überein") führt auf das Entscheidungsproblem, welche Alternative als "wahr" angenommen wird. Dies ist eine schwierige Entscheidung, da (a) mehrere Alternativen zu demselben Resultat bezüglich der Verteilung von \hat{P} führen können und (b) meistens nicht die Daten bereitstehen, um zwischen diesen Alternativen zu unterscheiden. Diese Problematik ist typisch für die Methode des Vergleichs der Schätzwerte mit Populationswerten. Aus diesem Grunde wurde dieses Verfahren auch als exploratives, d.h. Hypothesen generierendes Verfahren bezeichnet.

Bei den Ehescheidungen sind folgende Hypothesen möglich:

- (a) Die Befragten geben die Ehescheidung nicht im Fragebogen an. Ein

²⁸Die Null-Hypothese, daß die Unterschiede zwischen den Schätzungen \hat{M} und \hat{F} zufällig sind, kann nicht abgelehnt werden

derartiger Itemnonresponse ist bei dem gegebenen Fragebogendesign des SOEP nicht ohne weiteres erkennbar²⁹.

- (b) Die benutzten Hochrechnungsfaktoren basieren noch auf der Annahme, daß die Ausfälle von unterschiedlichen Haushalten unabhängig voneinander sind. Diese Annahme ist durch die hier präsentierte Analyse nicht bestätigt worden.
- (c) Das Ereignis "Scheidung" führt zu höheren Ausfallraten als es allein die Haushaltsmobilität impliziert.

Die Hypothese (b) verdient besondere Beachtung, da sie auf ein generelles Problem hinweist, das für die Schätzung aller Ereignisse relevant ist, die mehrere Personen in unterschiedlichen Haushalten betreffen. Dieses Problem läßt sich besonders einfach in einer ereignisorientierten Sichtweise darstellen. Hierbei werden die Inklusionswahrscheinlichkeiten von Ereignissen bestimmt³⁰. Zu jeder Scheidung s gehören eine Frau F_s und ein Mann M_s . Der letzte Haushalt, in dem das Ehepaar F_s und M_s vor ihrer Scheidung zusammen leben, sei H_s . W_{F_s} sei der Indikator für die erfolgreiche Weiterverfolgung der Frau F_s bis zur Scheidung und W_{M_s} sei der entsprechende Indikator für den Mann M_s . Es gilt dann:

$$\begin{aligned} \Pr(s \text{ wird beobachtet}) &= \Pr(F_s \text{ oder } M_s \text{ wird beobachtet}) \\ &= \Pr(F_s \text{ wird beob.}) + \Pr(M_s \text{ wird beob.}) \\ &\quad - \Pr(F_s \text{ und } M_s \text{ wird beob.}) \\ &= \Pr(H_s \text{ wird beob.}) (\Pr(W_{F_s} = 1) + \Pr(W_{M_s} = 1) \\ &\quad - \Pr(W_{F_s} = 1, W_{M_s} = 1)) \end{aligned}$$

Wir wollen nun abschätzen, in welchen Größenordnungen

$$p_{W_s} = \Pr(W_{F_s} = 1) + \Pr(W_{M_s} = 1) - \Pr(W_{F_s} = 1, W_{M_s} = 1)$$

hinsichtlich unterschiedlicher Annahmen über $\Pr(W_{F_s} = 1, W_{M_s} = 1)$ schwankt. Vereinfachend wollen wir dabei annehmen, daß die Weiterverfolgungswahrscheinlichkeit einer Person in einem alten Haushalt bei ca. 0.95 und in einem abgespaltenen Haushalt bei ca. 0.75 liegt. Diese Werte entsprechen etwa den in Tabelle 4 ausgewiesenen Werten. Geht man davon aus, daß einer der Geschiedenen in einem alten Haushalt und der andere in einem neuen Haushalt lebt, so erhält man bei Unabhängigkeit der Weiterverfolgung:

²⁹Auf einen Itemnonresponse kann man bei unplausiblen Übergängen bei der Variablen "Familienstand" schließen; etwa wenn eine Person nach der Angabe "verheiratet" die Angabe "ledig" macht. Jedoch sind derartige Schlüsse mit hohen Unsicherheiten behaftet.

³⁰Dieser Zugang dient nur der vereinfachten Darstellung. Die Erhebungseinheit "Scheidung" ist ein bloßes Konstrukt. Die einzig sinnvollen realen Erhebungseinheiten bleiben im SOEP immer noch Personen bzw. Haushalte.

$$\begin{aligned}
 p_{W_s} &= 0.95 + 0.75 - 0.95 \cdot 0.75 \\
 &= 0.99
 \end{aligned}$$

Bei einem gleichgerichteten Teilnahmeverhalten verringert sich p_{W_s} . Der Grad der Verkleinerung hängt vom Grad der Abhängigkeit und der Richtung der Beeinflussung ab. Der Wertebereich für $Pr(W_{F_s} = 1, W_{M_s} = 1)$ liegt zwischen 0.75 und 0.95. Entsprechend liegt p_{W_s} zwischen 0.95 und 0.75.

Für die Schätzung der Ehescheidungen wurde in Tabelle 13 ein Wert für p_{W_s} benutzt, der $1/0.99$ entspricht. Die Verwendung von $1/0.95$ bzw. $1/0.75$ führt zu einer Erhöhung der Anzahl der Ehescheidungen um den Faktor 1.05 bzw. 1.33. Da der Registerwert in Tabelle 13 das 1.3-fache des Schätzwertes (Version $\widehat{M + F}$) beträgt, kann diese Unterschätzung also allein das Resultat einer unzutreffenden Unabhängigkeitsannahme für das Teilnahmeverhalten von Haushalten sein.

6 Zusammenfassung

Diese Studie setzte sich mit dem Begriff der "Repräsentativität" einer Panelstichprobe auseinander. Es zeigte sich, daß eine naive Interpretation dieses Begriffs als Proportionalität von Stichprobe und Grundgesamtheit, die schon für Querschnittsurveys problematisch ist, für Panelerhebungen gänzlich unbrauchbar ist. Benutzt man "Repräsentativität" als Synonym für die Möglichkeit, auf Basis der Stichprobe valide Rückschlüsse für die Grundgesamtheit zu ziehen, so muß man die Art der Analyse und das Validitätskriterium spezifizieren. Die vorliegende Arbeit befaßte sich mit der erwartungstreuen Schätzung von Populationsmerkmalen. Die hierzu notwendige Qualitätsforderung ist die Kenntnis der Auswahlwahrscheinlichkeiten der Elemente der Panelstichprobe.

In dieser Operationalisierung von "Repräsentativität" bietet sich in einem Panel die Möglichkeit, eine Reihe von "Bias"-Hypothesen der Umfrageforschung direkt zu überprüfen.

Für das SOEP ergab sich dabei keine empirische Evidenz, daß die klassischen Kontrollvariablen das Teilnahmeverhalten nach der ersten Welle wesentlich beeinflussen. Vielmehr ergab sich das Bild, daß die unmittelbaren Umstände der Feldarbeit für das Teilnahmeverhalten wesentlich sind. Dieses Ergebnis steht nicht unbedingt im Widerspruch zu den klassischen Thesen der Umfrageforschung. Deren Ausgangspunkt war stets die unmittelbare Interviewsituation. Problematisch ist nur die Operationalisierung von "geringes Interesse" durch "geringes Bildungsniveau" und "keine Zeit" durch eine "hohe berufliche Position". In einem Panel ergeben sich jedoch präzise operationalisierbare Feldvariablen, z.B.: Interviewerwechsel, Anzahl der gegebenen Interviews und Haushaltsmobilität.

Solange die im Rahmen dieser Analyse geschätzten Ausfallwahrscheinlichkeiten dem tatsächlichen Antwortverhalten entsprechen, sind Populationsschätzungen auf Basis dieser Wahrscheinlichkeiten erwartungstreu für den Populationswert. Unter dieser Prämisse ist also ein Panel selbst noch dann "repräsentativ" in dem hier benutzten Sinn, wenn die Ausfälle während des Panels ungleichmäßig auftreten.

Die Analyse hat allerdings auch gezeigt, daß bestimmte Hypothesen über das Ausfallverhalten auch in einem Panel nicht direkt überprüft werden können. Dies betrifft alle Hypothesen, die ein erhöhtes Ausfallrisiko als Folge von Merkmalsveränderungen postulieren. Ausnahmen bestehen nur hinsichtlich solcher Veränderungen, die schon im Rahmen der Feldarbeit erkennbar sind, wie z.B. Umzüge und Haushaltsaufspaltungen.

Nur in wenigen Fällen kann man durch Vergleich mit den entsprechenden Populationswerten zumindest indirekt Rückschlüsse auf den Effekt demographischer Veränderungen von Haushalten auf deren Teilnahmeverhalten ziehen. Hierbei ergaben sich in zwei Fällen (Geburt und Tod) keine Hinweise auf ein verändertes Antwortverhalten. Im dritten Fall (Scheidung) kann die Unterschätzung des Populationswertes allein aufgrund eines Feldeffekts (Interaktionen im Antwortverhalten) zustande gekommen sein, der bei der Schätzung noch nicht berücksichtigt wurde.

Es spricht also einiges für die Hypothese, daß Haushaltsveränderungen nicht ursächlich die Antwortbereitschaft in einem Panel beeinflussen, solange sie nicht mit einer wesentlichen Veränderung der Feldsituation verbunden sind.

Weiterhin muß man als wesentliches Ergebnis der Ausfallanalyse des SOEP feststellen, daß in Haushaltspanels mit Interaktionen im Antwortverhalten zu rechnen ist. Neben der Abhängigkeit im Antwortverhalten innerhalb von Haushalten muß offensichtlich auch mit Abhängigkeiten zwischen Haushalten gerechnet werden, deren Bewohner in früheren Panelwellen zusammengelebt haben. Dieser Aspekt bedarf noch einer genaueren Untersuchung als dies hier geschehen ist.

Insgesamt kann man feststellen, daß auch nach 8 Befragungswelle sinnvolle Populationsschätzungen mit der SOEP-Stichprobe möglich sind. In diesem Sinne ist das SOEP auch weiterhin "repräsentativ".

Literatur

Basu, Debabrata 1971: An essay on the logical foundations of survey sampling, Part one, in: Godambe, V. und D. Sprott (Hrsg.), Foundations of Statistical Inferenc, Toronto, S. 203-242.

Beckett, Sean, William Gould, Lee Lillard und Finis Welch 1983: Attrition from the PSID, Santa Monica, CA.: Unicorn Research Corp.

- Beckett, Sean, William Gould, Lee Lillard und Finis Welch 1988*: The Panel Study of Income Dynamics After 14 Years: An Evaluation, *Journal of Labor Economics*, 6, S. 472-492.
- Blossfeld, Hans-Peter 1987*: Zur Repräsentativität der Sfb-3 Lebensverlaufsstudie. Ein Vergleich mit Daten der amtlichen Statistik, in: *Allgemeines Statistisches Archiv*, 71, S. 126-144.
- Böltken, Ferdinand 1976*: Auswahlverfahren. Eine Einführung für Sozialwissenschaftler, Stuttgart.
- Cochran, William 1977*: Sampling Techniques, New York 1977³.
- Coz, David R. und David V. Hinkley 1974*: Theoretical Statistics, London.
- Duncan, Greg und Daniel Hill 1989*: Assessing the Quality of Household Panel Data: The Case of the Panel Study of Income Dynamics, *Journal of Business and Economic Statistics*, October 1989, S. 1-11.
- Esser, Hartmut 1988*: Über die Teilnahme an Befragungen, in: *ZUMA-Nachrichten*, 18, S. 38-47.
- Hartmann, Peter H. und Bernhard Schimpl-Neimanns 1992*: Sind Sozialstrukturanalysen mit Umfragedaten möglich? Analysen zur Repräsentativität einer Sozialforschungsumfrage, in: *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 44, S. 315-340.
- Kruskal, William und Frederick Mosteller 1979a*: Representative sampling, I: Non-scientific literature, in: *International Statistical Review*, 47, S. 13-24.
- Kruskal, William und Frederick Mosteller 1979b*: Representative sampling, II: Scientific literature, excluding statistics, in: *International Statistical Review*, 47, S. 111-127.
- Kruskal, William und Frederick Mosteller 1979c*: Representative sampling, III: The current statistical literature, in: *International Statistical Review*, 47, S. 245-265.
- Rendtel, Ulrich 1988*: Panelmortalität. Eine Analyse der Antwortausfälle beim Sozio-ökonomischen Panel in der 2. und 3. Befragungswelle, in: *Vierteljahrshefte zur Wirtschaftsforschung*, Heft 1/2, S. 37-59.
- Rendtel, Ulrich 1990*: Teilnahmeentscheidung in Panelstudien: Zwischen Beeinflussung, Vertrauen und sozialer Selektion. Über die Entwicklung der Antwortbereitschaft im Sozio-ökonomischen Panel, in: *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 42, S. 280-299.
- Rendtel, Ulrich 1991*: Die Schätzung von Populationswerten in Panelerhebungen, in: *Allgemeines Statistisches Archiv*, 75, S. 225-244.
- Rendtel, Ulrich 1993*: Die Auswertung von Haushaltspanels unter Berücksichtigung von Stichprobenausfällen, Manuskript, in Vorbereitung.
- Rendtel, Ulrich und Ulrich Pötter 1992*: Über Sinn und Unsinn von Repräsentativitätsstudien, DIW-Diskussionspapier Nr. 61, Berlin.
- Rendtel, Ulrich und Ulrich Pötter*: "Empirie ohne Daten". Kritische Anmerkungen zu einer Repräsentativitätsstudie über den Allbus, erscheint in: *Kölner Zeitschrift für Soziologie und Sozialpsychologie*.
- Reuband, Karl-Heinz 1984*: Zur Rekrutierung und sozialen Zusammensetzung von Interviewerstäben, in: H. Meulemann und K.-H. Reuband (Hrsg.), *Soziale Realität im Interview*, Frankfurt/New York, S. 61-80.
- Statistisches Bundesamt 1989*: *Wirtschaft und Statistik*, Serie 1, S. 155.
- Statistisches Bundesamt 1992*: *Statistisches Jahrbuch 1992*, S. 73 und S. 84.
- Wagner, Gert, Jürgen Schupp und Ulrich Rendtel*: Das Sozio-ökonomische Panel — Methoden der Datenproduktion und -aufbereitung im Längsschnitt, in: *DIW Diskussionspapier Nr. 31*, Berlin 1991 (erscheint 1993 in: *Hauser, R. et al. (Hrsg.), Mikroanalytische Grundlagen der Gesellschaftspolitik — Erhebungsverfahren, Analyseverfahren und Mikrosimulation*).