

SOEPpapers

on Multidisciplinary Panel Data Research

129

Tobias Gramlich

**Analyse der Panelausfälle
im Sozio-oekonomischen Panel SOEP**

Berlin, Oktober 2008

SOEPpapers on Multidisciplinary Panel Data Research at DIW Berlin

This series presents research findings based either directly on data from the German Socio-Economic Panel Study (SOEP) or using SOEP data as part of an internationally comparable data set (e.g. CNEF, ECHP, LIS, LWS, CHER/PACO). SOEP is a truly multidisciplinary household panel study covering a wide range of social and behavioral sciences: economics, sociology, psychology, survey methodology, econometrics and applied statistics, educational science, political science, public health, behavioral genetics, demography, geography, and sport science.

The decision to publish a submission in SOEPpapers is made by a board of editors chosen by the DIW Berlin to represent the wide range of disciplines covered by SOEP. There is no external referee process and papers are either accepted or rejected without revision. Papers appear in this series as works in progress and may also appear elsewhere. They often represent preliminary studies and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be requested from the author directly.

Any opinions expressed in this series are those of the author(s) and not those of DIW Berlin. Research disseminated by DIW Berlin may include views on public policy issues, but the institute itself takes no institutional policy positions.

The SOEPpapers are available at
<http://www.diw.de/soeppapers>

Editors:

Georg **Meran** (Vice President DIW Berlin)
Gert G. **Wagner** (Social Sciences)
Joachim R. **Frick** (Empirical Economics)
Jürgen **Schupp** (Sociology)
Conchita **D'Ambrosio** (Public Economics)
Christoph **Breuer** (Sport Science, DIW Research Professor)
Anita I. **Drever** (Geography)
Elke **Holst** (Gender Studies)
Frieder R. **Lang** (Psychology, DIW Research Professor)
Jörg-Peter **Schräpler** (Survey Methodology)
C. Katharina **Spieß** (Educational Science)
Martin **Spieß** (Survey Methodology)
Alan S. **Zuckerman** (Political Science, DIW Research Professor)

ISSN: 1864-6689 (online)

German Socio-Economic Panel Study (SOEP)
DIW Berlin
Mohrenstrasse 58
10117 Berlin, Germany

Contact: Uta Rahmann | urahmann@diw.de

Tobias Gramlich

**Analyse der Panelausfälle im
SozioOekonomischen Panel SOEP**

**Institut für Soziologie
Universität Duisburg-Essen**

15. Oktober 2008

Inhaltsverzeichnis

1	Nonresponse und Panelausfälle	12
1.1	Definition Nonresponse und Nonresponsebias	13
1.2	Nonresponse-Bias	16
1.3	Ausfallmechanismen	18
1.3.1	Unsystematische Ausfälle, <i>Missing Completely At Random, MCAR</i> .	19
1.3.2	Systematische Ausfälle, <i>Covariate-Dependent Dropout</i>	20
1.3.3	Systematische Ausfälle, <i>Missing at Random, MAR</i>	21
1.3.4	Systematische Ausfälle, <i>Missing Not at Random, MNAR</i>	23
1.4	Arten von Nonresponse, Ausfallgründe	25
1.4.1	Erreichbarkeit, Kontakt	28
1.4.2	Ausfälle durch Verweigerung	30
1.4.3	Ausfälle durch Teilnahmeunfähigkeit	32
1.5	Nonresponse in Panelerhebungen, Panelausfälle	34
1.5.1	Absorbierende Befragungszustände	37
1.5.2	Nicht-absorbierende Befragungszustände	38
1.5.3	Ausfallmuster bei Panelausfällen	38
1.5.4	Teilnahmesequenzen	41
2	Theoretische Ansätze zur Erklärung von Teilnahmeverhalten	43
2.1	Leverage-Saliency-Theorie	44
2.2	Oversurveying	45
2.3	Rational-Choice-Theorien	46
3	Verfahren zur Behandlung von Nonresponse	53
3.1	Gewichtungsverfahren	55
3.2	Ausfallmechanismus als Omitted Variable	57
3.3	Gewichtung auf Basis der Ausfallwahrscheinlichkeiten: Propensity-Gewichtung	59
3.4	Imputation fehlender Werte	59

4	Datenbasis: Das Sozio-Oekonomische Panel (SOEP)	63
4.1	Grundgesamtheit und Zielpopulation	64
4.2	Die Substichproben des SOEP	65
4.2.1	Stichprobe A (Westdeutschland)	65
4.2.2	Stichprobe B (Ausländer)	66
4.2.3	Stichprobe C (DDR/Ostdeutschland)	67
4.2.4	Stichprobe D (Zuwanderer)	68
4.2.5	Stichprobe E (Stabilisierung, Querschnitt)	70
4.2.6	Stichprobe F (Innovation)	70
4.2.7	Stichprobe G (High Income)	71
4.3	Weiterverfolgungskonzept und Panelpflege des SOEP	74
4.3.1	Weiterverfolgungsregeln	75
4.3.2	Panelpflege und Tracking	78
4.4	Interviewereinsatz beim SOEP	80
5	Datenstruktur und -aufbereitung	86
5.1	Ausfallursachen im SOEP	87
6	Entwicklung der Ausfälle im SOEP	93
7	Ausfallmuster und Ausfallsequenzen	104
7.1	Ausfallmuster im SOEP	104
7.2	Teilnahmesequenzen im SOEP	105
8	Multinomiales Logit-Modell der Bearbeitungsergebnisse	112
8.1	Abfolge der Bearbeitungsergebnisse als Markov-Kette	113
8.2	Schätzung der Übergangswahrscheinlichkeiten - Multinomiales Logitmodell .	116
8.3	Vorhergesagte Übergangswahrscheinlichkeiten des Markov-Modells	123
8.4	Bemerkungen zur Bewertung des Modellfits und Regressionsdiagnostik . . .	127
9	Diskussion der Ergebnisse und Schlussfolgerungen	129

Literatur	130
A Anhang	143
A.1 Abbildungen	143
A.2 Tabellen	145

Tabellenverzeichnis

1.1	Ausfallmuster am Beispiel eines Panels mit 7 Wellen	39
1.2	Teilnahmesequenzen am Beispiel eines Panels mit 7 Wellen	43
4.1	Überblick über die SOEP-Substichproben	73
4.2	Überblick über die SOEP-Substichproben: Auswahlatz und Design	74
5.1	Ausfallursachen im SOEP, Variable \$PERGZ („Zweisteller“) aus dem Datensatz \$PBRUTTO	89
5.3	Aufbereitung der Ausfallursachen im SOEP: Datensätze und Variablen	91
5.4	Änderung der Ausfallursachen in \$PBRUTTO durch Angaben in YPBRUTTO	92
7.1	Teilnahmemuster: Vollzeit - monoton - nichtmonoton	104
7.2	Die 10 häufigsten Sequenzen, Sequenzlänge 20, 15, 10, 5 und 1	108
7.3	Sequenzen des Bearbeitungsergebnisses, Jahr vor dem Ausfall, 1984-2004	109
7.4	Sequenzen des Bearbeitungsergebnisses, 2 Jahre vor dem Ausfall, 1984-2004	111
8.1	Multinomiales Logit-Modell: Bearbeitungsergebnis - Bearbeitungsergebnis Vor- welle	117
A.1	Multinomiales Logit-Modell: Bearbeitungsergebnis - Bearbeitungsergebnis Vor- welle/Teilnahmejahr	145

Abbildungsverzeichnis

1.1	Datenlage bei Nonresponse: Frame-Information Z , Kovariaten X , abhängige Variable Y	19
4.1	Entwicklung der Fallzahlen im SOEP, Haushalte und Personen, 1985-2005 . . .	75
4.2	Entwicklung der Fallzahlen, nach Teilnahmejahren	75
4.3	Anzahl eingesetzte Interviewer, Anzahl Interviews pro Interviewer, 1985-2005	81
4.4	Durchschnittliche Anzahl Interviewerkontakte bis zum Kontakt, 1985-2005 bzw. Welle 1 bis 22	82
4.5	Durchschnittliche Anzahl Interviewerkontakte bis zum Ausfall, 1985-2005 bzw. Welle 1 bis 22	84
4.6	Anteile der Ausfälle an allen Interviews pro Interviewer (verweigert, nicht auffindbar, nicht durchführbar)	85
4.7	Anteile der Ausfälle an allen Interviews pro Interviewer (nicht auffindbar, nicht auswertbar)	86
5.1	Datenstruktur des SOEP, jährliche Querschnittsdatensätze	87
5.2	Datenstruktur des SOEP, Längsschnittsdatensätze	87
6.1	Entwicklung der Ausfälle im SOEP, 1985-2005, $\beta = 0.18$	95
6.2	Entwicklung der Ausfälle im SOEP, nach Teilnahmejahren, $\beta = -0.24$	95
6.3	Entwicklung Ausfälle „Nicht auffindbar“, 1985-2005, $\beta = -0.21$	96
6.4	Ausfälle „Nicht auffindbar“, nach Teilnahmejahren, $\beta = -0.12$	96
6.5	Entwicklung Ausfälle „Nicht erreichbar“, 1985-2005, $\beta = 0.11$	97
6.6	Ausfälle „Nicht erreichbar“, nach Teilnahmejahren, $\beta = -0.25$	97
6.7	Entwicklung Ausfälle „Nicht durchführbar“, 1985-2005, $\beta = -0.57$	98
6.8	Ausfälle „Nicht durchführbar“, nach Teilnahmejahren, $\beta = -0.1$	98
6.9	Entwicklung Ausfälle „verstorben“, 1985-2005, $\beta = -0.08$	99
6.10	Ausfälle „verstorben“, nach Teilnahmejahren, $\beta = 0.57$	99
6.11	Entwicklung Ausfälle „verzogen“, 1985-2005, $\beta = -0.34$	100
6.12	Ausfälle „verzogen“, nach Teilnahmejahren, $\beta = 0.04$	100
6.13	Entwicklung Ausfälle „nicht auswertbar“, 1985-2005, $\beta = 0.06$	101

6.14	Ausfälle „nicht auswertbar“, nach Teilnahmejahren, $\beta = 0.02$	101
6.15	Entwicklung neutraler Ausfälle, 1985-2005	102
6.16	Entwicklung Ausfälle „verweigert“, 1985-2005, $\beta = 1.81$	103
6.17	Ausfälle „verweigert“, nach Teilnahmejahren, $\beta = 0.01$	103
8.1	Vorhergesagte Übergangswahrscheinlichkeiten zwischen den Befragungsergebnissen	123
A.1	Entwicklung und Veränderung der Haushaltsgröße, 1985-2005	143
A.2	Sequenz-Index-Plot der Teilnahmesequenzen mit Sequenzlänge 20	144

Vorwort

Diese Arbeit ist als Abschlussarbeit zum Diplom-Verwaltungswissenschaftler an der Universität Konstanz entstanden. Ich danke meinen Betreuern und Gutachtern Prof. Dr. Thomas Hinz und insbesondere Prof. Dr. Rainer Schnell für seine Ideen und Hinweise zu Thema und Durchführung dieser Arbeit. Außerdem danke ich Stefan Baron, Iris Bräuning, Christina Fritz, Martin Gramlich und Gabriele Gramlich für die Durchsicht des Manuskripts, sowie Susi Gramlich für ihre Aufmunterungen und Ermutigungen.

Die ursprüngliche Fassung der Arbeit ist über die Universitätsbibliothek Konstanz zugänglich unter www.ub.uni-konstanz.de/kops/volltexte/2007/3293/.

Kontakt: Tobias Gramlich, Universität Duisburg, Institut für Soziologie, Lotharstraße 65, 47057 Duisburg.

E-Mail: tobias.gramlich@uni-due.de

Abstract

Nonresponse is a severe problem in sample surveys, especially in panel surveys, where nonresponse results not only in reduced efficiency of estimates compared to the full initial sample but the sample also becomes more and more selective with each wave of the panel since nonresponse is cumulative over all waves. Nonresponse threatens the possibility to make inference on the population the sample was drawn from if the mechanism that leads to nonresponse is nonrandom and (observed or unobserved) characteristics of nonrespondents differ from the characteristics of respondents. If nonrespondents differ systematically from respondents nonresponse can cause biased estimates. The analysis of nonresponse and the mechanisms that lead to nonresponse is therefore essential in order to examine possible differences between nonrespondents and respondents and to account for nonresponse in estimation to avoid nonresponse bias in estimates. However, the threat of possible nonresponse bias is complicated by the fact that nonrespondents can differ not only systematically from respondents, but also different types of nonrespondents can differ systematically, splitting the original sample not only in respondents and nonrespondents but in different types of nonresponse categories. Each of the different types of nonresponse can be a source of a possible nonresponse bias and has distinct effects. For instance, refusals may bias the mean downwards, whereas noncontacts bias the mean away from zero: If nonresponse is not analysed separately, possible effects of the different types can be mixed, making it difficult or impossible to account for possible nonresponse bias accurately.

This thesis analyses nonresponse in the German Socio-Economic Panel Study (SOEP) from 1984 until 2005. In the yearly gross data of individuals, the SOEP documents up to 29 different reasons for nonresponse; some of them lead to final drop out (deceased, moved abroad, explicit refusal), whereas other nonrespondents are contacted again the following year according to the tracking rules. Unfortunately, due to changes in fieldwork coding schemes or different coding behavior of the interviewers, the coding scheme for the 29 reasons for nonresponse is not consistent over time, yielding 10 distinctive and consistent categories of nonresponse over time after summarising the original 29 categories.

The thesis shows the development of the different categories of nonresponse in the SOEP.

By looking at the whole history of contact information for every single observation, drop-out sequences are constructed showing participation patterns and typical sequences one, two, and three years prior to leaving the panel. Typical drop-out sequences are, for example, definite refusals after successful previous interviews or two consecutive waves of refusing. The next most frequent drop-out sequences are leaving the panel due to death or moving abroad after being interviewed successfully. Refusals after missing the previous wave because of being incapable for an interview is another rather typical drop-out sequence.

Closely related to the language of sequences, the thesis estimates in the following transition probabilities between the different states using a first-order stationary Markov-Model. Since drop-out categories are unordered states, this can be done using multinomial logistic regression.

Both, the results from looking at typical sequences, as well as the estimated transition probabilities show, that there is a high probability being repeatedly interviewed successfully, but the overall probability decreases markedly if a person is missing the previous wave for any reason. Especially for refusals there is a high probability for refusing again the following wave and therefore leaving the panel definitely. The same holds for noncontacts (either not found at the old address or not reached during the fieldwork period) in the previous wave: there is a relatively high probability for dropping out again due to refusal if found and reached at the new address, or they get lost since they are not found repeatedly.

Whereas the (inverse of the) predicted probabilities can be used as rough propensity weights, the results point to clear implications for the fieldwork procedures of the interviewers: by all means, temporary drop-outs have to be avoided.

Vorgehen und Ziel der Arbeit

Nonresponse stellt ein ernstzunehmendes Problem für die Möglichkeit dar, von einer Stichprobe auf die Grundgesamtheit zu schließen. Durch Nonresponse verringert sich zunächst die Fallzahl der Stichprobe, sodass sich die Effizienz der Schätzer der Grundgesamtheitsparameter im Vergleich zu einer Stichprobe ohne Nonresponse verringert. Zudem besteht die Gefahr der Verzerrung der Schätzer, wenn sich Teilnehmer von Nichtteilnehmern systematisch unterscheiden (Kapitel 1.2) und der Mechanismus, der zu Nonresponse führt nicht zufällig ist (Kapitel 1.3). Diese Ausfallmechanismen können sich für verschiedene Arten von Nonresponse unterscheiden (Kapitel 1.4). Dabei besteht grundsätzlich kein Unterschied zwischen Nonresponse in einmaligen Querschnittserhebungen und Ausfällen bei Wiederholungsbefragungen; die Besonderheiten für Panelausfälle werden in Kapitel 1.5 beschrieben.

In Kapitel 2 werden theoretische Erklärungsansätze für Nonresponse diskutiert und anschließend Verfahren zur Korrektur von Nonresponse vorgestellt (Kapitel 3), um trotz Nonresponse unverzerrte Schätzer der Grundgesamtheitsparameter zu erhalten. Alle diese Korrekturverfahren treffen (zumindest implizit) eine Annahme über den Ausfallmechanismus. Da sich Ausfallmechanismen für unterschiedliche Ausfallarten unterscheiden können ist für eine Korrektur von Nonresponse eine Unterscheidung der Ausfallarten von zentraler Bedeutung.

Der empirische Teil der Arbeit analysiert Ausfälle aus dem SozioOekonomischen Panel (SOEP), dessen Grundgesamtheit und einzelne Substichproben in Kapitel 4 vorgestellt und beschrieben werden. Das SOEP ist eine jährliche Wiederholungsbefragung privater Haushalte und aller erwachsenen Personen in diesen Haushalten. In den Kontaktprotokollen der Interviewer werden bis zu 29 unterschiedliche Kontaktergebnisse festgehalten. Da diese über die Jahre nicht konsistent erfasst wurden, ist eine Aufbereitung der Kontaktergebnisse für jede Welle notwendig, die durch die komplexe Struktur des SOEP erschwert wird (Kapitel 5).

Der deskriptive Teil der Analyse der Panelausfälle im SOEP (Kapitel 6 ab Seite 93) zeigt die Entwicklung der verschiedenen Ausfallursachen seit Start des SOEP im Jahr 1984. Da das SOEP eine Wiederholungsbefragung derselben Personen ist, können die Abfolgen der jährlichen Kontaktergebnisse untersucht werden. Die Teilnahmemuster von Teilnahme und Nichtteilnahme der Befragungspersonen des SOEP werden in Kapitel 7.1 beschrieben. Be-

trachtet man für Nonrespondenten zusätzlich die Abfolge der detaillierten Ausfallursachen, ergeben sich Teilnahmesequenzen. Bei der deskriptiven Analyse dieser Teilnahmesequenzen in Kapitel 7.2 zeigen sich typische häufige Ausfallsequenzen aus dem SOEP für die letzten zwei bzw. letzten drei Jahre, bevor eine Beobachtung aus dem SOEP ausscheidet. Diese Ausfallsequenzen zeigen klare Implikationen für die Feldarbeit und Erhebungspraxis des SOEP auf.

Eng verbunden mit der Beschreibung der Sequenzen schließt sich im folgenden Kapitel eine Modellierung der Ausfälle aus dem SOEP als diskrete und stationäre Markov-Kette an (Kapitel 8 ab Seite 112). Ziel der Analyse ist die Schätzung der Übergangswahrscheinlichkeiten zwischen den Kontaktergebnissen zwischen den Wellen. Da die Ausfallursachen als ungeordnete Kategorien vorliegen, können diese Übergangswahrscheinlichkeiten mit einem multinomialen Logitmodell geschätzt werden. Die Ergebnisse des Markov-Modells bestätigen die Ergebnisse der deskriptiven Analyse.

Ziel der Arbeit liegt aber weniger in den Übergangswahrscheinlichkeiten im Sinne einer Propensity-Gewichtung, sondern es soll vielmehr gezeigt werden, dass die unterschiedlichen Ausfallarten verschiedene Wahrscheinlichkeiten für eine erneute Teilnahme bzw. einen erneuten Ausfall besitzen. Dadurch soll gezeigt werden, dass eine Korrektur von Nonresponse aufgrund einer einfachen Unterscheidung Respondenten - Nonrespondenten zu kurz greift (vor allem, wenn bestimmte Gruppen eine höhere Ausfallwahrscheinlichkeit durch eine bestimmte Ausfallursache haben) und einen möglichen Nonresponsebias nicht oder nur unzulänglich korrigieren kann.

1 Nonresponse und Panelfälle

Panelfälle sind eng mit Nonresponse verknüpft: „in the case of longitudinal surveys, sample attrition is the cumulative effect of non-response over repeated waves“ (Lynn, 2006, S. 393). Nonresponse bezeichnet das Problem, dass für eine zur Befragung ausgewählte Befragungseinheit keine Daten vorliegen; entweder weil einzelne Fragen nicht beantwortet werden (Item-Nonresponse) oder die gesamte Befragungseinheit ausfällt (Unit-Nonresponse) (vgl. bspw. Dillman et al., 2002, S. 3; oder Schnell, 1997, S. 17). Das Problem des Nonresponse ist dabei ungefähr genauso alt, wie Techniken zur Stichprobenziehung (Cochran, 1983, S. 12) und die Umfrageforschung selbst (de Leeuw und de Heer, 2002, S. 41) und ein alltägliches Phänomen: „Nonresponse is a normal but undesirable feature of a survey“ (Särndal und Lundström, 2005, S. 1)¹. Zu Nonresponse existiert eine ausführliche Literatur, aufgeteilt in Literatur darüber, wie Nonresponse durch verschiedene Maßnahmen vermieden oder minimiert werden kann und Literatur darüber, wie trotz Nonresponse Aussagen über die Zielpopulation und Grundgesamtheit gemacht werden können (Särndal und Lundström, 2005, S. 1; Groves et al., 1992, S. 476; vgl. auch Singer, 2006, S. 638f) und eben nicht nur über die Umfragegesamtheit (Population der Respondenten, „Surveypopulation“). Während die statistische Literatur zur Korrektur von Nonresponse theoriegeleitet ist, stellt sie jedoch keine Theorie der Surveyteilnahme dar, demgegenüber ist die Literatur zur Vermeidung von Nonresponse selbst wenig theoriegeleitet (Groves et al., 1992, S. 476).

Obwohl das Nonresponse-Problem nicht neu ist, finden sich seit einigen Jahren allerdings vermehrt Berichte in der Literatur, dass Nonresponseraten über die Zeit hinweg für viele Surveys und in vielen Ländern zugenommen haben (bspw. DeMaio, 1980; Steeh, 1981; de Heer, 1999; Atrostic et al., 2001; de Leeuw und de Heer, 2002; Curtin et al., 2005)². Allerdings

¹Zu diesen „unerwünschten Eigenheiten“ gehören weiter bspw. Under- und Overcoverage der Auswahlgrundlage und Stichprobenfehler durch die Stichprobenvarianz (Särndal und Lundström, 2005, S. 5); zu Fehlerarten von Surveys siehe ausführlicher bspw. Groves et al. (2004, v.a. S. 67-130).

²de Heer (1999) beschreibt auch das Problem des Vergleichs von Responseraten über die Zeit und verschiedene Länder: „Changes in respondent definition, content and design of the survey, fieldwork procedures and strategies as well as organizational changes of the survey fieldwork agency may distort comparisons“ (S. 130) Und weiter: „Differences in response levels may be caused by specific circumstances in some coun-

herrscht darüber auch Uneinigkeit: „Die Klage über das Ansteigen der Nonresponse-Raten scheint somit nur wenig älter zu sein als die Klage über mangelndes Datenmaterial um die behauptete Tendenz zu belegen“ (Schnell, 1997, S. 30). Dies trifft auch auf die Entwicklung des Nonresponse in verschiedenen Ländern zu (de Leeuw und de Heer, 2002, S. 41): „In sum: for several countries a downward trend in response can be discovered for several surveys. However, the trends are not always clear-cut and differences between countries exist for different surveys. Differences [...] may be caused by differences in survey design or by specific circumstances in some countries“ (de Heer, 1999, S. 135). Zu einem ähnlichen Ergebnis kommen Behr et al. (2005, S. 492) für das Europäische Haushaltspanel ECHP³: „there is no clear tendency across all countries in the wave specific response rates in terms of rising or falling“.

Insgesamt stehen wenig theoretische Ansätze zur Erklärung von Nonresponse zur Verfügung, am meisten sind Ausfälle durch Verweigerung (siehe Abschnitt 1.4.2) durch theoretische Arbeiten untermauert (de Leeuw und de Heer, 2002, S. 42). Diese Ansätze werden in Abschnitt 2 näher beschrieben.

1.1 Definition Nonresponse und Nonresponsebias

Nonresponse bezeichnet das Problem, das ein zur Befragung ausgewähltes Element der Grundgesamtheit aus der Stichprobe fällt und keine Informationen ausser den Informationen des Sampling-Frames für diese Elemente vorliegen. Ausgehend von einer einfachen Zufalls-

tries, for example whether or not a survey is mandatory or what specific survey design is used“ (S. 132). Zu den Problemen gehören auch gesellschaftliche Veränderungen, bspw. ein Trend zu verkleinerten Haushalten oder der steigende Anteil berufstätiger Frauen.

³Allerdings ist hier der Vergleich über Länder mit großen Problemen behaftet, da trotz genereller Vorgaben von Eurostat, Erhebungen dezentral in den Mitgliedsländern durchgeführt werden (vgl. Peracchi, 2002, S. 65); bspw. variiert die Anzahl der Interviewerbesuche zwischen den Ländern zwischen 1 und 3 und auch die Interviewerstabilität ist zwischen den Ländern sehr unterschiedlich (unter 30% der ECHP-Haushalte in Portugal werden in der nächsten Welle vom gleichen Interviewer befragt, gegenüber 80% der Haushalte in Irland); auch die Länge der Feldzeit ist nicht für alle teilnehmenden Länder am ECHP gleich (siehe Nicoletti und Peracchi, 2005, S. 768f).

stichprobe ist die Wahrscheinlichkeit π_i einer Beobachtung i in die Stichprobe zu gelangen entsprechend dem Quotient der Elemente der Grundgesamtheit N und der Anzahl der Stichprobenelemente n :

$$\pi_i = \pi = n/N \quad \text{mit} \quad (1.1)$$

$$\pi > 0 \quad (1.2)$$

Im Falle einer einfachen Zufallsstichprobe aus der Grundgesamtheit ist für alle Beobachtungen i die Auswahlwahrscheinlichkeit in die Stichprobe zu gelangen gleich groß; im Falle einer geschichteten Zufallsstichprobe mit j Schichten ergibt sich die Auswahlwahrscheinlichkeit für Beobachtung i_j entsprechend als Quotient der Anzahl der Grundgesamtheitselemente in Schicht j und der Anzahl der Stichprobenelemente in Schicht j :

$$\pi_{i_j} = n_j/N_j \quad (1.3)$$

Der Indikator I gibt nun an, ob eine ausgewählte Beobachtung in die Stichprobe gelangt

$$I_i = \begin{cases} 1, & \text{Beobachtung } i \text{ ist in der Stichprobe,} \\ 0, & \text{Beobachtung } i \text{ ist nicht in der Stichprobe} \end{cases} \quad (1.4)$$

und der Response-Indikator R , ob eine Beobachtung bei Auswahl in die Stichprobe in der Stichprobe beobachtet wurde, oder durch Nonresponse ausfällt

$$R_i(I = 1) = \begin{cases} 1, & \text{Beobachtung } i \text{ wird in der Stichprobe beobachtet} \\ 0, & \text{Beobachtung } i \text{ fällt durch Nonresponse aus.} \end{cases} \quad (1.5)$$

Er teilt Merkmale der Beobachtungen y_i in Merkmale der Beobachtungen, die in der Stichprobe beobachtet werden y_i^{obs} und Merkmale der Beobachtungen, die trotz Auswahl in die Stichprobe durch Nonresponse fehlen y_i^{miss} :

$$\begin{aligned} \sum_{i=1}^n y_{i,I=1} &= y_{i,R=1|I}^{obs} + y_{i,R=0|I}^{miss} \\ &= y_{ir} + y_{im} \end{aligned} \quad (1.6)$$

und der Umfang n der Stichprobe besteht entsprechend aus r Beobachtungen mit $R = 1|I$ und m Beobachtungen mit $R = 0|I$, gegeben $I = 1$:

$$n = r_{(R=1|I)} + m_{(R=0|I)} \quad (1.7)$$

Die Nonresponserate ist entsprechend definiert als Quotient aller ausgewählten Beobachtungen und der in der Stichprobe aufgrund von Nonresponse nicht beobachteten Elemente:

$$NR = \frac{m}{n} \quad (1.8)$$

Der Mittelwert der vollständigen Stichprobe vom Umfang n besteht dementsprechend zunächst aus mindestens zwei Teilen, dem Mittelwert der Beobachtungen $\bar{y}_{i,R=1|I}$ und dem Mittelwert der Beobachtungen $\bar{y}_{i,R=0|I}$:

$$\begin{aligned} \bar{y}_n &= \frac{r}{n} \left(\sum_{i=1}^r y_{ir}/r \right) + \frac{m}{n} \left(\sum_{i=1}^m y_{im}/m \right) \\ &= \frac{r}{n} \bar{y}_r + \frac{m}{n} \bar{y}_m \end{aligned} \quad (1.9)$$

Erster Teil des Nonresponse-Problems besteht nun darin, dass alle Parameter der Stichprobe nur auf Basis der Beobachtungen aus r mit $R = 1$ berechnet werden können, der zweite Teil in Gleichung 1.6 fällt aus allen Schätzungen der Grundgesamtheitsparameter auf Basis der Stichprobenparameter heraus, da in der Stichprobe nur diejenigen Elemente der Stichprobe mit $R = 1|I$ beobachtet, Elemente der Stichprobe mit $R = 0|I$ hingegen nicht beobachtet werden; dadurch werden bspw. Totalwerte T eines Merkmals Y der Grundgesamtheit auf Basis der Stichprobenmerkmale y einer Stichprobe mit Nonresponse durch den Horvitz-Thompson-Schätzer \hat{T} unterschätzt:

$$\hat{T} = \sum_{i=1}^N y_i I_i R_i 1/\pi_i \quad \text{mit: } \sum_{i=1}^n I_i > \sum_{i=1}^n R_i \quad (1.10)$$

$$\hat{T} < T \quad (1.11)$$

Einfache Gewichtungsverfahren können diese Ausfälle leicht korrigieren (siehe dazu ausführlicher Abschnitt 3.1 ab Seite 55): bezeichnet r die Anzahl der beobachteten Elemente der Stichprobe $R = 1|I$ und n die Stichprobengröße ohne Nonresponse, können im einfachsten Fall einer einfachen Zufallsstichprobe Ausfälle beispielsweise durch eine Ausfallgewichtung als Inverse der Responserate $\phi = r/n$ korrigiert werden:

$$w_i = w = 1/(\pi_i * \phi_i) = 1/(\pi * \phi) \quad (1.12)$$

Entsprechend lautet dann bspw. der geschätzte Totalwert der Grundgesamtheit auf Basis der beobachteten Elemente der Stichprobe anhand eines korrigierten Horvitz-Thompson-Schätzers im Vergleich zu Gleichung 1.10:

$$\begin{aligned}\hat{T} &= \sum_{i=1}^r y_i \pi_i^{-1} \phi_i^{-1} \\ &= \sum_{i=1}^r y_i w_i\end{aligned}\tag{1.13}$$

Der durch den beobachteten Stichprobenmittelwert \bar{y}_r geschätzte Mittelwert der Grundgesamtheit \bar{Y} lautet nach einer Ausfallgewichtung dann:

$$\begin{aligned}\bar{Y} &= \left(\sum_{i=1}^N y_i I_i R_i \right) / \left(\sum_{i=1}^N I_i R_i \right) \\ &= \left(\sum_{i=1}^r y_{ir} \pi_i^{-1} \phi_i^{-1} \right) / \left(\sum_{i=1}^r \pi_i^{-1} \phi_i^{-1} \right) \\ &= \left(\sum_{i=1}^r y_{ir} w_i \right) / \left(\sum_{i=1}^r w_i \right)\end{aligned}\tag{1.14}$$

1.2 Nonresponse-Bias

Schwerwiegender und schwieriger zu lösen ist das Problem einer möglichen Verzerrung der Schätzer durch Nonresponse: „moreover, possible biases exist because the respondents are often systematically different from the nonrespondents“ (Rubin, 1987, S. 1; vgl. auch Kautter et al., 2006, S. 72) und wenn Ausfälle (direkt oder indirekt) mit dem Thema der Befragung zusammenhängen. Ein Nonresponse-Bias bezeichnet dann den Unterschied der Mittelwerte des vollständigen Samples \bar{y}_n und dem durch Nonresponse verringerten beobachteten Samples \bar{y}_r (vgl. Groves, 2006, S. 648f). Ausgehend von Gleichung 1.9 ist der Mittelwert der beobachteten Stichprobe \bar{y}_r :

$$\bar{y}_r = \bar{y}_n + \frac{m}{n}(\bar{y}_r - \bar{y}_m)\tag{1.15}$$

Ein möglicher Bias durch Nonresponse ist dann die Abweichung des beobachteten Mittelwerts in der beobachteten Stichprobe \bar{y}_r vom Mittelwert des vollständigen Samples ohne Nonresponse \bar{y}_n :

$$\text{Bias}(\bar{y}_r) = \frac{m}{n}(\bar{y}_r - \bar{y}_m)\tag{1.16}$$

Eine probabilistische Definition eines möglichen Nonresponse-Bias bezieht zusätzlich die Kovarianzstruktur der interessierenden Variable Y_i und dem Responseindikator R_i mit ein (Groves et al., 2004, S. 182):

$$\text{Bias}(\bar{y}_r) = \text{Cov}(R_i, Y_i) + E \left[\left(\frac{m}{n} \right) (\bar{y}_r - \bar{Y}) \right] \quad (1.17)$$

Die Größe des möglichen Nonresponse-Bias hängt also nicht nur vom Ausmaß des Nonresponse, der Nonresponserate $\frac{m}{n}$ ab, sondern auch davon, wie stark sich Nonrespondenten von den Respondenten unterscheiden, d.h. auch bei hohen Responseraten kann Nonresponse große Verzerrungen verursachen⁴, wenn sich der Mittelwert der durch Nonresponse nicht beobachteten Elemente \bar{y}_m sehr stark vom beobachteten Mittelwert \bar{y}_r unterscheidet, aber auch, wenn nur geringe Unterschiede in den Mittelwerten bestehen, aber die Kovarianz zwischen dem Ausfallindikator und der interessierenden Variable hoch ist. Andererseits ist aus Gleichung 1.16 auch deutlich, dass auch bei hoher Nonresponserate kein Nonresponse-Bias vorliegen muss, wenn sich Respondenten nicht von Nonrespondenten unterscheiden und die Differenz der beiden Mittelwerte $\bar{y}_m - \bar{y}_r = 0$ oder nahe bei 0 bzw. ausgehend von Gleichung 1.17 die Kovarianz zwischen R_i und Y_i sehr klein ist. Die probabilistische Definition eines Nonresponse-Bias ist für die Beschreibung der Ausfallmechanismen (siehe Abschnitt 1.3) besonders hilfreich.

Das Problem bei der Bestimmung eines möglichen Nonresponsebias ist aber, dass \bar{y}_m meist nicht berechnet werden kann: „the tricky part is that the values for the nonrespondents are usually not known“ (Groves et al., 2004, S. 182), lediglich die Nonresponserate kann berechnet werden⁵.

⁴„The lack of any evidence that nonresponse causes a problem does not justify ignoring it, because the appropriate interpretation of no evidence is 'do not know'. [...] The relatively high response rate does not justify ignoring the problem of nonresponse altogether, we should [...] play the devil's advocate and contemplate what impact the 20% of the subjects might have had on the planned or intended inferences, had they all responded“ (Longford, 2005, S. 21).

⁵Groves et al. (2004, S. 183) zeigen, dass in einigen Fällen jedoch auch eine Berechnung der Nonresponserate nicht ohne Schwierigkeiten ist, bspw. wenn in einem mehrstufigen Auswahlverfahren Sampling Points vollständig ausfallen und die Anzahl der Elemente in den ausgefallenen Sampling Points nicht bekannt ist;

1.3 Ausfallmechanismen

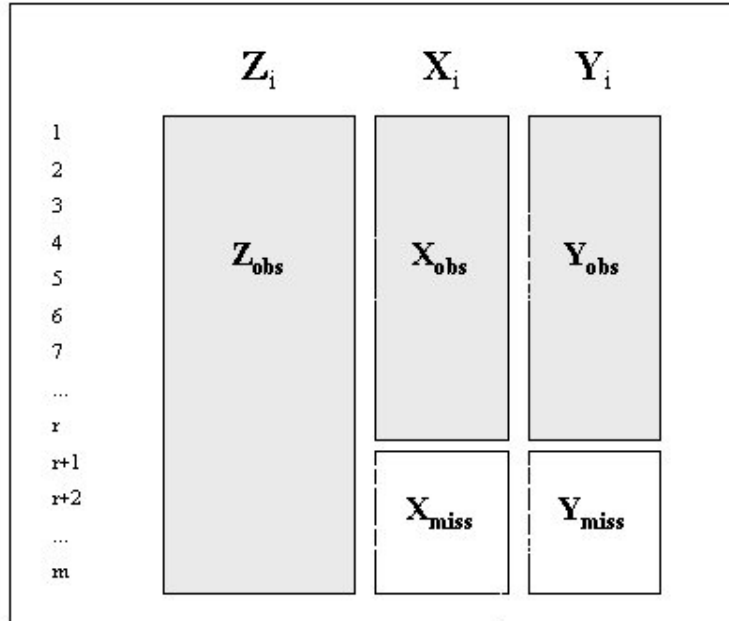
Ein Kriterium für das Vorliegen eines Nonresponse-Bias ist der Mechanismus, der zum Ausfall der n_m Beobachtungen der vollständigen Stichprobe n führt. Danach bestimmt sich eine Unterscheidung danach, ob Nonresponse „ignorierbar“ oder „nicht ignorierbar“ ist. Generell gilt Nonresponse als ignorierbar, wenn der Mechanismus, der zu Nonresponse führt unabhängig von den fehlenden, unbeobachteten Variablen ist; entsprechend ist Nonresponse nicht ignorierbar, wenn der Mechanismus nicht unabhängig von fehlenden Variablen ist (vgl. Fay, 1986, S. 354). Je nach Mechanismus sind Ausfälle unsystematisch und Unterschiede zwischen \bar{y}_r und \bar{y}_m sind ignorierbar oder Unterschiede sind systematisch und müssen bei einer Analyse berücksichtigt und durch geeignete Verfahren korrigiert werden. Alle Maßnahmen zur Korrektur⁶ von Nonresponse (siehe Abschnitt 3 ab Seite 53) treffen zumindest implizit eine Annahme über diese Ausfallmechanismen. Hält diese Annahme über einen Ausfallmechanismus nicht, können Ausfälle nicht korrigiert werden und eine mögliche Verzerrung durch einen Nonresponse-Bias bleibt bestehen (vgl. Ahern und Le Brocque, 2005, S. 55; Goodman und Blum, 1996, S. 629). Ausfallmechanismen unterscheiden sich danach, ob Ausfälle von der interessierenden abhängigen Variable selbst abhängen, von beobachteten Kovariaten⁷ oder unsystematisch und unabhängig von beiden als Zufallsauswahl stattfinden: „the important question is whether missingness depends on missing values, after conditioning on recorded values for a case. If this is the case, the mechanism is nonignorable“ (Little, 1995, S. 1119,

ausserdem ist unklar, ob bei einem Auswahlverfahren mit unterschiedlichen Auswahlätzen für bestimmte Elemente diese unterschiedlichen Auswahlwahrscheinlichkeiten bei der Berechnung der Nonresponseraten berücksichtigt werden müssen.

⁶Nicht nur die Techniken zur Korrektur treffen eine Annahme über den Ausfallmechanismus, auch das Ignorieren von Nonresponse, bspw. durch Ausschluß der unvollständigen Beobachtungen bei Item-Nonresponse legt implizit einen (vollständig zufälligen) Ausfallmechanismus zugrunde.

⁷Das bedeutet auch, dass eine Verzerrung durch Nonresponse und Ausfälle für jede Forschungsfrage und abhängige Variable untersucht werden muss: „there is evidence to suggest, that attrition bias is model specific and must be explored within each sample, population or study. Lack of evidence of bias in one model does not imply that there will no bias in any other model“ (Ahern und Le Brocque, 2005, S. 59; vgl. auch Zabel, 1998, S. 495 und Groves, 2006, S. 649).

Abbildung 1.1: Datenlage bei Nonresponse: Frame-Information Z , Kovariaten X , abhängige Variable Y



vgl. auch Goodman und Blum, 1996, S. 631). Im Folgenden bezeichnet Y die abhängige, interessierende Variable, X die erhobenen Kovariaten, Z bezeichnet Informationen des Sampling Frames oder auch Informationen aus frühere Wellen; Y und X können für Nonrespondenten fehlen, wogegen Z vollständig für Respondenten und Nonrespondenten vorliegt (vgl. Longford, 2005, S. 32 und siehe Abbildung 1.1).

1.3.1 Unsystematische Ausfälle, *Missing Completely At Random, MCAR*

Weitestgehende Annahme bezeichnet unsystematische Ausfälle, die eine Zufallsstichprobe aus dem vollständigen Sample ohne Nonresponse darstellen: eine Teilnahme ist unabhängig von der Auswahl in die Stichprobe zu gelangen, der interessierenden abhängigen Variable Y oder beobachteten oder unbeobachteten Kovariaten X :

$$R \perp\!\!\!\perp (Y, X, I) \quad (1.18)$$

$$Pr(R = 0|y, x, z) = Pr(R = 0) \quad (1.19)$$

Die resultierende Stichprobe vom Umfang n_r ist sowohl eine Zufallsstichprobe aus der Grundgesamtheit, als auch aus der angestrebten Stichprobe vom Umfang n . Der Ausfallmechanismus ist lediglich eine weitere Stufe im Auswahlprozess. Little und Rubin (1987, S. 14) bezeichnen diesen Ausfallmechanismus als *Missing Completely at Random (MCAR)*; sind Ausfälle MCAR kommt es zu keinen systematischen Unterschieden zwischen Respondenten und Nonrespondenten und es liegt kein Nonresponse-Bias vor. Einfache Ausfallgewichte (siehe Gleichung 1.12) treffen bspw. die Annahme vollständig zufälliger Ausfälle, ebenso Imputationsverfahren, die für fehlende Werte lediglich unbedingte Mittelwerte einsetzen (siehe Abschnitt 3.4). Sind Ausfälle vollständig zufällig, *MCAR*, sind Schätzer auf Basis der vollständigen Beobachtungen nicht verzerrt, lediglich die Effizienz der Schätzer verringert sich im Vergleich zum Schätzer auf Basis der vollständigen Stichprobe ohne Nonresponse. Ein Beispiel für einen vollständig zufälligen Ausfallmechanismus stellt die Anwendung des sog. *Double-Samplings* von Nonrespondenten dar (vgl. Madow und Singh, 1983): Aus der Grundgesamtheit wird eine Stichprobe vom Umfang n gezogen; davon können n_1 Elemente befragt werden, n_2 Elemente fallen durch Nonresponse aus; wird aus den n_2 Nonrespondenten eine weitere Stichprobe vom Umfang n_3 gezogen, um sie mit erhöhtem Aufwand doch noch zu befragen und für alle n_3 Elemente liegen danach vollständige Informationen vor, kann für die restlichen $n_2 - n_3$ Elemente der ursprünglichen Stichprobe n ein *MCAR*-Ausfallmechanismus angenommen werden. Die Annahme, Ausfälle seien MCAR ist in der Regel jedoch nicht zutreffend (vgl. Longford, 2005, S. 30): „The assumption that the process acts like simple random sampling (observations missing completely at random) is in most contexts rather optimistic“ (Longford et al., 2006, S. 509).

1.3.2 Systematische Ausfälle, *Covariate-Dependent Dropout*

Little (1995) beschreibt Ausfallmechanismen bei wiederholten Messungen und bezieht für eine Beschreibung der verschiedenen Ausfallmechanismen zusätzlich einen unbeobachteten Zufallsvektor β_i mit ein, der die Veränderung der individuellen Mittelwerte \bar{y}_i über die Zeit für jede Beobachtung i widerspiegelt. Ein Ausfallmechanismus, der unabhängig von der abhängigen Variable und von individuellen Veränderungen der abhängigen Variable über Zeit gege-

ben bestimmte Kovariaten ist, stellt einen Spezialfall der *MCAR*-Annahme im Längsschnitt dar:

$$R \perp\!\!\!\perp (Y, \beta_i) | X \quad (1.20)$$

Diesen Spezialfall bezeichnet Little (1995) als „covariate-dependent dropout“ (S. 1114): „A strong assumption is that the drop-out mechanism does not depend on the outcome values y_i or the random coefficients β_i but is allowed to depend on the values of the fixed covariates X_i “. Obwohl ursprünglich als ein Spezialfall der *MCAR*-Annahme auch als *MCAR* bezeichnet, „I suggest the term ‘covariate-dependent dropout’ because I think the term ‘missing completely at random’ should be reserved for the case when missingness does not depend on y_i , β_i , or X_i “ (S. 1114). Als Beispiel führt er fehlende Beobachtungen aufgrund des Designs der Erhebung an: bei einer wiederholten Messung der Lungenfunktion bei Kindern zwischen 3 und 12 Jahren fehlen für Kinder, die zu Beginn der Studie älter als 3 Jahre sind bzw. jünger als 12 Jahre bei Ende der Studie Messwerte: „This mechanism depends on cohort, but plausibly does not depend on outcomes if cohort is a covariate in the model“ (S. 1114). Der Ausfallmechanismus ist ignorierbar und führt nicht zu verzerrten Schätzern; allerdings verringert sich die Effizienz der Schätzer.

1.3.3 Systematische Ausfälle, *Missing at Random, MAR*

Eine weniger strenge Annahme über den Ausfallmechanismus ist die Annahme, dass Ausfälle unabhängig von der interessierenden Variable Y ⁸ sind, aber von beobachteten Kovariaten X abhängen können: „the observed values of Y are not necessarily a random subsample of the sampled values, but they are a random subsample of the sampled values within subclasses defined by values of X “ (Little und Rubin, 1987, S. 14).

$$R \perp\!\!\!\perp Y | X \quad (1.21)$$

$$Pr(R = 0 | y, x, z) = Pr(R = 0 | x, z) \quad (1.22)$$

⁸Bei Little (1995, S. 1114) heisst es treffender: „missingness depends only on observed data; that is, the distribution of R_i depends on y_i only through its observed components $y_{obs,i}$ “.

Little und Rubin (1987, S. 14) bezeichnen diesen Mechanismus als *Missing at Random*, *MAR*⁹. Korrekturverfahren wie Gewichtungsverfahren nach Gewichtungszellen, die Imputation bedingter Mittelwerte oder mit Regressionsmodellen, sowie eine Ausfallgewichtung anhand von Propensity-Gewichten (zu Korrekturverfahren siehe Abschnitt 3 ab Seite 53) basieren auf der *MAR*-Annahme (vgl. Kalton und Flores-Cervantes, 2003, S. 85 und 89). *MAR*-Mechanismen umfassen die Auswahl anhand beobachtbarer Variablen; die Teilnahmewahrscheinlichkeit hängt nur von beobachteten Kovariaten ab; dasselbe gilt zusätzlich zu Gleichung 1.22 für

$$Pr(R = 0|x, z) = Pr(R = 0|x), \quad (1.23)$$

das heisst, die Teilnahmewahrscheinlichkeit ist zusätzlich zu y auch unabhängig von z , (Alderman et al., 2001, S. 85).

Särndal und Lundström (2005, S. 104) kritisieren diese vereinfachte Annahme: zum Einen bezeichnet die *MAR*-Annahme einen starken Zusammenhang zwischen X und der Teilnahmewahrscheinlichkeit θ und gleichzeitig keinen oder nur schwachen Zusammenhang zwischen Y und θ ; damit dürften auch Y und X nur einen schwachen Zusammenhang zeigen: „But that would go contrary to our preference for a strong relation [...] which is essential [...] in order to control nonresponse“. Damit stehen sie auch der Bezeichnung „ignorierbarer“ Ausfälle kritisch gegenüber; ignorierbar sind Ausfälle dann, wenn ein Modell diese Ausfälle beschreiben und vorhersagen kann (wie bei Gewichtungsverfahren oder der Propensity-Gewichtung gefordert; siehe Abschnitt 3). Allerdings: „complete explanation of θ in terms of a vector x is virtually never possible. All situations are nonignorable. We can never hope to have an auxillary vector liable to correct, with full effectiveness, for bias due to nonresponse. [...] to hope that this x will achieve a complete explanation of θ is utopian“ (S. 105).

Eine Variante der *MAR*-Annahme im Längsschnitt ist die Annahme, dass Ausfälle zusätzlich zu den Kovariaten X auch von den Random-Effects-Koeffizienten β_i (siehe auch Abschnitt 1.3.2) abhängen können (Little, 1995, S. 1118):

$$R \perp\!\!\!\perp Y|(X, \beta_i). \quad (1.24)$$

⁹Little (1995) bezeichnet diesen Mechanismus als *random dropout* (S. 1114).

Little (1995, S. 1118) bezeichnet diesen Mechanismus „Random-Effect-Dependent Dropout“.

1.3.4 Systematische Ausfälle, *Missing Not at Random, MNAR*

Schwerwiegender sind systematische Ausfälle, die nicht unabhängig von einer Ausprägung der abhängigen Variable selbst sind; d.h. wenn die Teilnahme als eine Funktion der abhängigen Variable darstellbar ist:

$$R \not\perp (Y_{miss})|X \quad (1.25)$$

$$Pr(R = 0|y, x, z) \neq Pr(R = 0|x, z) \quad (1.26)$$

Ausfälle sind insbesondere nicht unabhängig vom unbeobachteten Teil Y_{miss} von Y . Little (1995, S. 1115) bezeichnet diesen Mechanismus als *Nonignorable Outcome-Based Dropout*, bei Diggle et al. (1994, S. 209) wird dieser Mechanismus als *informative missing value mechanism* bezeichnet. Eine Auswahl findet hier also anhand der nichtbeobachteten Variablen statt (*missing on unobservables*).

Hängt der Ausfallmechanismus nicht direkt von $y_{miss,i}$ sondern indirekt über den unbekanntem Koeffizient β_i bezeichnet Little (1995, S. 1116) dies als *Nonignorable Random-Coefficient-Based Dropout*. Als Beispiel führt er eine Längsschnittstudie von AIDS-Patienten an; die abhängige Variable bezeichnet die Anzahl bestimmter Immunzellen, Ausfälle geschehen hauptsächlich durch Tod der Patienten, gemessen als überlebenszeit, die als eine lineare Funktion von individuellen, unbekanntem Eigenschaften β_i und der Kovariaten X modelliert wird.

Dauer von Armutsepisoden oder Dauer von Arbeitslosigkeitsepisoden sind Beispiele „klassischer“ sozioökonomischer Fragestellungen, die anhand von Panelbefragungen untersucht werden können, allerdings sind Schätzungen der Dauer dann verzerrt, wenn Ausfälle aus dem Panel von der interessierenden Variable selbst abhängen, bspw. Opportunitätskosten aufgrund eines Übergangs von Arbeitslosigkeit in eine Erwerbstätigkeit steigen und eine Person daher aus dem Panel ausscheidet. Zabel (1998, S. 495f) beschreibt, dass sich bspw. das Arbeitsmarktverhalten und Arbeitskräfteangebot für Teilnehmer und Nonrespondenten unterscheidet: „The results [...] show that attritors are more likely to come from smaller family

sizes, not be married, and not own houses. These characteristics are also indicative of a lower level of labor force attachment [...]. Given this evidence [...] it is likely that these two groups will exhibit different labor supply behavior“. Ist das Arbeitsmarktverhalten oder das Arbeitskräfteangebot abhängige Variable, sind Ausfälle „Missing not at random“ und Schätzungen durch Nonresponse verzerrt. van den Berg et al. (2006) unterscheiden beim Zusammenhang zwischen Arbeitslosigkeitsdauer und Nonresponse zwei Arten; einmal einen direkte kausalen Zusammenhang (wenn eine Person bspw. nicht (mehr) erreichbar ist oder umgezogen ist, weil sie wieder arbeitet) und einen Zusammenhang, der durch die gleichen, (unbeobachteten) individuellen Merkmale bestimmt ist, die sowohl das Arbeitskräfteangebot und die Arbeitsaufnahme bestimmen, als auch die Teilnahme an der Befragung (bspw. wenn Personen nach langer Arbeitssuche wenig motiviert sind, an einer Befragung zum Thema Arbeitssuche teilzunehmen): Sie kommen zu dem Schluss, dass durch Nonresponse verzerrte Schätzungen die Arbeitslosigkeitsdauer unterschätzen. Daneben besteht für viele epidemiologische oder medizinische Fragestellungen die Gefahr für einen Nonresponse-Bias (Cohen und Duffy, 2002, S. 13): „using panel data [...] to analyse longitudinal models of health creates a risk that results will be contaminated by bias that is associated with longitudinal nonresponse. There are drop-outs of the panel at each wave and some of these may be related directly to health [...] others may be indirectly related to health“. Daraus schliessen sie: „the long-term survivors who remain in the panel are likely to be healthier on average compared with the sample at wave 1“ (Jones et al., 2006, S. 543f); zudem wird die Stichprobe im Durchschnitt auch gesünder sein als die Grundgesamtheit, aus der sie ursprünglich stammte (vgl. auch Cohen und Duffy, 2002, S. 14); hier besteht die Gefahr für eine Verzerrung durch nichtzufällige Ausfälle bspw. durch zwei Arten von Ausfällen: zum Einen sind gerade alte und kranke Personen nicht teilnahmefähig, obwohl durch den hohen empfundenen Nutzen besonders zur Teilnahme motiviert, andererseits haben gesunde Personen kein Interesse an einer Befragung zum Thema Gesundheit und verweigern die Teilnahme.

1.4 Arten von Nonresponse, Ausfallgründe

Die Bezeichnung „Nonresponse“ ist dahingehende ungenau, da es unterschiedliche Formen von Nonresponse gibt; das heisst, Nonrespondenten unterscheiden sich evtl. nicht nur von Respondenten, sondern unterscheiden sich auch je nach Form des Nonresponse und können so auch unterschiedliche Auswirkungen auf Ergebnisse und einen möglichen Nonresponse-bias haben (vgl. Singh, 1995, S. 44). Daher ist es wichtig, nicht nur nach Teilnehmern und Nicht-Teilnehmern zu unterscheiden, sondern jede Form des Nonresponse getrennt zu betrachten. In der Literatur werden vor allem drei Arten¹⁰ von Nonresponse beschrieben (vgl. Groves et al., 2004, S. 169f); jede mit unterschiedlichen Ursachen und Konsequenzen. Diese drei Arten von Nonresponse sind Ausfälle durch Nichterreichbarkeit (Abschnitt 1.4.1), Ausfälle durch Verweigerungen (Abschnitt 1.4.2) und Ausfälle durch Teilnahmeunfähigkeit (siehe Abschnitt 1.4.3)¹¹. Für eine Analyse und Korrektur von Nonresponse ist eine getrennte Betrachtung der Nonresponsearten von zentraler Bedeutung, da sich Ursachen und Konsequenzen der einzelnen Ausfälle unterscheiden (Schnell, 1997, S. 133)¹²: „The circumstances

¹⁰„Dem tatsächlichen Geschehen bei Erhebungen werden diese Kategorien aber kaum gerecht“, dem gegenüber stehen „viele verschiedene Ursachen die sich nach den Details des Auswahlverfahrens, der Datenerhebungsmethode und den Feldprozeduren unterscheiden“ (Schnell, 1997, S. 18). Für den Fall einer mündlichen Befragung im Rahmen einer Zufallsstichprobe zählt Schnell (1997) 19 verschiedene Ausfallgründe auf. van den Berg et al. (2006, S. 590f) zählen für eine Befragung zum Arbeitssuchverhalten 22 verschiedene Ausfallgründe auf, die sie für ihre Analyse jedoch in vier Kategorien (verweigert, umgezogen, kein Kontakt, sonstige) zusammenfassen; allerdings ist eine Zusammenfassung nicht ohne Probleme: „It is not always obvious to which type a specific code should be assigned“.

¹¹Eine weitere Form des Nonresponse sind bspw. Versäumnisse bei der Feldarbeit (vgl. Singh, 1995, S. 45), bspw. durch Ausfall einzelner Interviewer bzw. Sampling Points. Oftmals werden zusätzlich noch „neutrale Ausfälle“ und „sonstige Ausfälle“ als Restkategorien verwendet, bspw. unterscheiden Atrostic et al. (2001, S. 211f) lediglich Verweigerungen und Nichterreichbare; Ausfälle durch Tod oder Teilnahmeunfähigkeit fassen sie in eine Residuenkategorie der sonstigen Ausfälle. Zusätzlich gilt, „Causes of nonresponse are not restricted solely to the behavior of respondents but may also arise from other aspects of the survey execution“ (Fay, 1986, S. 354). Djerf (2004, S. 46) beschreibt, dass eine vollständige Welle der (monatlichen) finnischen Arbeitskräfteerhebung im April 1986 aufgrund eines Streiks im öffentlichen Dienst ausfiel.

¹²„[...] muss betont werden, dass die Erreichbarkeit der Befragten andere Ursachen besitzt als die Befragungsbereitschaft. Nonresponse-Studien, die diese beiden Aspekte nicht trennen können, konfundieren zwei

of nonresponse are thus potentially varied and complex. It is easily argued, then, that no single model can correctly reflect the implications of nonresponse in all instances“ (Fay, 1986, S. 354). Auch Honggao und Hill (2005) betonen die Trennung der verschiedenen Nonresponsearten, unterscheiden dann in ihrer Arbeit aber lediglich Ausfälle durch Tod („attrition due to uncontrollable causes“, „passive attrition“) und „attrition due to choice“ (S. 1f): Ihr Ergebnis auf Basis der Britischen Health and Retirement Study (HRS): Ausfälle durch Tod unterscheiden sich von sonstigen Ausfallursachen: passive Ausfälle sind häufiger männlich, zu Beginn des Panels schon älter, nicht erwerbstätig, besitzen geringere kognitive Fähigkeiten; sie sind vermehrt alleinstehend und sie geben häufiger Proxy-Interviews; ihr (berichteter) Gesundheitszustand ist schlechter und verschlechtert sich weiter: „All these results were intuitively reasonable“ (S. 14)¹³. Neben der „Klärung“ und „Beseitigung“ (Schnell, 1997, S. 156) benennen Honggao und Hill (2005) die Reduzierung des Nonresponse als weiteren Grund zur getrennten Betrachtung und Analyse: „If someone drops out of a survey because he or she cannot be reached, for example, finding an effective way to contact respondents may be important to reduce attrition. On the other hand, if someone drops out because of lack of motivation, improving incentive schemes may help him or her to stay“ (S. 2).

Einflüsse auf eine Kontakt- und Teilnahmewahrscheinlichkeit werden auf vier Ebenen beschrieben (vgl. Groves et al., 2004, S. 176 und Groves et al., 1992; ausführlicher siehe dazu Groves und Couper, 1998):

- **soziales Umfeld**; darunter fallen bspw. Stadt-Land-Unterschiede, Bevölkerungsdichte, das Wohnumfeld aber auch Kriminalitätsraten
- **persönliche Merkmale** der Befragungsperson, bspw. Geschlecht, Alter, Einkommen,

Mechanismen mit unterschiedlichen Korrelaten. Dies kann weder zur Klärung noch zur Beseitigung der Folgen von Nonresponse beitragen“ (Schnell, 1997, S. 156). Steeh (1981) untersucht bspw. nur Verweigerungen und sonstige Ausfallgründe („the two major sources of nonresponse“ (S. 41)) und vermischt bei sonstigen Ausfallgründen auch Effekte durch Nichterreichbarkeit. Sie kommt zu dem Ergebnis, dass eine Zunahme der Nonresponseraten vor allem auf gestiegene Verweigerungsraten zurückzuführen ist (S. 52).

¹³Dabei vermischen sie die Effekte der verschiedenen „aktiven“ Ausfallarten: bspw. dürften sich Personen mit nicht durchführbaren Interviews ebenfalls und eventuell auf die selbe Art von bspw. Verweigerern oder Nichterreichbaren unterscheiden.

Erwerbstätigkeit und Bildung, gesellschaftliche Partizipation, oder Merkmale der Haushalte wie Haushaltsgröße und -zusammensetzung

- **Merkmale der Interviewer**, vor allem Erfahrung und Erwartungen der Interviewer¹⁴; „Viewed from the the non-response perspective, interviewers can be seen to vary systematically in terms of their achieved response rates“ (O’Muircheartaigh und Campanelli, 1999, S. 438)
- **Merkmale der Befragung**, darunter sind bspw. Befragungsdauer und Belastung, Thema, Auftraggeber und Erhebungsmodus der Befragung zu fassen.

Die meisten theoretischen Erklärungsansätze zur Teilnahme beziehen sich auf eine oder mehrere Ebene dieser Einflüsse (vgl. Groves et al., 2004, S. 176 und de Leeuw und de Heer, 2002, S. 42). Diese Ansätze werden in Abschnitt 2 näher beschrieben. Aus der Literatur sind viele Variablen bekannt, die mit Nonresponse korrelieren, aber, so Schnell (1997, S. 199), „von einer tatsächlichen Erklärung der Ausfälle [...] kann jedoch kaum eine Rede sein“. Er bezieht sich bei dieser Aussage auf in der Literatur häufig beschriebene demographischen Korrelate von Nonresponse, die „über eine Vielzahl von Mechanismen mit einer Vielzahl von Phänomenen gekoppelt sind“ (S. 199), aber keine vollständige Erklärung darstellen, da sie den zugrundeliegenden Mechanismus nicht angeben. Auch Groves und Couper (1996, 1998) gehen in die gleiche Richtung, wenn sie feststellen: „we do not hypothesize that many of the socio-demographic variables are direct *causal* influences [...]. Rather they are indirect measures of what are essentially social psychological constructs“ (Groves und Couper, 1998, S. 120, Hervorhebung im Original). Zabel (1998, S. 492) kommt bei der Analyse der Pane-

¹⁴Allerdings besteht hier das Problem, dass Interviewereffekte zunächst nicht von regionalen Effekten der Sampling Points getrennt werden können. O’Muircheartaigh und Campanelli (1999) beschreiben ein Experiment innerhalb des Britischen Haushaltspanel BHPS, bei dem Interviewer nicht auf Sampling Points, sondern zufällig auf Befragte aufgeteilt wurden (allerdings auch hier innerhalb „geographic pools“ (S. 439)). Sie unterscheiden zwei Arten von Nonresponse, Verweigerungen und Nichterreichbarkeit, und kommen nach der Analyse anhand multinomialer Logitregressionen zu dem Ergebnis, dass „gute Interviewer“ sowohl niedrigere Verweigerungsraten als auch höhere Kontaktraten besitzen, auch wenn für beide Aufgaben unterschiedliche Fähigkeiten notwendig sind (S. 443).

lausfälle aus dem amerikanischen Survey of Income and Program Participation (SIPP) und der Panels Study of Income Dynamics (PSID) zu dem Ergebnis: „generally, the demographic variables tend to have little impact on attrition“, auch wenn es zwischen Teilnehmern und Nonrespondenten zahlreiche signifikante Unterschiede gäbe.

1.4.1 Erreichbarkeit, Kontakt

Eine erste Gruppe von Nonrespondenten besteht aus Personen oder Haushalten, zu denen kein Kontakt hergestellt werden kann; Ursache dafür kann sein, dass der Interviewer die gewählte Adresse nicht auffindet, der Zugang durch Zugangsbeschränkungen verhindert ist¹⁵, oder zu befragende Personen sind nicht erreichbar. Nichterreichbar sind Personen, die aufgrund ihres Lebensstils häufig nicht zu Hause sind (bspw. Erwerbstätige oder bei Freizeitaktivitäten ausser Haus sowie bei politischer oder gesellschaftlicher Partizipation) oder vorübergehend längere Zeit nicht zu Hause sind (bspw. durch Urlaub oder Krankenhausaufenthalte)¹⁶. Auch wenn keine Einigkeit in der Literatur darüber besteht, ob Nonresponder tatsächlich zunehmen, herrscht doch weitgehende Einigkeit darüber, dass für einen Kontakt durchschnittlich mehr Kontaktversuche unternommen werden müssen als früher. Groves et al. (2004, S. 172) berichten, dass die meisten Haushalte in den USA vorhersehbare Zeitpläne haben, sodass über entsprechende Feldprozeduren Kontaktraten erhöht werden können: „there are very few households in the United States where no one is ever at home in the evening“. Allerdings bestehen Unterschiede in diesen „at-home-pattern“ (S. 173) für verschiedene Subgruppen. Entsprechend steigt die Kontaktwahrscheinlichkeit, werden die Kontaktzeiten variiert und die Anzahl der Kontaktversuche erhöht. Auch Interviewereinflüsse scheinen weniger mit den Fähigkeiten der Interviewer als solchen zusammenzuhängen, als vielmehr mit den von ihnen den gewählten Kontaktzeiten (Schnell, 1997, S. 225). Die Erreichbarkeit hängt neben der Anzahl der Kontaktversuche auch wesentlich von der Länge

¹⁵Das scheint insbesondere bei telefonischen Erhebungen zunehmend Probleme zu bereiten, bspw. durch Anrufbeantworter und Rufnummernanzeige.

¹⁶Eine dritte kleine Gruppe besteht aus Personen, die zwar prinzipiell zu Hause erreichbar sind, aber die nicht auf Kontaktversuche reagieren (Schnell, 1997, S. 219).

der Feldzeit (und der Verteilung der Kontaktversuche über die Feldzeit) ab; je länger die Feldzeit, desto höher die Chance, dass Kontakt auch zu schwererreichbaren Haushalten oder Personen hergestellt werden kann¹⁷, die temporär bspw. in Urlaubszeiten oder aufgrund von Krankenhausaufenthalten oder Dienstreisen nicht zu Hause anzutreffen sind¹⁸. Einflüsse auf eine Kontaktwahrscheinlichkeit beziehen sich hauptsächlich auf Merkmale der sozialen Umwelt und soziodemographische Merkmale der Zielhaushalte bzw. Zielpersonen. Entsprechend einer Lebenszyklushypothese gelten junge Personen und kleine, vor allem Ein-Personen-Haushalte als schwierig zu kontaktieren (ohne sich allerdings bei ihrer Teilnahmebereitschaft zu unterscheiden), bei großen Haushalten und Haushalten mit (kleinen) Kindern und Haushalten mit nicht (mehr) erwerbstätigen Personen ist dagegen die Kontaktwahrscheinlichkeit höher. Groves und Couper (1998, S. 85ff) berichten zudem, dass für dichtbesiedelte und städtische Gebiete zum einen häufiger kein Kontakt hergestellt werden kann als auch durchschnittlich mehr Kontaktversuche bis zu einem erfolgreichen Kontakt unternommen werden müssen. Gleiches gilt für Gebiete mit hoher Kriminalitätsrate¹⁹. Zusammenfassend stellen Groves und Couper (1998, S. 94) fest: „Many of the population trends over the last few decades have conspired against easy contact with sample households. There are more households with all adults employed outside the home. There are more single-person households. There are fewer households with young children at home. [...] Urban areas are more filled with

¹⁷Djerf (2004, S. 41f) berichtet für die finnische Arbeitskräfteerhebung, dass dort der Anteil der Nicht-erreichbaren rund zwei Drittel der Nonrespondenten ausmacht. „The noncontact rate has always been the dominant part of the nonresponse because the fieldwork period is relatively short“.

¹⁸Schnell (1997, S. 237ff) gibt den Anteil der Personen im Urlaub je nach Monat zwischen 1% und 10% an; den Anteil der zu einem Stichtag aufgrund eines Krankenhausaufenthalts nicht zu Hause erreichbaren Personen schätzt er auf 1%.

¹⁹Allerdings bemerken Groves und Couper (1998) dazu, dass die Kriminalitätsrate gemessen auf einer hohen Aggregationsebene möglicherweise ein schlechter Indikator für Kriminalitätsraten auf der Ebene der Auswahlbezirke ist, zudem sind auch gegenläufige Effekte der Kriminalitätsrate auf eine Kontaktwahrscheinlichkeit denkbar, wenn Personen in Gebieten mit hoher Kriminalitätsrate häufiger zu Hause anzutreffen sind (S. 87). Allerdings verschwinden Effekte der Umgebung in der multivariaten Analyse, berücksichtigt man zusätzliche Haushaltsmerkmale wie Haushaltsgröße. Zudem, so Schnell (1997, S. 176) korreliere „die subjektive Kriminalitätsfurcht zwar mit der Ortsgröße, aber kaum mit der tatsächlichen Kriminalitätsbelastung“.

housholds and housing units that pose problems for contact“.

1.4.2 Ausfälle durch Verweigerung

Ist eine Kontaktadresse aufgefunden und ein Kontakt hergestellt, können Personen oder Haushalte durch eine explizite Verweigerung der Teilnahme ausfallen²⁰. Auch hier bestehen Hinweise, dass, wenn auch kein Rückgang der Kooperation stattfindet, größere Anstrengungen notwendig sind, Personen oder Haushalte zur Teilnahme zu bewegen²¹. Der Interaktion zwischen Interviewer und Kontakt- bzw. Zielpersonen wird hierbei eine zentrale Rolle zugeschrieben. Die Entscheidung an einer Befragung teilzunehmen wird in wenigen Augenblicken gefällt²². In dieser kurzen Zeit muss der Interviewer eventuelles Misstrauen gegenüber Fremden auf Seiten der Kontakt- oder Befragungsperson abbauen, Ziel, Auftraggeber und Inhalt der Befragung vorstellen und auf Fragen und äusserungen des Gegenüber reagieren. Groves et al. (2004, S. 177): „No one introduction is suitable to adress the concerns of diverse sample persons. Interviewers must have ways of learning the concerns to make salient those attributes given positively leverage by the sample persons“²³. Dies ist ein Grund, weshalb häufig für erfahrene Interviewer niedrigere Verweigerungsraten berichtet werden und spezielle Interviewtechniken zur Vermeidung und Minimierung von Verweigerungsraten diskutiert werden. In der Literatur werden diese Techniken als „*Tailoring*“ und „*Maintaining Interaction*“ beschrieben (Groves und Couper, 1998, S. 248ff). Während *Tailoring* Maßnahmen umfasst,

²⁰Razafindratsima und Kishimba (2004, S. 359/366) unterscheiden bei Verweigerungen zusätzlich zwischen Verweigerungen, die am Ende des Interviews der letzten Welle stattfinden und Verweigerungen, nachdem in einer folgenden Welle wieder Kontakt hergestellt wurde.

²¹„In sum, we believe that cooperation with surveys is declining in the United States, or at least it is becoming harder to maintain the same levels of cooperation“ (Groves und Couper, 1998, S. 166).

²²„Decisions to decline a request are made quickly (on the telephone most refusals take place in less than 30 seconds)“ (Groves et al., 2004, S. 175). In persönlichen Befragungen ist der erste Kontakt etwas länger, O’Brien et al. (2002, S. 2502) berichten von 1 bis 5 Minuten bis zum Start des Interviews oder Ende des Kontakts.

²³Damit ist das Verhalten der Interviewer einem Widerspruch ausgesetzt: zum Einen ist der Prozess vom Kontakt bis zur Teilnahmebereitschaft unstrukturiert, das Erhebungsinstrument selbst wiederum ist in der Regel hoch strukturiert (vgl. auch Groves und Couper, 1998, S. 195).

wie den Befragungspersonen möglichst „maßgeschneidert“ entgegengetreten werden kann, umfasst *Maintaining Interaction* Maßnahmen, wie ein Kontakt möglichst lange Aufrecht erhalten und möglichst viele Information für den Interviewer generiert werden kann, damit dieser sein Auftreten an Sorgen und Nöte der Befragungsperson anpassen kann²⁴. Interviewern werden vor allem aufgrund ihrer Erfahrung als Interviewer und ihren Erwartungen an ihre Aufgabe Einflüsse auf eine Kontakt- und Teilnahmewahrscheinlichkeit unterstellt. Erfahrene Interviewer können auf ein größeres Repertoire an Verhaltensweisen zurückgreifen, um im Kontakt mit Befragungspersonen eine Teilnahme zu erreichen²⁵, neu eingesetzte, unerfahrene Interviewer können nicht auf dieses Repertoire an bewährten Strategien, besonders im Umgang mit Verweigerern zurückgreifen und führen so eventuell zu einem hohen Anteil von „weichen Verweigerern“ (O’Brien et al., 2002, S. 2502)²⁶.

Gründe für Nichterreichbarkeit und Verweigerung können sich unterscheiden, ebenso können sich verschiedenen Gruppen in ihrer Erreichbarkeit und Teilnahmebereitschaft unterscheiden; junge Personen gelten bspw. aufgrund ihres Lebensstils schwierig zu erreichen, unterscheiden sich dann aber kaum in ihrer Teilnahmebereitschaft, im Gegenteil, DeMaio (1980, S. 227f) kommt zu dem Ergebnis, dass gerade jüngere Personen unter 30 Jahren eine höhere Teil-

²⁴Im *Maintaining Interaction* liegen auch große Unterschiede zwischen kommerziellen und wissenschaftlichen Umfragen; eine Strategie des *Maintaining Interaction* lautet, die Anzahl der Interviews in einer Zeiteinheit zu maximieren, die andere, gegensätzliche Strategie lautet, die Anzahl der Verweigerungen zu minimieren und jede der ausgewählten Befragungspersonen zu interviewen; Groves et al. (1992, S. 489) bemerken dabei, „[...] how similar the goals of a quota sample interviewer are to those of any salesperson, but how different are those of the probability sample interviewer“.

²⁵Allerdings ist die Untersuchung von Interviewereffekten auch mit Problemen behaftet: Schnell (1997, S. 224) nennt hier u.a. eine mögliche Selbstselektion erfolgreicher Interviewer oder dass Interviewereffekte je nach Zuteilung der Interviewer zu den Befragungspersonen nicht von Stichprobenklumpen, bspw. Sampling-Points getrennt werden können. Dafür müssten Interviewer zufällig den Befragungspersonen zugeteilt werden (siehe auch Groves und Couper, 1998, S. 192f).

²⁶Das wird auch dann zu einem Problem, wenn unerfahrene Interviewer Reaktionen der Befragungspersonen falsch interpretieren und daraufhin Ausfallgründe nicht korrekt erfassen: „inexperienced interviewers often interpret respondent concerns expressed as questions as signaling reluctance. Questions, however, often lead to complete interviews for a number of reasons“ (O’Brien et al., 2002, S. 2502).

nahmebereitschaft besitzen²⁷. In städtischen Gebieten ist dagegen die Kontakt- als auch Teilnahmebereitschaft niedriger als in ländlichen Gebieten.

1.4.3 Ausfälle durch Teilnahmeunfähigkeit

Ist eine Zieladresse aufgefunden und sind Personen grundsätzlich teilnahmebereit fallen einige Personen trotzdem aus der Stichprobe aufgrund mangelnder Fähigkeit, an der Befragung teilzunehmen. Diese Ausfälle spielen vor allem bei speziellen Populationen eine Rolle. Teilnehmern ist es bspw. aufgrund von Sprachschwierigkeiten nicht möglich, mit dem Interviewer zu kommunizieren oder sind nicht in der Lage, den Fragebogen zu beantworten²⁸. Zudem können Teilnahmeschwierigkeiten aufgrund von Alter und körperlicher und geistiger Einschränkung auftreten und eine Teilnahme unmöglich machen²⁹. Schließlich erschwert oder verhindert eine Lese- und Rechtschreibschwäche und Analphabetismus eine Teilnahme bei schriftlich-postalischen Befragungen³⁰. Für nicht teilnahmefähige Personen muss eine Regelung getroffen (und eingehalten) werden, ob hier Proxy-Interviews erlaubt und durchgeführt werden müssen (und als solche gekennzeichnet werden), oder ob jegliche Proxy-Interviews untersagt sind und lediglich der zutreffende Ausfallgrund erfasst wird.

Entsprechend dieser Unterscheidung der unterschiedlichen Nonresponsearten besteht das vollständige Sample nun nicht mehr nur aus Teilnehmern und Nichtteilnehmern (vgl. Gleichung 1.1).

²⁷Im Verlaufe der Wiederholungsinterviews des Current Population Survey CPS verweigern junge Personen allerdings gleich häufig wie ältere (DeMaio, 1980, S. 228).

²⁸Da das SOEP seit Beginn auch aus einer separaten Ausländerstichprobe bestand, stehen seit Beginn auch fremdsprachige Fassungen der Fragebögen als auch der Übersetzungshilfen in den Interviewerhandbüchern zur Durchführung der Interviews zur Verfügung. Übersetzungen sind in englisch, türkisch, serbokroatisch, griechisch, italienisch und spanisch vorhanden.

²⁹Hier kann vor allem im Anstaltsbereich ein Gatekeeper-Problem auftreten: nicht der Interviewer stellt die Teilnahmeunfähigkeit fest, sondern eine dritte Person verweigert den Zugang zur Befragungsperson mit dem Hinweis auf Teilnahmeunfähigkeit (vgl. Schnell, 1997, S. 178f).

³⁰Goyder et al. (2002, S. 6) berichten bspw., dass in Kanada über 30% (!) der Bevölkerung zwischen 16 und 69 Jahre Leseschwächen in alltäglichen Leseanforderungen besitzen.

chung 1.7), sondern vielmehr mindestens aus

$$n = \overbrace{n_{\text{Interview}}^r} + \overbrace{n_{\text{verweigert}} + n_{\text{nichterreichbar}} + n_{\text{nicht teilnahmefähig}} + n_{\text{sonstige}} + \dots}^m$$

und eine Nonresponse-Quote besteht entsprechend aus

$$\begin{aligned} NR &= 1 - \frac{r}{n} \\ &= \frac{m}{n} \\ &= \frac{\text{verweigert}}{n} + \frac{\text{nichterreichbar}}{n} + \frac{\text{nicht teilnahmefähig}}{n} + \frac{\text{sonstige}}{n} + \dots \end{aligned}$$

Damit können die unterschiedliche Ausfallarten auch unterschiedliche Ausfallmechanismen besitzen und auf unterschiedliche Art zu einem möglichen Nonresponse-Bias beitragen (vgl. Groves, 2006, S. 649f). Ausgehend von Gleichung 1.15 ergibt sich der beobachtete Mittelwert \bar{y}_r als (vgl. auch Groves und Couper, 1998, S. 12f):

$$\bar{y}_r = \bar{y}_n + \left(\frac{m_{rf}}{n}(\bar{y}_r - \bar{y}_{rf}) + \frac{m_{nc}}{n}(\bar{y}_r - \bar{y}_{nc}) + \frac{m_{nf}}{n}(\bar{y}_r - \bar{y}_{nf}) + \dots \right) \quad (1.27)$$

mit

m_{rf} = Anzahl Verweigerungen

m_{nc} = Anzahl Nichterreichbare

m_{nf} = Anzahl Nichtteilnahmefähige

n = Umfang Stichprobe ohne Nonresponse

\bar{y}_n = Mittelwert der Stichprobe ohne Nonresponse

\bar{y}_r = beobachteter Mittelwert

\bar{y}_{rf} = Mittelwert Verweigerer

\bar{y}_{nc} = Mittelwert Nichterreichbare

\bar{y}_{nf} = Mittelwert Nichtteilnahmefähige

Damit ist es leicht zu zeigen, dass die unterschiedlichen Ausfallarten unterschiedliche Auswirkungen auf einen Nonresponse-Bias haben können; aus diesem Grund ist es von zentraler Bedeutung für die Untersuchung auf einen möglichen Nonresponse-Bias, unterschiedliche

Ausfallarten zu unterscheiden und getrennt auf ihre Unterschiede zu teilnehmenden Elementen der Stichprobe zu untersuchen: „ideally, the $(nr/n)(y_r - y_{nr})$ component of the equation should be replaced by separate terms of each type of non-response“ (Gray et al., 1996, S. 165). Groves und Couper (1998, S. 13) beschreiben dies an einem Beispiel eines Viktimisierungssurveys. Sie beschreiben, dass bspw. nichterreichbare Personen öfter Opfer eines Verbrechens wurden, und aus diesem Grund auch nicht erreichbar sind (bspw. durch Zugangsbeschränkungen zum Grundstück), wohingegen Personen, die seltener Opfer eines Verbrechens wurden kein Interesse am Thema „Viktimisierung“ haben und daher auch eine Teilnahme an der Befragung verweigern; hier können die Auswirkungen auf einen möglichen Nonresponse-Bias in entgegengesetzte Richtungen zeigen:

$$\frac{m_{rf}}{n}(\bar{y}_r - \bar{y}_{rf}) > 0 \quad \text{für die Verweigerer}$$

$$\frac{m_{nc}}{n}(\bar{y}_r - \bar{y}_{nc}) < 0 \quad \text{für die Nichterreichbaren}$$

Ein weiteres anschauliches Beispiel sind Befragungen mit Gesundheitsbezug: hier kann vermutet werden, dass Personen mit schlechterem Gesundheitszustand verstärkt nicht erreichbar (bspw. durch Krankenhausaufenthalte) oder aber verstärkt nicht teilnahmefähig sind, wohingegen (überdurchschnittlich) gesunde Personen kein Interesse an der Befragung haben und daher eine Teilnahme verweigern (vgl. auch Cohen und Duffy, 2002). Werden die verschiedenen Ausfallursachen also nicht unterschieden, können sich ihre Auswirkungen vermischen; eine Korrektur eines möglichen Nonresponse-Bias ist dann allerdings nicht möglich, da sich auch die Kovariaten und Prädiktoren für die verschiedenen Ausfallursachen vermischen (siehe dazu Abschnitt 3 ab Seite 53).

1.5 Nonresponse in Panelerhebungen, Panelausfälle

Im Längsschnitt ist eine Unterscheidung zwischen Unit- und Item-Nonresponse nicht immer klar: Nonresponse in einer Welle stellt sich im Längsschnitt zunächst einmal als eine Form des Item-Nonresponse dar, betrachtet man das Panel aus einer Querschnittsperspektive Welle für Welle, ist Nonresponse in einer Welle als Unit-Nonresponse zu betrachten (vgl. Kalton, 1986, S. 304). Je nach Blickwinkel bieten sich unterschiedliche Behandlungen

des Wellen-Nonresponse an: während für Item-Nonresponse in der Regel Imputationsverfahren angewendet werden, wird Unit-Nonresponse in der Regel durch Gewichtung behoben (S. 304). Taris (2000, S. 20f) unterscheidet im Längsschnitt „Initial Nonresponse“ in der ersten Welle, wenn Personen oder Haushalte nicht in das Panel aufgenommen werden können, „Attrition“, wenn Panelteilnehmer nach einer oder mehrerer Wellen endgültig aus dem Panel durch Nonresponse ausscheiden und schließlich „Wave-Nonresponse“ als temporäre Ausfälle von Panelteilnehmern, die nach einem Ausfall wieder in das Panel zurückkehren (vgl. auch Lepkowski und Couper, 2002, S. 259f; Hawkes und Plewis, 2006, S. 479).

Ausfallursachen und Auswirkungen von Nonresponse können sich bei Panelerhebungen im Vergleich zu Querschnittserhebungen unterscheiden (Lynn, 2006, S. 393), zudem sind „Ausfallwahrscheinlichkeiten in den verschiedenen Kategorien bei Panelerhebungen nach der 1. Welle kaum mit denen bei Querschnittsstudien vergleichbar“ (Schnell, 1997, S. 53). Für Panelteilnehmer späterer Wellen sind Namen, Adresse und Informationen aus der vorhergehenden Welle bekannt, so dass eine Kontaktaufnahme erleichtert wird; zudem sind Teilnehmern bei einer erneuten Kontaktaufnahme Inhalt und das Prozedere der Befragung sowie in der Regel der Interviewer aus der vorhergehenden Welle bekannt, daher können Nichterreichbarkeit und Verweigerung bei Querschnittserhebungen und Panelerhebungen durch unterschiedliche Mechanismen bedingt sein (vgl. Schnell, 1997, S. 143f; oder Lepkowski und Couper, 2002): „cooperation on a subsequent wave is generally predicted by prior cooperation“ (Singer, 2002, S. 167). Ein besonderes Problem von Nonresponse bei Panelerhebungen besteht darin, dass Nonresponse über die einzelnen Wellen kumuliert, also von Welle zu Welle Personen aus dem Panel durch Nonresponse ausscheiden (der Begriff *Panel Attrition* oder „Abnützung“ ist dafür besonders anschaulich). Dadurch verringert sich zuerst einmal die Fallzahl von Welle zu Welle³¹. Führt Nonresponse zu Verzerrung durch Nonresponse-Bias, wird das Panel durch diese Ausfälle von Welle zu Welle immer selektiver³²: „Even if

³¹Zumindest die Anzahl der ursprünglichen Stammpersonen verringert sich; je nach Weiterverfolgungsregel (siehe Abschnitt 4.3) können allerdings auch neu zu befragende Haushalte und Personen in das Panel eintreten, bspw. durch Zuzüge von Nicht-Stammpersonen in Stammhaushalte oder durch das erstmalige Erreichen des Befragungsalters von Kindern in Stammhaushalten.

³²Laurie et al. (1999, S. 278) befürchten, dass das Sample durch den Ausfall von Verweigerern über die

a panel starts out with a good representation of the population of interest, small attrition rates can quickly accumulate and render the sample unrepresentative“ (Hill und Willis, 2001, S. 416). Für die Analyse von Panelausfällen liegen zudem mehr Informationen vor als bei Querschnittsbefragungen (vgl. Bose und West, 2002). Im Querschnitt ist lediglich spärliche Information des Sampling-Frames vorhanden³³, um Ausfälle zu untersuchen und einen Nonresponse-Bias abzuschätzen. Für die Analyse von Panelausfällen liegt für Ausfälle in einer späteren Welle aber detailliert Information aus den vorhergehenden Wellen vor (zumindest aus dem Basisjahr der ersten Welle). Unterschiede der Ergebnisse auf Basis der Teilnehmer einer früheren Welle und der Respondenten der folgenden Welle können dann Hinweise auf einen Nonresponse-Bias durch Ausfälle zwischen diesen Wellen liefern (vgl. Bose und West, 2002, S. 280); allerdings wenden Gray et al. (1996, S. 164) ein, dass auch ein Vergleich der Merkmale von Panelteilnehmern späterer Wellen mit denen früherer Wellen zur Untersuchung auf einen möglichen Nonresponse-Bias problematisch ist: „the description of non-respondents is limited to individuals who initially participate in the survey but do not do so subsequently. These people may be different from those who never participate at all“.

Wellen immer kooperativer wird; das gilt aber nur, wenn man annimmt, das Verweigerungsverhalten eine über die Zeit stabile Verhaltensdisposition ist; dann müsste sich auch zeigen lassen, dass mit späteren Wellen immer weniger Ausfälle aufgrund von Verweigerung stattfinden. Vergleiche dazu auch Little et al. (2000, S. 188): „Those who continue may engender characteristics that also bias the study’s results [...] Therefore, not only do the characteristics of dropouts affect outcomes of a study, the characteristics of the continuers do as well“.

³³Eventuell liegen aber auch bei Querschnittsbefragungen zusätzliche Informationen für Nonrespondenten vor, bspw. durch Beobachtungen der Interviewer des Wohnumfelds. Goyder et al. (2002) verwenden in einem Survey im kanadischen Ontario bspw. Fotografien aller ausgewählten Haushalte, die bei der Zustellung der Ankündigungsschreiben erstellt wurden, um den sozioökonomischen Status auch der Nonrespondenten zu schätzen und validieren die Angaben der Rater anhand öffentlich zugänglicher Register: „our results indicate that such photos can be rated for SES with acceptable consensus between judges“ (S. 7); Angaben der Rater und Registerdaten zum Wert der Häuser teilen 60% der gesamten Varianz (S. 10). Allerdings gibt es auch Frame-Daten, die vergleichsweise umfangreich sind; ein Beispiel dafür beschreibt Kennickell (1999, S. 283) bei der Analyse von Nonresponse des amerikanischen Survey of Consumer Finances SCF. Der SCF ist eine Registerstichprobe und „the survey is fortunate in having extensive frame data on income and some other characteristics for the entire list sample“ (S. 283).

Zudem wird bei einem Vergleich der Teilnehmer späterer Wellen mit denen früherer Wellen die Möglichkeit einer „natürlichen Veränderung“ der Merkmale über die Zeit ignoriert.

Ausfälle in der ersten Welle sind größer als Ausfälle zwischen den Wellen, zudem spielt die Bestimmung der Ausfallwahrscheinlichkeiten in der ersten Welle eine besondere Rolle, da auf dieser Basis Ausfallwahrscheinlichkeiten über die nächsten Wellen fortgeschrieben werden (Rendtel und Projektgruppe Sozio-Oekonomisches Panel, 1995, S. 138).

Grundsätzlich kommen auch bei Panelerhebungen die oben genannten Gründe für Nonresponse vor (siehe Abschnitte 1.4.1 bis 1.4.3), während bei einmaligen Querschnittserhebungen Ausfälle per Design „absorbierend“, das heißt, endgültig sind, können bei Panelerhebungen Ausfälle temporär sein, das heißt, ein Haushalt oder eine Befragungsperson kehrt nach einem Ausfall in das Panel zurück. Little et al. (2000, S. 191) unterscheiden unterschiedliche Subgruppen der Ausfälle; das sind neben Ausfällen vor der ersten Welle frühe Ausfälle nach wenigen Wellen im Panel, späte Ausfälle, Ausfälle, die ins Panel zurückkehren und neue Teilnehmer; eine Identifizierung und Unterscheidung dieser Gruppen, so Little et al., ist bspw. hilfreich, um die Selektivität des Panels über die Zeit anhand der Unterschiede dieser Gruppen zu untersuchen. Diese Unterscheidung führt Little et al. (2000, S. 193f) schließlich auch zu einer Analyse der Ausfälle auf Basis der Teilnahmeprofile.

1.5.1 Absorbierende Befragungszustände

Die Definition von absorbierenden Zuständen hängt unter anderem von den Weiterverfolgungsregeln ab, also davon, wie lange und wie weit Haushalte und Personen in der nächsten Welle gefolgt wird. Absorbierende Zustände des Befragungsstatus in einem Panel sind nicht (mehr) auffindbare Haushalte, die an der alten Adresse nicht mehr auffindbar sind (bspw. durch eine Haushaltsauflösung oder durch Umzug des Haushalts) und eine neue Adresse nicht zu ermitteln ist oder die neue Adresse außerhalb des durch die Weiterverfolgungsregeln festgelegten Erhebungsgebietes liegt³⁴. Daneben stellen auch Verweigerungen je nach

³⁴Im deutschen SOEP bspw. werden Haushalte und Personen bei einem Umzug ins Ausland nicht weiterverfolgt.

Weiterverfolgsregel einen absorbierenden Zustand dar³⁵; andauernde Krankheit oder Tod der Befragungsperson ist demgegenüber ein natürlicher absorbierender Zustand.

1.5.2 Nicht-absorbierende Befragungszustände

Temporäre Ausfälle in einer Welle werden demgegenüber in der nächsten Welle wieder kontaktiert, und es wird versucht, eine Teilnahme zu erreichen. Temporäre Ausfälle umfassen je nach Weiterverfolgsregel bspw. Haushalte oder Personen, die zwar an der alten Adresse oder nach einem Umzug an der neuen Adresse aufgefunden, dort auch nach mehrmaligen Kontaktversuchen aber nicht erreichbar sind. Temporäre Ausfälle sind bspw. auch durch Krankheit bedingt, wenn eine Befragungsperson vorübergehend nicht teilnahmefähig ist. Hinzu kommen temporäre Ausfälle durch sonstige Gründe, die eine Befragung verhindern, aber bei denen Haushalte oder Personen nicht endgültig eine weitere Teilnahme verweigern.

1.5.3 Ausfallmuster bei Panelausfällen

Tabelle 1.1 zeigt einige Ausfallmuster, die aufgrund endgültiger und temporärer Ausfälle auftreten (vgl. Burkam und Lee, 1998, S. 556/558), dabei lassen sich drei Muster unterscheiden: den Idealfall des Panels stellen „Vollzeit-Teilnehmer“ dar, die über alle Wellen des Panels erfolgreich interviewt wurden (die ersten drei Zeilen der Tabelle 1.1).

Monotone Ausfallmuster entstehen, wenn ein Ausfall in einen absorbierenden Zustand führt, ein Teilnehmer bis zu diesem Ausfall aber über alle vorherigen Wellen erfolgreich interviewt wurde³⁶; nichtmonotone Ausfallmuster ergeben sich durch temporäre Ausfälle, d.h. Teilneh-

³⁵Im SOEP werden harte Verweigerungen und wiederholte Verweigerungen in zwei aufeinanderfolgenden Wellen in der nächsten Welle nicht mehr kontaktiert, diese Ausfälle sind endgültig.

³⁶„Block 1 Variables are consequently *more observed* than Block 2 variables“ (Little und Rubin, 1987, S. 6, Hervorhebung im Original); *monotone* Teilnahmemuster werden auch als *nested*, *verschachtelte* Teilnahmemuster bezeichnet (vgl. bspw. Fay, 1986, S. 355). Ein monotones Muster ergibt sich natürlich auch für Teilnehmer, die bis zu einem endgültigen, absorbierenden Ausfall immer temporär ausgefallen sind; allerdings sieht bspw. das deutsche SOEP vor, dass nur erfolgreich interviewte Haushalte in der ersten Welle in der nächsten Welle wieder zu interviewen sind, d.h. sie besitzen mindestens ein erfolgreiches Interview in der ersten Welle.

Tabelle 1.1: Ausfallmuster am Beispiel eines Panels mit 7 Wellen

Nr.	Person	Welle							Muster
		W_1	W_2	W_3	W_4	W_5	W_6	W_7	
1	I_{11}	•	•	•	•	•	•	•	Full Time
2	I_{12}	•	•	•	•	•	•	•	Full-Time
3	I_{1n}	•	•	•	•	•	•	•	Full-Time
4	I_{21}	•	•	•	•	•	---	*	monoton
5	I_{22}	•	•	•	---	*	-	-	monoton
6	I_{23}	•	•	---	*	-	-	-	monoton
7	I_{2n}	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
8	I_{31}	•	---	+	•	•	---	*	nichtmonoton
9	I_{31}	•	•	---	+	•	•	•	nichtmonoton
10	I_{32}	•	---	+	---	+	•	•	nichtmonoton
11	I_{3n}	•	---	+	•	---	+	---	nichtmonoton

•: Teilnahme; ---+: temporärer Ausfall; ---*: endgültiger Ausfall

mer kehren nach einem Ausfall wieder in das Panel zurück und können bspw. in der folgenden Welle erfolgreich interviewt werden (vgl. auch Burkam und Lee, 1998, S. 558). Zusätzlich kommen in Panelbefragungen je nach Weiterverfolgungsregel auch „Späteinsteiger“ vor (vgl. Kalton, 1986, S. 307), bspw. durch das erstmalige Erreichen des Befragungsalters bei Jugendlichen im Haushalt. Für eine Imputation monotoner Ausfälle können iterative Imputationsverfahren angewandt werden und Welle für Welle fehlende Werte eingesetzt werden, allerdings bereiten iterative Imputationsverfahren auch bei einfachen Imputationen beträchtlichen Aufwand (vgl. Marini et al., 1980, S. 322). Nichtmonotone Ausfallmuster bereiten für die Gewichtung und Imputation fehlender Werte spezielle Probleme (vgl. Kalton, 1986, S. 306ff): „Nonmonotone missingsness complicates the modeling task, the estimation task, and the imputation task“ (Rubin, 1987, S. 188). Bei nichtmonotonen Mustern erhöht sich zunächst einmal die Anzahl möglicher kausaler Mechanismen, die zu einem Ausfall führen, eine Annahme eines ignorierbaren Ausfallmechanismus ist schwerer zu halten (vgl. Fay, 1986, S. 360). Allerdings bestehen mehrere Lösungen des Problems bei nichtmonotonen Mustern. Eine erste und einfache Lösung besteht darin, die Information derjenigen Wellen zu ignorieren³⁷, welche monotone Muster zu nichtmonotonen Muster machen (Marini et al., 1980,

³⁷Burkam und Lee (1998) untersuchen anhand der amerikanischen High-School and Beyond-Studie (HS&B)

S. 322), bspw. bei dem Muster mit der Nummer 9 in Tabelle 1.1: der temporäre Ausfall in Welle 3 zerstört hier zunächst ein Vollzeit-Muster; ignoriert man jedoch die die Wellen nach dem Ausfall, ergibt sich aus dem nichtmonotonen Muster über alle Wellen ein monotonen Ausfallmuster mit vollständiger Information bis Welle 3; gilt die Annahme, dass Ausfälle auch nach dem Ignorieren noch zufällig, (zumindest) *Missing at Random*, (*MAR*) sind, sind Ergebnisse zwar konsistent aber weniger effizient (Marini et al., 1980, S. 323). Wie groß ein Informationsverlust ist, hängt davon ab, wieviele Wellen ignoriert werden müssen, aber auch von der Abhängigkeit der Information zwischen den Wellen; können Werte der Welle 3 aus Werten früherer Wellen gut vorhergesagt werden, ist der Informationsverlust entsprechend geringer (vgl. Rubin, 1987, S. 190). Eine weitere Möglichkeit ist die Annahme bedingter Unabhängigkeit³⁸ einzelner Wellen gegeben die Werte der vorherigen Wellen³⁹, so dass für eine Imputation eine monotone Struktur der Ausfälle unterstellt wird (vgl. Rubin, 1987, S. 190f). Betrachtet man Muster 10 in Tabelle 1.1, heisst das, Werte von Welle 3 und Welle 4 sind bedingt unabhängig, gegeben Welle 1 und 2. Fehlende Werte der Welle 2 werden auf Basis von Welle 2 gegeben Welle 1 imputiert, fehlende Werte in Welle 3 werden imputiert auf Basis von Welle 3 gegeben Welle 1 und 2, schließlich werden Werte des nichtmonotonen Musters

die Auswirkungen auf Parameterschätzungen, wenn bei der Analyse Beobachtungen mit nichtmonotonen (und monotonen) Mustern ignoriert werden und die Analyse lediglich auf Vollzeit-Teilnehmern basiert. Sie kommen zu dem Ergebnis, dass sich Vollzeit-Teilnehmer von Teilnehmern mit monotonen und nichtmonotonen Teilnahmemustern unterscheiden: Frauen, Teilnehmer mit höherem sozio-ökonomischen Status, weißer Hautfarbe und mit geringerem Item-Nonresponse sind häufiger Vollzeit-Teilnehmer; Teilnehmer mit monotonen bzw. nichtmonotonen Mustern sind sich im Vergleich zu Vollzeit-Teilnehmern ähnlicher (S. 564f). Ein Nonresponse-Bias durch Ausschluss der Beobachtungen mit monotonen oder nichtmonotonen Mustern zeigt sich nach ihren Ergebnissen zu Schulleistungen in einer überschätzung des Effektes der Herkunft der Absolventen.

³⁸ „In some cases, such assumptions of conditional independence may be perfectly reasonable“, allerdings „multiply-imputed data sets created under this independence assumption will tend to confirm this independence even if it is not an accurate reflection of reality“ (Rubin, 1987, S. 191). Das ist ein weiterer Grund, weshalb imputierte Werte als solche im Datensatz markiert werden müssen.

³⁹Bei Rubin (1987, S. 190) ist das die Variable bzw. der Variablenblock „X“, beispielsweise können das auch Informationen des Sampling-Frames oder Interviewerbeobachtungen, Kontakt- oder Quartiersinformationen sein.

in Welle 4 durch Welle 4, gegeben Welle 1 und 2 imputiert. Marini et al. (1980) stellen eine alternative Lösung des Problems bei nichtmonotonen Ausfallmuster vor, die eine Maximum-Likelihood-Schätzung der Parameter ohne Imputationen sondern durch Manipulation der Varianz-Kovarianzmatrix ermöglicht.

1.5.4 Teilnahmesequenzen

Betrachtet man nicht nur das Muster von Teilnahme - Nichtteilnahme einer Beobachtung i über alle Wellen t , sondern unterscheidet (zumindest) die verschiedenen vorübergehenden Ausfallgründe (*vorübergehend*) *nichterreichbar*, *Verweigerung* und (*vorübergehend*) *nicht teilnahmefähig* (siehe Abschnitt 1.4) sowie die absorbierenden, endgültigen Ausfälle *ins Ausland verzogen*, *dauerhaft nicht teilnahmefähig*, *endgültig verweigert* und schließlich auch *verstorben*, dann ergibt sich für jede Beobachtung eine Sequenz des Befragungsstatus über alle Wellen. Diese Sequenzen sind also entweder vollständig bzw. im Falle eines laufenden Panels rechtszensiert (eine Beobachtung hat an allen Wellen erfolgreich teilgenommen) oder enden vorzeitig in einem der absorbierenden Zustände *verstorben*, *unbekannt verzogen*, *dauerhaft nicht teilnahmefähig* oder *endgültig verweigert*. Für eine Betrachtung der Teilnahmesequenzen ist zentral, dass Ausfallgründe von den Interviewern im Feld möglichst detailliert erhoben und in den Kontaktprotokollen verzeichnet werden; zudem müssen Interviewer Ausfallkategorien zum Einen für alle Beobachtungen einer Welle, zum Anderen auch über alle Wellen gleich verwenden und nicht bestimmte Kategorien als „Restkategorien“ oder bspw. als „versteckte Verweigerungen“ verwenden. Schnell (1997, S. 62f) beschreibt, dass es über die Zeit bspw. in den Erhebungsinstituten zu Änderungen der Klassifikation der Ausfallursachen, als auch zu Umkodierungen durch die Interviewer kommen kann, zudem seien Kontaktprotokolle „durch Nachlässigkeiten der Interviewer“ (S. 151) unvollständig⁴⁰. Zum korrekten Klassifizieren der Ausfallursachen sind Interviewer teilweise auch auf Angaben dritter Personen angewiesen; hier kann es bspw. durch *Gatekeeper* zu (bewusst oder unbewusst) falschen Klassifikatio-

⁴⁰Kennickell (1999, S. 287) beschreibt für den amerikanischen Survey of Consumer Finances SCF einen Interviewerfragebogen für eine Verbesserung der Datenlage bei Nonresponse und kommt zu dem Ergebnis: „The completion rate for the interviewer questionnaire was 100 percent, and missing information problems there are fairly small“.

nen („nicht teilnahmefähig“) kommen (siehe zum Gatekeeperproblem auch Abschnitt 1.4.3). Andererseits sind Auskunftspersonen in einigen Fällen die einzige Möglichkeit zur korrekten Klassifikation des Ausfallgrundes, bspw. bei Krankenhausaufenthalten (vgl. Schnell, 1997, S. 241): wird bei einem Kontaktversuch auch keine Auskunftsperson angetroffen, wird ein Krankenhausaufenthalt nicht als solcher erfasst, sondern als „nicht erreichbar“ klassifiziert. Besonders deutlich wird die Bedeutung einer korrekten Klassifikation des Ausfalls für Mortalitätsanalysen auf Basis von Panelstudien: werden Ausfälle durch Tod nicht korrekt erfasst, besteht die Gefahr, Mortalitätsraten zu unter- und Lebenserwartung zu überschätzen⁴¹. Betrachtet man die nach einzelnen Ausfallursachen aufgeschlüsselten Teilnahmesequenzen über alle Wellen hinweg, zeigen sich eventuell auch „de facto“ absorbierende Zustände, nämlich dann, wenn beispielsweise Kontaktpersonen über mehrere Wellen „weiche Verweigerungen“ aussprechen und in den folgenden Wellen immer wieder kontaktiert, aber nicht erfolgreich interviewt werden; eventuell zeigen sich auch Tendenzen der Interviewer, bestimmte Ausfallkategorien häufiger zu verwenden und bspw. Verweigerung über mehrere Wellen zu verschleppen⁴².

Sequenzen können als solche analysiert werden; Ziel einer Sequenzanalyse ist die Entdeckung von Mustern in den Sequenzen und der Reihenfolge der einzelnen Sequenzbestandteile: „There are two goals in sequential analysis. The first goal [...] is to discover stochastic (i.e., probabilistic) patterns in the data. The goal is equivalent to cracking a secret code“ (Gottman und Roy, 1990, S. 19). Zudem versucht eine Sequenzanalyse, die Sequenzstruktur durch Kovariaten zu erklären. Zentraler Ausgangspunkt ist die *Abfolge* der einzelnen Sequenzbestandteile

⁴¹Für das SOEP wurde zuletzt 2001 eine Verbleibstudie für alle Ausfälle zwischen 1985 und 1998 (ausser Ausfälle durch Tod und Wegzug ins Ausland) durchgeführt, um aktuelle Vitalinformation der Ausfälle zu erhalten (Infratest Sozialforschung, 2002b). Schnell und Trappmann (2006, ohne Seitenangaben) kommen auf Basis der aktualisierten Todesfälle im SOEP jedoch zu dem Ergebnis, dass sich „keine schwerwiegenden Auswirkungen auf die Schätzungen der Lebenserwartung“ und der Survivorfunktion durch die Berücksichtigung der Ergebnisse der Verbleibstudie ergeben.

⁴²Kennickell (1999, S. 301) äussert allerdings auch die Vermutung, „some respondents may make themselves hard to contact rather than have to deal with an interviewer, and such people may be more likely to have refused had they been contacted further“.

Tabelle 1.2: Teilnahmesequenzen am Beispiel eines Panels mit 7 Wellen

Nr.	Person	Welle							Sequenz	Sequenzlänge
		W_1	W_2	W_3	W_4	W_5	W_6	W_7		
1	I_{11}	4	4	4	4	4	4	4	4 4 4 4 4 4 4	7
2	I_{12}	4	4	4	4	4	4	4	4 4 4 4 4 4 4	7
3	I_{1n}	4	4	4	4	4	4	4	4 4 4 4 4 4 4	7
4	I_{21}	4	4	4	4	4	1	–	4 4 4 4 4 1	6
5	I_{22}	4	4	4	2	–	–	–	4 4 4 2	4
6	I_{23}	4	4	1	–	–	–	–	4 4 1	3
7	I_{2n}	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
8	I_{31}	4	0	4	4	1	–	–	4 0 4 4 1	5
9	I_{31}	4	4	0	4	4	4	4	4 4 0 4 4 4 4	7
10	I_{32}	4	3	1	4	0	3	4	4 3 1 4 0 3 4	7
11	I_{3n}	4	3	4	3	4	1	–	4 3 4 3 4 1	6

0: nicht erreichbar; 1: verweigert; 2: nicht teilnahmefähig; 3: sonstiger Ausfall; 4: Teilnahme

über die Zeit: „Anyone who has collected data over time and ignores time is missing an opportunity“ (Gottman und Roy, 1990, S. 3).

Vergleicht man Tabelle 1.2 mit Tabelle 1.1, zeigt sich zunächst, dass sich in Tabelle 1.2 durch Permutation der Zeilen immer monotone Ausfallmuster herstellen lassen. In Tabelle 1.2 zeigen sich zudem deutlich Tracking- und Weiterverfolgungsregeln: offensichtlich führen in diesem Beispiel explizite Verweigerungen (1) und Ausfälle aufgrund von Teilnahmeunfähigkeit (2) zu einem endgültigen Ausfall aus dem Panel, während Ausfälle durch Nichterreichbarkeit und durch sonstige Gründe in der nächsten Welle erneut kontaktiert werden; andere Tracking- und Weiterverfolgungsregeln können andere Muster erzeugen.

2 Theoretische Ansätze zur Erklärung von Teilnahmeverhalten

Auch wenn in der Literatur zu Nonresponse „lange Zeit ein nahezu vollkommen empiristisches Vorgehen“ (Schnell, 1997, S.133) zu beobachten war, existieren Ansätze, das Teilnahmeverhalten in einen theoretischen Rahmen einzubetten, auch wenn es, so Schnell (1997, S. 157), „nach mehr als 50 Jahren Methodenforschung zum Interview immer noch keine

einheitliche Theorie des Interviews gibt“⁴³. Problematisch ist dabei, dass viele Arbeiten lediglich uni- und bivariate Einflüsse auf Nonresponse untersuchen und zum Teil nicht zwischen den verschiedenen Ausfallursachen unterscheiden; zudem ist problematisch, dass sich viele Zusammenhänge auf demographische Variablen beziehen, die eigentlichen Mechanismen, die diesen Zusammenhängen zugrunde liegen aber nicht deutlich werden, dies ist bspw. deutlich bei Geschlecht, Alter oder Einkommen als Einflußvariable auf Nonresponse (vgl. Schnell, 1997, S. 198ff). Tatsächlich können die meisten der Einflußgrößen im Rahmen einer Rational-Choice-Theorie des Teilnahmeverhaltens erklärt werden⁴⁴. Daneben existieren aber auch andere Erklärungsansätze zur Erklärung eines Teilnahmeverhaltens.

2.1 Leverage-Saliience-Theorie

Ein solcher Ansatz ist die sogenannte *Leverage-Saliience-Theorie* (vgl. Groves et al., 2004, S. 176ff); danach werden verschiedene Aspekte der Befragung von Befragten unterschiedlich bewertet und gegeneinander abgewägt⁴⁵. Solche Aspekte umfassen bspw. die Dauer der Befragung (und dementsprechend die Opportunitätskosten der Befragung), ob das Thema als interessant oder als sensitiv empfunden wird, ob eine Belohnung oder Anerkennung in Form eines *Incentives* ausgehändigt wird oder ob der Auftraggeber der Befragung positiv oder negativ für den Befragten besetzt ist. Je nach *Saliience* eines Aspektes fällt der Befragte

⁴³Für das Teilnahmeverhalten im Längsschnitt treffen Goodman und Blum (1996, S. 649) eine ähnliche Schlußfolgerung: „We were working with no theory of attrition to guide. A theory of attrition, like any other theory, will take time to develop. A theory of attrition would encourage longitudinal researchers to design their studies in light of anticipated attrition and to collect data on particular variables known to affect subject attrition“.

⁴⁴Schnell (1997, S. 200f) bemerkt bspw., dass Unterschiede in der Teilnahme zwischen den Geschlechtern vor allem auf Unterschiede in der Erreichbarkeit zurückzuführen sind und bspw. in einem höheren Anteil berufstätiger Männer begründet ist; einen ähnlichen Hinweis auf steigende Nonresponseraten, verursacht durch steigende Erwerbstätigkeit von Frauen geben Groves und Couper (1998, S. 169).

⁴⁵Groves et al. (2004, S. 177) nutzen das Bild der Balkenwaage um die Leverage-Saliience-Theorie zu veranschaulichen: Abhängig davon, wo (positiv oder negativ) ein Aspekt (bspw. das Thema, der empfundene Nutzen, der Auftraggeber, usw.) auf der Waage verortet ist und welchen Stellenwert dieser Aspekt für den Befragten besitzt, schlägt die Waage zugunsten oder entgegen einer Teilnahme aus.

die Entscheidung zur Teilnahme oder verweigert eine Mitarbeit. Aufgabe des Interviewers ist nun, Aspekte, die für den Befragten positiv besetzt sind zu betonen und Sorgen und ängste negativ besetzter Aspekte abzubauen. Damit lässt sich *Tailoring* und *Maintaining Interaction* gut in den Rahmen einer Leverage-Saliency-Theorie einbetten. Den Einfluss des Erhebungsmodus auf Nonresponse durch eine unterschiedliche Bedeutung des Themas der Befragung für Befragte betonen Gallagher et al. (2005): während Befragte bei schriftlichen Befragungen nach Durchsicht des gesamten Fragebogens eine Teilnahmeentscheidung treffen können, ist das bei telefonischen Befragungen nicht der Fall; dadurch, so Gallagher et al. (S. 74), sei bei telefonischen Befragungen weniger als bei postalischen Befragungen mit systematischen themenbezogenen Ausfällen zu rechnen.

2.2 Oversurveying

Der theoretische Aspekt eines „*Oversurveying-Effekts*“ (Groves et al., 2004, S. 176) bezieht sich auf die Tatsache, dass die Anzahl von amtlichen, akademischen und kommerziellen Befragungen stark angestiegen ist und viele Personen mittlerweile Befragungserfahrungen gemacht haben⁴⁶; durch die damit verbundene „Teilnahmemüdigkeit“ werden sinkende Teilnahmeraten begründet, jedoch „ob sich eine Befragung auf die Wahrscheinlichkeit auswirkt, an späteren Befragungen teilzunehmen, ist empirisch bislang ungeklärt“ (Schnell, 1997, S. 172)⁴⁷. Auch Groves et al. (2004, S. 176) attestieren einem Oversurveying-Effekt „spotty support“. Allerdings bleibt durch die gestiegene Anzahl von Befragungen auch eine erhöhte Wahrscheinlichkeit, tatsächlich einmal negative Befragungserfahrungen zu machen. Empfohlene *Tailoring-Techniken* zur Betonung der Möglichkeit, an einer seltenen Gelegenheit teilzuhaben, seine Meinung zu äussern und seiner Stimme Gehör zu verschaffen („*Sie*

⁴⁶Schnell (1997, S. 171f) zeigt an einem Beispiel ausgehend von 1990, dass rund 80% der Bevölkerung über Befragungserfahrung verfügen müssten, allerdings werden ein Großteil dieser Erfahrungen offensichtlich vergessen oder nicht als Interviewerfahrung erinnert. Zudem würden die meisten Befragungen „von der überwiegenden Mehrheit der Befragten als angenehm empfunden“ (S. 173).

⁴⁷Bei Panelbefragungen scheint eine erfolgreiche Befragung in der vorhergehenden Welle die Teilnahmebereitschaft in der nächsten Welle zu erhöhen: „Of the entire group of new refusers, more people refused Interview I than any other interview in the cycle“ (DeMaio, 1980, S. 227).

wurden unter 10.000 Personen ausgewählt“) verlieren durch eine steigende Anzahl von Befragungen allerdings an Gewicht. Dieses „Knappheitsprinzip“ wird auch im Rahmen der Rational-Choice-Theorien (siehe Abschnitt 2.3) diskutiert. Lipps (2006, S. 3) berichtet für das vergleichsweise junge Schweizer Haushaltspanel SHP, dass sich dort ein negativer Effekt eines Oversurveying („prevalent in Switzerland“) in im Vergleich zu anderen Haushaltspanels deutlich niedrigeren Responseraten zeigt.

2.3 Rational-Choice-Theorien

Viele der theoretischen Aspekte zur Erklärung des Teilnahmebereitschaft lassen sich als Rational-Choice-Theorien rekonstruieren (die folgende Darstellung folgt Schnell (1997, S. 157ff) und Groves und Couper (1998, S. 121ff)): Letzten Endes entscheidet sich die Teilnahme vor dem Hintergrund der Abwägung von Kosten der Teilnahme und Nutzen aus einer Teilnahme. Das Grundprinzip der Rational-Choice Theorie besagt, dass diejenige (subjektiv vorstellbare) Handlungsalternative gewählt wird, die unter gegebenen Umständen am ehesten der Zielerreichung dient (Schnell, 1997, S. 158f): „Die Multiplikation der subjektiven Wahrscheinlichkeiten mit den Intensitäten der Ziele (deren erwartete Nützlichkeit) ergibt für jede Handlungsalternative einen Vektor subjektiver Zielrelevanzen“ (S. 159). Allerdings besteht der Alltag aus typischen Situationen, in dem stabile Ziele als auch stabile, „bewährte“ Strategien zu Erreichung der Ziele existieren (S. 160); dadurch müssen nicht bei jeder Handlungswahl alle Handlungsoptionen berücksichtigt werden, sondern die automatische Wahl einer bewährten Handlung ist effizient zur Zielerreichung; erst wenn eine nicht-typische (und für die Zielerreichung wichtige) Situation eintritt, müssen eventuell neue Handlungsalternativen und Handlungspräferenzen berücksichtigt werden; ist diese Situation für die Zielerreichung nicht von Belang, kommt es nach Schnell (1997, S. 161) zu „mental coin-flipping“, also Situationen, in denen Verhalten nicht vorhersehbar ist, weil kleinste Änderungen der Situation die Wahl der Handlungsalternative verändern. Im Unterschied zu gewohnheitsmäßigem Verhalten steht skriptgesteuertes Verhalten und Handlungswahl. Aufgrund bestimmter Auslöser der Situation wird ein wissensbasiertes Handlungsmuster abgerufen und in Verhalten umgesetzt. Eine solche Situation ist bspw. ein Interviewerbesuch an der Haustüre: aufgrund

bestimmter Schlüssel wird die Person an der Haustüre als Vertreter identifiziert und ein Handlungsskript für „unerwünschter Vertreterbesuch“ ausgeführt und das Anliegen der Person an der Haustüre abgelehnt. Tatsächlich spielt der erste Kontakt zwischen Interviewer und Kontakt- bzw. Befragungsperson in der Literatur eine große Rolle und wird im Rahmen der Minimierung von Nonresponse durch Verweigerungen diskutiert; das Verhalten der Interviewer im Rahmen des *Tailoring* zielt gerade darauf ab, entsprechende Schlüssel zu einer korrekten Skriptidentifizierung auf Seiten der Befragungspersonen bereitzustellen.

Opportunitätskosten Die Opportunitätskostenhypothese (vgl. bspw. Dillman et al., 2002, S. 8) beschreibt dieses Abwägen zwischen Kosten und Nutzen der Teilnahme: die Kosten der Befragung sind für diejenigen höher, die knappe Zeitressourcen und dadurch höhere Opportunitätskosten haben; knappe Zeitressourcen spielen für die Erreichbarkeit zu Hause als auch für Verweigerungen aufgrund mangelnder Zeitressourcen eine Rolle. Die Dauer der Befragung ist zentral für das Konzept der Opportunitätskosten: je länger die (erwartete) Dauer, desto höher sind prinzipiell die Opportunitätskosten⁴⁸. Aufgrund höherer Opportunitätskosten wird auch eine höhere Ausfallwahrscheinlichkeit durch Verweigerung für berufstätige Personen erwartet (Schnell, 1997, S. 204); Vor diesem Hintergrund spielt auch eine zunehmende Erwerbstätigkeit von Frauen eine Rolle (vgl. Groves und Couper, 1998, S. 169). Bei der Analyse von Nonresponse im amerikanischen Survey of Income and Program Participation (SIPP) kommen Galvin et al. (2000) zu dem auf den ersten Blick überraschenden Ergebnis, dass längere Interviews in den Vorwellen ein niedrigeres relatives Risiko besitzen, aus dem Panel herauszufallen⁴⁹. Vor dem Hintergrund der Opportunitätskostenhypothese ist dieses Ergebnis allerdings plausibel: hohe Opportunitätskosten und knappe Zeitressourcen der Befragungsperson führen bei einer Teilnahme dazu, möglichst wenig Zeit für die

⁴⁸Allerdings, so Schnell (1997, S. 168), spiele dies nur für die Kontaktphase eine Rolle, Abbrüche während der Befragung seien selten.

⁴⁹Zu dem selben Ergebnis kommen auch Hill und Willis (2001, S. 432) auf Basis der amerikanischen Health and Retirement Study (HRS): erst nach einer Dauer der Befragung von über zwei Stunden sinkt eine Teilnahmebereitschaft in der nächsten Welle.

Befragung aufzuwenden⁵⁰. Um Befragungspersonen mit hohen Opportunitätskosten zu einer Teilnahme zu bewegen, ist es evtl. effektiv, dieser Gruppe besondere, bspw. höhere monetäre Anreize⁵¹ zu setzen als anderen Befragten (Hill und Willis, 2001, S. 421).

Theorie des sozialen Tauschs Weniger als Entschädigung der Kosten⁵² als Anreiz im Rahmen einer Theorie des sozialen Tauschs (Dillman et al., 2002, S. 8) sind monetäre *Incentives* oder Sachgeschenke zu sehen, deren Wert im allgemeinen nicht so hoch ist, um dadurch tatsächlich Opportunitätskosten auszugleichen⁵³. In der Theorie des sozialen Tauschs spielt das Konzept der *Reziprozität* eine zentrale Rolle (vgl. bspw. Dillman et al., 2002, S. 8f; Groves und Couper, 1998, S. 33; oder Schnell, 1997, S. 162): Durch kleine Aufmerksamkeiten, die Betonung der positiven Auswirkungen und Ergebnisse der Befragung für die Befragten, Ausdruck der Wertschätzung einer Teilnahme oder den Appell an eine „Bürgerpflicht“ soll ein reziprokes Verhalten der Befragungspersonen ausgelöst werden (damit spricht das Reziprozitätsprinzip beide oben beschriebenen Mechanismen an: einmal das Entsprechen einer gesellschaftlichen Norm zur gewohnheitsmäßigen Erwidierung positiven Verhaltens, als auch das Auslösen eines individuellen Skripts zu reziprokem Verhalten bspw. als vertrauensbildende Maßnahme, dass dann vom Befragten mit einer Teilnahme erwidert werden soll).

⁵⁰Galvin et al. (2000) vermuten allerdings keine niedrigeren Opportunitätskosten als Ursachen für das niedrige Ausfallrisiko auch bei längerer Interviewdauer, sondern ein verstärktes Interesse am Thema und ein Gefühl der „Bürgerpflicht“ bei der Befragung (S. 638).

⁵¹Groves et al. (1999) untersuchen diese „differenzierten Anreize“ für Verweigerer: sie kommen auf Basis ihres Feldexperiments zu dem Ergebnis, dass eine Differenzierung der monetären Anreize weder von der dann monetär benachteiligten Gruppe als ungerecht empfunden wird, noch dass eine Teilnahme für diese Gruppe bei Kenntnis der Differenzierung der Anreize weniger wahrscheinlich wird (S. 261ff).

⁵²Hill und Willis (2001) formalisieren für eine Erklärung des Teilnahmeverhaltens die Theorie des sozialen Tauschs als ökonomisches Tauschmodell, in dem allein der erwartete Nutzen der Befragung selbst und der erwartete monetäre Nutzen der Teilnahme eine Teilnahmeentscheidung ausmacht.

⁵³Im SOEP erhalten Befragte keinen finanziellen Ausgleich; jede Befragungsperson erhält nach einem erfolgreichen Interview allerdings als Dankeschön ein Monatslos der „Aktion Mensch“ sowie beim Besuch des Interviewers ein kleines Sachgeschenk, bspw. eine Tasse.

Theorie der sozialen Isolation Mit einer Theorie des sozialen Tauschs sind auch Ansätze der sozialen Isolation verbunden (Dillman et al., 2002, S. 9): Sozial isolierte Personen sind demnach nicht durch dieselben gesellschaftlichen Normen geleitet wie sozial integrierte Personen (Groves und Couper, 1998, S. 33); dementsprechend sind die beschriebenen Mechanismen der Reziprozität nicht wirksam. Folglich verweigern sozial isolierte Personengruppen häufiger eine Teilnahme an Befragungen wie auch eine generelle soziale, politische und gesellschaftlichen Teilhabe (vgl. Groves und Couper, 1998, S. 131f): Waterton und Lievesley (1987, S. 274) zeigen bspw. anhand der Ausfälle des British Social Attitudes Survey 1983 bis 1986⁵⁴, dass Ausfälle aus dem Panel weniger politisch aktiv sind als beständige Panelteilnehmer⁵⁵. Auch Djerf (2004, S. 44f) bemerkt, dass eine politische Teilhabe stark zurückgegangen ist und verbindet dies mit sinkenden Teilnahmeraten an der finnischen Arbeitskräfteerhebung⁵⁶. Unterschiedliches Teilnahmeverhalten von ethnischen Gruppen wird ebenfalls im Rahmen der Isolationshypothese diskutiert (Groves und Couper, 1998, S. 132) als auch das Teilnahmeverhalten älterer Personen; Groves und Couper (1998, S. 133) berichten von einer verringerten Teilnahme älterer Personen, allerdings vermischen sich hier Effekte der Erreichbarkeit und der Kooperation: generell sind ältere Personen häufiger zu Hause anzutreffen als jüngere, allerdings gibt es unterschiedliche Effekte durch die verringerte Teilnahmefähigkeit (bspw. aufgrund von Krankheit) und verringerter Teilnahmewahrscheinlichkeit durch einen im Vergleich niedrigeren Bildungsstand (der die Wahrscheinlichkeit einer falschen Skriptidentifizierung erhöht); andererseits, so argumentiert Schnell (1997) besitzen ältere Personen ein

⁵⁴Der British Social Attitudes Survey besteht eigentlich aus mehreren unabhängigen Querschnitten, für 1983 bis 1986 wurde jedoch auch eine kleinere Panelstudie im Rahmen des Attitudes Survey durchgeführt (Waterton und Lievesley, 1987, S. 267f).

⁵⁵Zudem zeigen Waterton und Lievesley (1987) „the attitudes which do discriminate between those who drop out and those who continue are related to attitudes which discriminate between libertarians and conservatives“ (S. 274). Beständige Panelteilnehmer seien insbesondere toleranter in Bezug auf eine freie Meinungsäußerung und ihren Ansichten zu Sexualverhalten; dieser Effekt sei nur teilweise dadurch zu erklären, „that more of the older age groups - who hold more stringent moral views - are lost“.

⁵⁶„The correlation between the yearly response rate and the voting turnout is over 0.90 for both types [kommunale und landesweite, TG] of elections“ (Djerf, 2004, S. 45).

im Vergleich verkleinertes soziales Umfeld, sodass eine Teilnahme an einer Befragung durchaus als nutzensteigernd betrachtet werden kann. Im Rahmen der Tausch- und Isolations-
theorie diskutieren Groves und Couper (1998) auch widersprüchliche Einflüsse ökonomischer
Faktoren (S. 126ff). Der in der Literatur häufig beschriebene „Mittelstandsbias“ bezeichnet
die Unterschiede für unterschiedliche sozio-ökonomische Gruppen in ihrem Teilnahmeverhal-
ten. Von der Theorie des sozialen Tauschs wäre zu erwarten, dass Personen mit niedrigem
sozio-ökonomischen Status am meisten von einer Teilnahme an bspw. amtlichen Befragungen
profitieren und bspw. als Ausgleich für staatliche Leistungen an der Befragung teilnehmen;
entsprechend sehen Personen mit hohem Status wenig Nutzen in einer Teilnahme. Entspre-
chend sinkt die Teilnahmewahrscheinlichkeit mit steigendem sozio-ökonomischem Status.
Allerdings, ausgehend von einer Theorie der sozialen Isolation, kann argumentiert werden,
dass Personen mit niedrigem sozio-ökonomischen Status sich als durch den Staat benach-
teiligt ansehen, ebenso wie Personen mit hohem sozio-ökonomischen Status sich aufgrund
ihrer Beiträge bspw. in Form von Steuern und Abgaben an den Staat nicht zu weiteren Bei-
trägen in Form einer Befragung motiviert sehen. Dies würde einem umgekehrt u-förmigem
Zusammenhang zwischen sozio-ökonomischen Status und Teilnahmewahrscheinlichkeit be-
schreiben⁵⁷. Empirische Ergebnisse, die für oder gegen einen Mittelstandsbias sprechen sind
widersprüchlich (vgl. Groves und Couper, 1998, S. 127). Schnell (1997, S. 202ff) argumen-
tiert, dass, wenn es einen Effekt gibt, dieser sich zunächst einmal als Bildungseffekt darstellt
(S. 206): „Eine höhere Bildung erlaubt die (korrekte) Einschätzung der völligen individuellen
Konsequenzlosigkeit einer Survey-Teilnahme“, die Wahrscheinlichkeit zur Identifikation und

⁵⁷Einen im Gegensatz dazu u-förmigen Zusammenhang beschreibt bspw. DeMaio (1980, S. 227f) für Ver-
weigerungsraten nach Einkommen: Haushalte mit niedrigem und hohem Einkommen verweigerten weniger
häufig eine Teilnahme beim amerikanischen Current Population Survey CPS als Haushalte mit mittleren
Einkommen. Allerdings fehlt für 10% der kooperativen Haushalte eine Einkommensangabe und für 60% der
konvertierten Verweigerer, daher sind die Ergebnisse DeMaio (1980) mit Vorsicht zu betrachten; für die Re-
gisterstichprobe des amerikanischen SCF liegen aus dem Frame vollständige Einkommensinformation für eine
Reihe von Einkommensarten vor, Kennickell (1999, S. 290) findet für eine Reihe von Einkommensarten eine
niedrigere Teilnahmewahrscheinlichkeiten bei höheren Einkommen, Goyder et al. (2002) finden eine höhere
Teilnahmewahrscheinlichkeit bei hohem sozioökonomischen Status.

Abrufen des korrekten Skripts durch die Befragungsperson ist größer (vgl. auch Kennickell, 1999, S. 298).

Im Rahmen der Isolationstheorie werden auch höhere Teilnahmewahrscheinlichkeiten für jüngere Personen, Haushalte mit kleinen Kindern, sowie Haushalte in kleineren Wohngebäuden (bspw. Einfamilienhäuser) aufgrund einer besseren Integration in die Nachbarschaft oder höherer Kontaktfreudigkeit und -möglichkeit erwartet; überraschend sind Kontaktraten auch unter mobilen Haushalten höher (vgl. Groves und Couper, 1998, S. 138f); Erklärungen eines Effektes der Stadt-Land-Unterschiede (*Urbanicity*, „one of the most consistently documented ecological correlates of survey cooperation“; Groves und Couper, 1998, S. 176) basieren auf Argumenten höherer Anonymität und höherem Misstrauen, Fremden gegenüber, der soziale Zusammenhalt⁵⁸ in Ballungsgebieten sei geringer. Dementsprechend sei in ländlichen Gebieten die Isolation niedriger und soziale Kontrolle größer und mit ihr, die Bereitschaft der Teilnahme an sozialen Aktivitäten und Teilnahme an Aktivitäten zur Bereitstellung eines öffentlichen Gutes zum Nutzen einer sozialen Gemeinschaft (vgl. Groves und Couper, 1998, S. 176ff). Steeh (1981) bspw. betont v.a. Urbanisierungstendenzen und ihre Auswirkungen auf eine Teilnahmebereitschaft als Hauptgrund für gestiegene Nonresponseraten über die Zeit.

Im Rahmen der Rational-Choice-Theorien werden auch Erklärungen des Teilnahmeverhaltens durch das Prinzip des sekundären Vergleichs (Schnell, 1997, S. 162) diskutiert; danach werden Verhaltensweisen anderer Personen kopiert und als Orientierung für eigenes Verhalten verwendet (Groves et al., 1992, S. 482); dieses Prinzip findet bspw. beim *Tailoring* und Aufrechterhalten eines Kontakts durch den Interviewer Anwendung („Alle anderen angesprochenen Personen haben ebenfalls teilgenommen“)⁵⁹. Das Konsistenzprinzip andererseits beschreibt das Bestreben konsistenter Verhaltensweisen von Personen; eine schon einmal gezeigte Verhaltensweise soll unter gleichen Umständen fortgesetzt werden. Besonders bei

⁵⁸Groves und Couper (1998, S. 179) operationalisieren diese „soziale Kohäsion“ über den Anteil der Personen, die in Gruppen- oder Gemeinschaftsunterkünften wohnen und den Anteil selbstgenutzten Wohneigentums und Einfamilienhäuser.

⁵⁹Allerdings muss dabei die Privatsphäre gewahrt bleiben und ein Hinweis auf bestimmte Personen oder Personengruppen („Ihr Nachbar hat ebenfalls schon teilgenommen“) vermieden werden.

Panelbefragungen besteht darin eine Möglichkeit, weiterbestehendes Teilnahmeverhalten zu erklären; andererseits kann das Konsistenzprinzip auch zur Erklärung „habituellem Verweigerer“ herangezogen werden: Verweigerungen finden demnach aufgrund der (auch fälschlicherweise) identifizierten Interviewsituation statt und nicht aufgrund stabiler Verhaltensdispositionen. Eine weitere *Tailoring-Technik* wendet das Knappheitsprinzip an (Groves und Couper, 1998, S. 33), um eine Teilnahme zu erreichen. Die Betonung, dass die Teilnahme eine seltene Chance darstellt, seine Meinung zu äussern oder einen wichtigen Beitrag zu einem bestimmten wichtigen Thema zu leisten, soll die Befragung als knappes positives Gut (Schnell, 1997, S. 162) darstellen⁶⁰. Eng verbunden mit der Theorie des sozialen Tauschs (und dabei vor allem dem Appell an die „Bürgerpflicht“) ist das Autoritätsprinzip, das eine höhere Teilnahmebereitschaft vor allem für staatliche oder wissenschaftliche Auftraggeber postuliert, da ihnen eine Berechtigung zur Durchführung von Befragungen am ehesten zugeschrieben wird (vgl. Groves und Couper, 1998, S. 33): eine klare Kennzeichnung bspw. durch ein Logo und explizite Betonung des Auftraggebers bei der Vorstellung durch den Interviewer soll eine Teilnahmebereitschaft erhöhen: „In Hinsicht auf akademische Surveys kann die Hoffnung bestehen, dass allein die Kennzeichnung als „wissenschaftliche Erhebung“ ein entsprechendes Skript bei den Kontaktpersonen auslöst“ (Schnell, 1997, S. 170). Entgegenesetzt zum Appell an eine „Bürgerpflicht“ im Rahmen einer Theorie des sozialen Tauschs bestehen allerdings auch Einwände, dass eine besondere Betonung eines amtlichen Surveys auch zu einer niedrigeren Teilnahmebereitschaft führen könnte, bspw. deutlich in äusserungen wie „Für den Staat habe ich schon genug getan“, „Das geht den Staat nichts an“ oder „Der Staat weiss schon genug über seine Bürger“. Wird ein Auftraggeber als nutzensteigernd für Befragte identifiziert ist generell eine Teilnahmewahrscheinlichkeit höher (vgl. Dillman et al., 2002, S. 11).

Die Belastung durch eine Befragung und vor allem die empfundene Belastung ist neben der Dauer auch abhängig vom Thema der Befragung; ist das Thema und Zweck der Befragung von Bedeutung für Befragte, schreiben sie einer Teilnahme einen hohen empfundenen Nutzen

⁶⁰Wenn es tatsächlich einen „Oversurveying-Effekt“ gibt (siehe Abschnitt 2.2), dann verliert das Knappheitsprinzip zunehmend seine Grundlage.

zu; durch einen hohen empfundenen Nutzen steigt auch die Teilnahmebereitschaft (Dillman et al., 2002, S. 9; vgl. dazu auch Abschnitt 2.1). Entsprechend können sensitive Themen einen empfundenen Nutzen verringern⁶¹, als auch zunehmende Sorgen um eine Verletzung der Privatsphäre durch eine Teilnahme⁶², auch wenn Schnell (1997, S. 167) bemerkt, dass Befragte trotz bekundetem Misstrauen gegenüber der Verwendung der erhobenen Daten tatsächlich keine Gefährdung ihrer Ziele sehen; DeMaio (1980, S. 230) berichtet demgegenüber, dass zwischen 17% und 20% der Verweigerer im amerikanischen Current Population Survey CPS die Verletzung ihrer Privatsphäre als Grund für die Verweigerung angeben.

3 Verfahren zur Behandlung von Nonresponse

Ahern und Le Brocque (2005, S. 57) geben drei Schritte vor, anhand derer entschieden werden muss, wie mit Nonresponse in Panelerhebungen umgegangen werden muss: „First [...] determine the extent of loss of follow-up in the data for the variables under question. Second,

⁶¹Bspw. gelten Fragen nach Einkommen und Vermögen als „sensitiv“ (vgl. Groves und Couper, 1998, S. 36). In der 4. Welle des SOEP 1988 wird ein höherer Anteil der Verweigerer u.a. auch dem wellenspezifischen Befragungsschwerpunkt zu Einkommen und individuellem Vermögen zugeschrieben; Razafindratsima und Kishimba (2004, S. 368) vermuten andererseits nur einen kleinen Teil der Ausfälle in einem Panel zum Verhütungsmittelgebrauch von Frauen in sensitiven Fragen zum Sexualverhalten begründet.

⁶²Als Beispiele für eine zunehmende Bedeutung des Schutzes der Privatsphäre führen Groves und Couper (1998, S. 173f) das „Project Metropolit“ in Schweden an (siehe dazu auch Lyberg und Lyberg, 1991, S. 82), und die Diskussion um die deutsche Volkszählung 1981, die 1983 schließlich in das „Volkszählungsurteil“ des Bundesverfassungsgerichts mündete. Nachdem in Schweden publik wurde, dass 15.000 Personen einer Geburtskohorte im Raum Stockholm mehrere Jahre ohne ihr Wissen und Einverständnis untersucht wurden, verdoppelten sich die Responseraten bspw. der schwedischen Arbeitskräfteerhebung im Anschluss an die öffentliche Diskussion nahezu. Interessanterweise geht diese Steigerung offensichtlich nur zu einem kleinen Teil auf Verweigerungen zurück, (vgl. Abbildungen 6.15 und 6.16 bei Groves und Couper (1998, S. 174f)). Ein weiteres Beispiel beschreibt Djerf (2004, S. 22) für Finnland, als 1988 eine Studie bekannt wurde, die (legal) das Wahlregister mit anderen Bevölkerungsregistern verband; in den Medien wurde diskutiert, ob eine Verletzung und Gefährdung der Privatsphäre stattgefunden hat: „However, after publication it was found that the study did not involve any such threat and the discussion about 'big brother watching you' ceased rapidly“, und auch die Nonresponseraten fielen wieder auf ihr früheres Niveau.

if attrition is evident, [...] determine if attrition biases the results. Third, if bias is evident [...] adjust for attrition“. Dabei betonen sie, dass diese Schritte für jede Forschungsfrage durchgeführt werden müssen: „there is evidence to suggest, that attrition bias is model specific and must be explored within each sample, population, or study. Lack of evidence of bias in one model does not imply that there will be no bias in any other model“ (S. 59-60). Alle Techniken zur Behandlung von Nonresponse stellen eine Form der Schadensbegrenzung durch Nonresponse dar: „All such methods should be regarded as a form of statistical damage control, or as statistical 'fire-fighting' equipment. [...] Other professions adhere to a similar creed. Avoiding emergency surgery by encouraging safer everyday conduct and healthier lifestyle is an integral part of a good health care system“ (Longford, 2005, S. 14).

Alle diese „Notoperationen“ erfordern eine Annahme über den Ausfallmechanismus (siehe Abschnitt 1.3).

Als „Daumenregel“ stellt Kalton (1986, S. 304) fest, dass Unit-Nonresponse in der Regel durch Gewichtungsverfahren, Item-Nonresponse durch Imputationsverfahren behandelt wird; wie in Abschnitt 1.5 beschrieben, ist im Längsschnitt diese Unterscheidung zwischen Item- und Uni-Nonresponse nicht immer deutlich; allerdings sind auch Gewichtung und Imputation eng miteinander verbunden (Kalton, 1986, S. 304). Der größte Unterschied besteht darin, dass für Unit-Nonrespondenten in der Regel nur wenig Information (bspw. aus dem Sampling-Frame) vorliegt, wohingegen für Beobachtungen mit Item-Nonresponse mehrere Variablen der eigentlichen Befragung auch tatsächlich erhoben werden konnten. Für Gewichtungsverfahren können die wenigen spärlichen Informationen, die für Respondenten als auch Nonrespondenten vorliegen genutzt werden, um Hochrechnungszellen zu bilden, da diese Informationen aber kaum dazu ausreichen, um alle fehlenden Werte der Nonrespondenten vorherzusagen, erhalten Respondenten derselben Hochrechnungszelle ein höheres Gewicht, um die Nonrespondenten auszugleichen (vgl. Kalton, 1986, S. 304f; Kalton und Flores-Cervantes, 2003, S. 84). Ein Nachteil der Gewichtung bei Item-Nonresponse lässt sich aus der unterschiedlichen Datenlage bei Item- und Unit-Nonresponse erklären: bei Gewichtung erhalten Nonrespondenten ein Gewicht von 0 (und Respondenten der gleichen Hochrechnungszelle entsprechend ein höheres Gewicht), wird bei Item-Nonresponse nun gewichtet, anstatt fehlende

Werte zu imputieren, geht vergleichsweise viel Information der Variablen ohne fehlende Werte verloren.

3.1 Gewichtungsverfahren

Häufig werden Ausfälle durch Nonresponse durch Gewichtungsverfahren korrigiert; vor allem in der ersten Welle liegen vergleichsweise wenig Informationen für Nonrespondenten vor; dabei wird die Stichprobe auf Basis der durch das Design vorgegebenen Auswahlwahrscheinlichkeiten an Randverteilungen einzelner Merkmale oder Merkmalskombinationen mehrerer Variablen⁶³ der Grundgesamtheit⁶⁴ angepasst. Ausgehend von einer geschichteten Zufallsstichprobe mit j Schichten ist die Wahrscheinlichkeit einer Beobachtung in die Stichprobe zu gelangen entsprechend dem Quotient der Elemente der Grundgesamtheit in Schicht j und der Anzahl der Stichprobenelemente in Schicht j

$$\pi_{ij} = n_j/N_j \quad (3.1)$$

Indikator I gibt nun an, ob eine ausgewählte Beobachtung in die Stichprobe gelangt

$$I_i = \begin{cases} 1, & \text{Beobachtung } i \text{ ist in der Stichprobe,} \\ 0, & \text{Beobachtung } i \text{ ist nicht in der Stichprobe,} \end{cases} \quad (3.2)$$

und Indikator R , der „Response-Indikator“, gibt an, ob eine Beobachtung, gegeben $I = 1$, in der Stichprobe beobachtet wurde, oder durch Nonresponse ausfällt

$$R_i = \begin{cases} 1, & \text{Beobachtung } i \text{ wird in der Stichprobe beobachtet} \\ 0, & \text{Beobachtung } i \text{ fällt durch Nonresponse aus.} \end{cases} \quad (3.3)$$

Unter der Annahme, Ausfälle stellen eine Zufallsstichprobe aus R dar (*Missing Completely at Random*) ist R unabhängig von beobachteten Variablen Y und davon, in die Stichprobe

⁶³„Postulating a uniform global response mechanism makes the mathematics trivial but generally is not very realistic“ (Oh und Scheuren, 1983, S. 148).

⁶⁴Oh und Scheuren (1983, S. 149) unterscheiden diese Anpassung an die Grundgesamtheit (*Poststratifikation*) von einer Anpassung, wenn die Randverteilung der Grundgesamtheit nicht bekannt ist und an die Randverteilung des ursprünglichen Samples ohne Nonresponse angepasst wird (*Weighting-class-approach*).

zu gelangen

$$R \text{ II } (I, Y), \quad (3.4)$$

und ein Nonresponse-Gewicht ergibt sich aus dem Quotient der Elemente N_j der Grundgesamtheit in Schicht j und der Anzahl der Stichprobenelemente n_j in Schicht j

$$\phi = n_j/N_j, \quad (3.5)$$

geschätzt durch den Quotient der Anzahl n_j der gezogenen Elemente in Schicht j und der Anzahl m_j der teilnehmenden Elemente in Schicht j

$$\hat{\phi}_j = m_j/n_j. \quad (3.6)$$

Dadurch ergibt sich der Horvitz-Thompson-Schätzer als Inverse des Produkts der beiden Auswahlwahrscheinlichkeiten $1/(\pi_{ij} * \hat{\phi}_{ij})$

Allerdings erfordert ein Gewichtungsverfahren nun nicht die strenge Annahme, dass Ausfälle *MCAR* sind, allerdings basieren sie auf der Annahme, dass Ausfälle innerhalb bestimmter beobachtbarer Gruppen C zufällig (*MAR*) sind (vgl. Little und Rubin, 1987, S. 54)

$$R \text{ II } (I, Y)|C. \quad (3.7)$$

Diese Gruppen werden zur Anpassung der Stichprobe an die Grundgesamtheit verwendet und Ausfälle durch Nonresponse korrigiert. Generell gilt, sind die durch die GewichtungsvARIABLEN C gebildeten Klassen in Bezug auf ihre Ausfallwahrscheinlichkeit homogen, ist ein möglicher Nonresponse-Bias umso kleiner, je mehr Klassen zur Gewichtung gebildet werden (Oh und Scheuren, 1983, S. 155). Häufige Gewichtungsmerkmale sind demographische Variablen wie Geschlecht, Alter, Familienstand, ethnische Zugehörigkeit, Bildung und Einkommen, Haushaltsgröße, und Gemeindegößenklassen; ihre Auswahl richtet sich meist mehr nach der praktischen Verfügbarkeit als ihrer Fähigkeit, einen Ausfall gut vorherzusagen. Um Ausfälle durch Gewichtung korregieren zu können, müssen die für die Gewichtung verwendeten Variablen C tatsächlich den Ausfallmechanismus abbilden, d.h. in den durch die Kombination der Merkmalsausprägungen entstehenden Klassen dürfen sich Teilnehmer und Nichtteilnehmer nicht unterscheiden (Kalton und Flores-Cervantes, 2003, S. 85); wie oben schon beschrieben

sind diese Klassen in Hinsicht auf die Ausfallwahrscheinlichkeiten häufig nicht homogen⁶⁵; im Ergebnis stimmen damit Randverteilungen zwar mit der Grundgesamtheit überein, das Gewichtungsverfahren ist aber auch bei vielen Gewichtungsvariablen nicht in der Lage, einen Nonresponsebias für die interessierenden Variablen zu korregieren: „in general, it is not possible to guarantee a bias reduction by employing great many poststrata. The reason for this is that, if the uniform response mechanism postulated within the strata do not hold, then there will be a residual bias within each group“ (Oh und Scheuren, 1983, S. 155-156).

Annahme bei diesen Gewichtungsverfahren ist ein zufälliger Ausfallmechanismus (*Missing Completely At Random, MCAR* oder *Missing at Random, MAR*), so dass die Teilnehmer als Zufallsstichprobe aus dem ursprünglichen Sample ohne Nonresponse⁶⁶ angesehen werden (Oh und Scheuren, 1983, S. 147): „this method assumes that, conditional on covariates known for both, the respondents and nonrespondents, nonresponse is ignorable. Once we have adjusted for nonresponse based on the covariates used to generate nonresponse weights, the respondents are then assumed to be a random sample of respondents plus nonrespondents“ (McGuigan et al., 1995, S. 403). Schätzung der Grundgesamtheitsparameter sind unverzerrt, aber weniger effizient als auf Basis des vollständigen Samples: Gewichtung erhöht die Varianz und damit auch die Standardfehler der Schätzer.

Gewichtungsverfahren haben gegenüber Imputationsverfahren den Vorteil, dass sie die Korrelationsstruktur zwischen den Variablen bewahren (vgl. Kalton, 1986, S. 305).

3.2 Ausfallmechanismus als Omitted Variable

Heckman (1979) beschreibt eine Lösung eines Nonresponse-Bias als fehlerhafte Spezifikation eines statistischen Modells (vgl. Ahern und Le Brocq, 2005, S. 61); ausgehend von einem

⁶⁵Um diese, in Hinsicht auf die Ausfallwahrscheinlichkeiten, homogenen Gruppierung zu erreichen schlagen Little und Rubin (1987, S. 57) vor, Propensity-Gewichte (siehe Abschnitt 3.3) für alle Beobachtungen zu schätzen und eine Gruppierung aufgrund der Propensity-Gewichte vorzunehmen.

⁶⁶Beim *Weighting-class-approach* stellen die realisierten Beobachtungen entsprechend eine Zufallsstichprobe aus den durch die Gewichtungsvariablen vorgegebene Klassen dar: Ausfälle sind *Missing at Random, MAR*.

Regressionsmodell für das vollständige Sample ohne Ausfälle

$$\begin{aligned}
 Y_{1i} &= X_{1i}\beta_1 + U_{1i} & i = 1, \dots, n & \quad \text{bzw.} \\
 E(Y_{1i}|X_{1i}) &= X_{1i}\beta_1 & & \quad (3.8)
 \end{aligned}$$

ergibt sich für das durch Ausfälle reduzierte Sample

$$E(Y_{1i}|X_{1i}, \text{Auswahlregel}) = X_{1i}\beta_1 + E(U_{1i}|\text{Auswahlregel}). \quad (3.9)$$

Ist $E(U_{1i}|\text{Auswahlregel}) = 0$ ergibt sich kein Bias durch die Verwendung des realisierten, reduzierten Samples anstelle des vollständigen Samples, lediglich die Effizienz verringert sich aufgrund verringerter Fallzahlen (Heckman, 1979, S. 155); ist $E(U_{1i}|\text{Auswahlregel}) \neq 0$, liegt ein *Omitted Variable Bias* vor und das Modell ist fehlspezifiziert. Ob eine Beobachtung Y_{1i} vorliegt, ist bspw. abhängig von einem weiteren $Y_{2i} \leq 0$. Davon ausgehend schätzt⁶⁷ Heckman (1979) mittels Probit-Regression für jede Beobachtung i eine Auswahlwahrscheinlichkeit Λ_i in Abhängigkeit von $Y_{2i} = X_{2i}\beta_2 + U_{2i}$ und nimmt sie als zusätzlichen Regressor in die Regressionsgleichung mit auf. Dadurch verschwindet der *Omitted Variable Bias* und der Schätzer ist unverzerrt (vgl. Ahern und Le Brocq, 2005, S. 61). Annahme bei Heckmans *Sample-Selection-Modell* ist, dass Ausfälle nicht vollständig zufällig sind, sondern abhängig von beobachteten Variablen (*Missing on observables*); diese Variablen bestimmen die Auswahlregel, auf deren Basis die Teilnahmewahrscheinlichkeit Λ geschätzt und zur Vermeidung des *Omitted Variable Bias* als Regressor in ein Modell zur Schätzung der Grundgesamtheitsparameter aufgenommen werden kann. Heckmans Modell bezieht also die Propensity-Gewichte (siehe folgender Abschnitt) direkt in die Parameterschätzung mit ein; fehlen diese Propensity-Gewichte, ist das Modell fehlspezifiziert.

Ein Problem bereitet das Sample-Selection-Modell Heckman dann, wenn Prädiktoren für Λ_i mit den Kovariaten des Regressionsmodells 3.8 hoch korrelieren. Multikollinearität liegt bspw. dann vor, wenn für beide Schritte im Heckman-Modell die gleichen Prädiktoren verwendet werden (McGuigan et al., 1995, S. 403): „ideally, the variables used to predict attrition are not a subset of the variables used to estimate the outcome variable“.

⁶⁷Das Problem des Nonresponse ist hier, dass für $Y_{2i} \leq 0$ kein Y_{1i} vorliegt, allerdings ist für $Y_{2i} \leq 0$ X_{2i} bekannt, beispielsweise aus früheren Wellen.

3.3 Gewichtung auf Basis der Ausfallwahrscheinlichkeiten: Propensity-Gewichtung

Grundlegende Idee der Propensity-Verfahren ist die direkte Schätzung der Ausfallwahrscheinlichkeiten und eine - ähnlich der Designgewichtung - Gewichtung mit den Inversen der geschätzten Ausfallwahrscheinlichkeiten. Ausfallwahrscheinlichkeiten werden bspw. durch logistische Regression der „Responseindikatoren“ R auf die im Sampling-Frame verfügbaren Design-Variablen (ähnlich der zur Bildung der Hochrechnungszellen verwendeten Variablen) berechnet (vgl. bspw. Kalton, 1986, S. 307; Kalton und Flores-Cervantes, 2003, S. 89). ähnlich zum Problem bei einfachen Gewichtungsverfahren muss das zur Schätzung der Ausfallwahrscheinlichkeit verwendete (logistische) Regressionsmodell eine gute Vorhersagekraft der Ausfallwahrscheinlichkeit haben. Zudem ist eine Unterscheidung der verschiedenen Ausfallgründe (siehe Abschnitt 1.4) von zentraler Bedeutung, da sie unterschiedliche Kovariaten für das Regressionsmodell benötigen. Ausfallwahrscheinlichkeiten werden dann analog der Designgewichte als Inverse der Auswahl- bzw. Ausfallwahrscheinlichkeit verwendet (McGuigan et al., 1995, S. 403). Annahme ist wie bei den Gewichtungsverfahren ein zufälliger Ausfallmechanismus, hier jedoch abhängig von vollständig beobachteten Kovariaten X (vgl. Little und Rubin, 1987, S. 57)

$$R \parallel (I, Y) | X. \quad (3.10)$$

Mittels logistischer oder Probitregression wird nun die „Teilnahmeneigung“ (Response Propensity) $\hat{p}(x_i)$ für alle Beobachtungen i gegeben die Kovariaten X geschätzt

$$\hat{p}(x_i) = Pr(R_i = 1 | x_i) \quad (3.11)$$

Diese Propensity-Gewichte können nun analog der Ausfallgewichtung für eine Korrektur der Ausfälle durch Nonresponse verwendet werden: $w_i = 1/(\pi_i * \hat{p}(x_i))$

3.4 Imputation fehlender Werte

Imputation bezeichnet allgemein das Ersetzen fehlender Werte durch Einsetzen gültiger Werte. Der Vorteil von Imputationsverfahren liegt unter anderem darin, dass alle übrigen Infor-

mationen der Beobachtung erhalten bleiben und sich, im Gegensatz zu Gewichtungsverfahren, die Fallzahl der ungewichteten Stichprobe nicht verringert: eine Analyse basiert auf allen Beobachtungen der Stichprobe. Im Falle von über die Zeit vergleichsweise stabilen Merkmalen können Merkmale der letzten Welle W_{t-1} als Imputationen in der folgenden Welle W_t direkt übernommen werden („Last Observation Carried Forward, LOCF“). Als „Spender“ für einen fehlenden Wert einer Beobachtung i_1 bei einer „Hot-Deck-Imputation“ kann eine Beobachtung i_2 dienen, die in der vorherigen Welle gleiche oder ähnliche Ausprägungen auf einer oder mehreren Variablen hat⁶⁸; dazu werden aus Variablen der Welle W_{t-1} Imputationszellen ähnlich den Hochrechnungszellen gebildet; als Spender für eine Beobachtung i_1 wird dann eine Beobachtung i_2 aus den gleichen Imputationszellen verwendet⁶⁹.

Sehr einfache Imputationsverfahren setzen für fehlende Werte y_i^m einen unbedingten Mittelwert \bar{y}_j^o aller j beobachteten, nicht-fehlenden Werte ein⁷⁰, oder bedingte Mittelwerte nach festgelegten Imputationsschemata vergleichbar den Hochrechnungszellen⁷¹. Die Panelstruktur hat dabei gegenüber der Querschnittsstruktur den Vorteil, dass nicht nur unbedingte oder bedingte Mittelwerte innerhalb der gleichen Welle („Column-Imputation“) gebildet werden können, sondern auch Mittelwerte über verschiedene Wellen hinweg („Column-and-Row-Imputation“), bspw. bei nichtmonotonen Ausfallmuster durch temporäre Ausfällen. Mit der Imputation durch bedingte Mittelwerte eng verbunden ist der Regressionsansatz; hierbei werden bedingte Mittelwerte bspw. durch Regressionsmodelle auf Basis der vollständigen Beobachtungen in Welle W_{t-1} vorhergesagt und als Imputation für Beobachtungen mit feh-

⁶⁸ „Hot Deck within Adjustment Cells“ und „Nearest Neighbour Hot Deck“ (Little und Rubin, 1987, S. 65f).

⁶⁹ Im Falle von Unit-Nonresponse entspricht diese Hot-Deck-Imputation einer Nonresponse-Gewichtung: für Unit-Nonresponse müssen alle Merkmale von einem Spender übernommen werden, der Gewichtungsfaktor (bspw. ein Designgewicht) w_i des Spenders erhöht sich dadurch um den Faktor 2 (allgemein: $w_i^* = w_i + w_i * \text{Anzahl der Empfänger}$).

⁷⁰ Eingesetzt werden Werte aus der Mitte der beobachteten Verteilung. Dadurch wird die Varianz und Kovarianz unterschätzt (vgl. Little und Rubin, 1987, S. 44).

⁷¹ Eine Berechnung der Mittelwerte ist nur für kontinuierliche Merkmale sinnvoll, bei Merkmalen mit diskreten, kategorialen Ausprägungen kann anstelle des Mittelwerts bspw. der Modus zur Imputation verwendet werden (Kalton, 1986, S. 310).

lenden Werten in Welle W_t verwendet⁷². Prädiktoren im Regressionsmodell sind Variablen des Sampling-Frames (für Nonresponse in der ersten Welle) oder Variablen aus der letzten Welle, in der Variablen für Nonrespondenten und Respondenten vorliegen. Voraussetzung für eine Korrektur von Nonresponse ist eine gute Vorhersagekraft des Regressionsmodells, allerdings „a high correlation for the respondents does not guarantee that the nonrespondents’ values will be predicted well“ (Kalton, 1986, S. 309), gleiches gilt analog für die Imputation fehlender Beobachtungen auf Basis der Informationen vorheriger Wellen. Annahme dieser Imputationsverfahren ist, dass Ausfälle nur von den im Regressionsmodell bzw. zur Bildung der Imputationszellen verwendeten Variablen abhängen (*Missing at Random, MAR*). Ein Problem bei diesen Imputationsverfahren besteht darin, dass sie fehlende Werte mit Werten aus der Mitte der (bedingten oder unbedingten) Verteilung der Beobachtungen ohne fehlende Werte ersetzen; dadurch wird die Varianz (und dadurch Standardfehler und Signifikanzniveaus) unterschätzt. eine Möglichkeit besteht zumindest beim Regressionsansatz darin, auf den zur Imputation für Beobachtung i geschätzten Wert \hat{y}_i ein Residuum e_j einer Beobachtung j zu addieren (Taris, 2000, S. 33); dadurch erhalten nicht alle Beobachtungen mit imputierten Werten *einen* bedingten Mittelwert \hat{y} , sondern der bedingte Mittelwert erhält selbst eine Varianz.

Ein Nachteil der Korrektur von Nonresponse durch Imputation besteht darin, dass die imputierten Werte nicht die Unsicherheit des Imputationsprozesses widerspiegeln: „They will thus attribute greater precision to the survey estimates than is justified“ (Kalton, 1986, S. 305); Imputationen verursachen unter Umständen Artefakte (dann, wenn die zur Imputation verwendeten Variablen kaum oder nur schlecht die zu imputierenden Werte vorher-sagen können), daher müssen imputierte Werte als solche im Datensatz zu erkennen sein (und Ergebnisse auf ihre Sensitivität gegenüber Imputationen überprüft werden). Zudem bewahren Imputationsverfahren im Gegensatz zu Gewichtungsverfahren lediglich die Korrelationsstruktur der zur Imputation verwendeten Variablen, aber nicht notwendigerweise

⁷²Imputierte Werte liegen dann auf der Regressionsgeraden, die sich durch die Regression von Y^o auf die bedingenden Variablen X_k (Kovariaten) ergibt; dadurch werden zwar auch Varianz und Kovarianz unterschätzt, allerdings nicht so stark wie bei einer Imputation mit unbedingten Mittelwerten; die Unterschätzung ist umso kleiner, je besser Y durch X_k vorhergesagt werden kann (Little und Rubin, 1987, S. 46).

zu allen anderen Variablen der Beobachtungen. Schafer (1997, S. 2) fasst die Probleme der Imputationsverfahren folgendermaßen zusammen: „Ad hoc methods of imputation are [...] problematic. Imputing averages on a variable-by-variable basis preserves the observed sample means, but distorts the covariance structure, biasing estimated variances and covariances towards zero. Imputing predicted values from regression models, on the other hand, tends to inflate correlations, biasing them away from zero“.

Als Lösung des Problems der fehlenden Unsicherheit bei der Imputation fehlender Werte schlägt Rubin (1987, S. 15ff) multiple Imputationen vor. Die Idee dabei ist, für einen fehlenden Wert nicht *einen* gültigen Wert einzusetzen, sondern *mehrere* Imputationen eines fehlenden Wertes vorzunehmen; Multiple Imputationen haben gegenüber einfachen Imputationen mehrere Vorteile, vor allem besteht bei multiplen Imputationen die Möglichkeit, Unsicherheit bei der Imputation der fehlenden der Imputationen zu berücksichtigen; das ist einmal die Unsicherheit bei der Imputation eines fehlenden Wertes aufgrund der Stichprobenvarianz; die Stichprobenvarianz kann dadurch berücksichtigt werden, dass nicht nur ein Wert, sondern mehrere Werte bei jedem Imputationsschema verwendet werden. Zweiter Bestandteil der Unsicherheit betrifft die Unsicherheit des der Imputation zugrunde gelegten Imputationsschemas, nach dem fehlende Werte eingesetzt werden; sie wird dadurch ausgedrückt, dass für fehlende Werte nicht nur ein Imputationsschema angewandt wird, sondern mehrere; dadurch kann zudem die Sensitivität der Ergebnisse aufgrund der verschiedenen imputierten Werte überprüft werden⁷³. Für multiple Imputationen können grundsätzlich die gleichen Techniken wie für einfache Imputationen verwendet werden: „often the multiple-imputation version of an existing single-imputation scheme is not difficult to implement“ (Rubin, 1987, S. 18).

⁷³Daneben beschreibt Rubin (1987, S. 17f) auch den Nachteil, den multiple Imputationsverfahren haben: mehr Aufwand und Arbeit zur Imputation und mehr Speicherplatz für die imputierten Datensätze. Schließlich erhöht sich auch der Aufwand, da jeder der imputierten Datensätze analysiert werden muss; allerdings bleiben auch hier die Vorteile gegenüber einfachen Imputationen bestehen: „These advantages are not serious when m [die Anzahl der multiplen Imputationen, TG] is modest [...] When fractions of missing information are large, modest- m multiple imputation is not fully satisfactory, but then single imputation can be disastrous“ (S. 18).

Ein besonderes Problem für Imputationen stellen nichtmonotone Ausfallmuster (siehe Abschnitt 1.5.3 ab Seite 38) dar; Marini et al. (1980) stellen jedoch einen Ansatz vor, der eine Maximum-Likelihood-Schätzung der Parameter bei nichtmonotonen Ausfallmustern ohne Imputationen allein durch Manipulation der Varianz-Kovarianzmatrix ermöglicht. Schaffer (1997) schlägt Imputationen auf Basis des EM-Algorithmus und Monte-Carlo-Markov-Simulationen vor.

Insgesamt bleibt für alle Verfahren zur Korrektur von Nonresponse die Kritik: „With messy data, the confidence should be reduced as compared to the hypothetical data, that would have contained no mess. Such a reduction in the assessed confidence should be indicated in the conclusions of the analysis“ (Longford, 2005, S. 15).

4 Datenbasis: Das Sozio-Oekonomische Panel (SOEP)

Das Sozio-Oekonomische Panel (kurz SOEP) ist eine der größten nichtamtlichen wissenschaftlichen Wiederholungsbefragungen in Deutschland. Hervorgegangen ist das SOEP aus dem Sonderforschungsbereich 3 „Mikroanalytische Grundlagen der Gesellschaftspolitik“ an den Universitäten Frankfurt/Main, Mannheim und Berlin und ist am Deutschen Institut für Wirtschaftsforschung (DIW) beheimatet. Die Feldarbeit führte ununterbrochen Infratest Sozialforschung durch. Das SOEP stellt Paneldaten vor allem für zwei theoretische Bereiche zur Verfügung: das sind zum einen klassische ökonomische Variablen im Rahmen der Humankapitaltheorie sowie Arbeitsmarktsegmentation, zum anderen sind soziologische Variablen „vom Konzept der Sozialberichterstattung“ und Lebensverlaufsanalyse (Wagner et al., 1994, S. 71) bestimmt. Das SOEP ist ein Haushaltspanel, stellt aber durch die Befragung jeder erwachsenen Person des Haushalts auch Information über die Mikroeinheiten des Haushalts zur Verfügung. Das Fragenprogramm umfasst standardmäßig folgende Bereiche (vgl. Wagner et al., 1994, S. 72): Demographie und Bevölkerung, Arbeitsmarkt und Beschäftigung, Einkommen, Steuern und soziale Sicherung, Wohnen, Gesundheit, Bildung und Qualifikation, Leistungen privater Haushalte sowie Grundorientierungen, Partizipation und Integration. Ergänzt wird dieses Standarderhebungsprogramm des SOEP durch wellenspezifische Erhebungsschwerpunkte (bspw. 1988 sowie 2002 der Schwerpunkt „individuelles Vermögen und

Vermögen der privaten Haushalte“).

Im Folgenden werden die Grundgesamtheit (Abschnitt 4.1) und die einzelnen Stichproben des SOEP (Abschnitt 4.2.1-4.2.7) und das für Panelerhebungen zentrale Weiterverfolgungskonzept (Abschnitt 4.3) vorgestellt. Die Darstellung der Datenstruktur und der Datenaufbereitung folgt schließlich in Kapitel 5.

4.1 Grundgesamtheit und Zielpopulation

Ziel von Umfragen sind Aussagen über eine Zielpopulation und Grundgesamtheit, aus der die Befragungseinheiten stammen; daher ist es notwendig, zunächst eine Grundgesamtheit zu bestimmen, über die Aussagen gemacht werden sollen (Zielpopulation). Dies ist im Falle des SOEP „die wohnberechtigte Bevölkerung in der Bundesrepublik Deutschland einschließlich West-Berlin im Jahr 1984 und die deutsche Wohnbevölkerung in der DDR im Juni 1990“ (Wagner et al., 1994, S.75). Die Bezeichnung „wohnberechtigt“ schließt Ausländer in Deutschland mit ein, allerdings sind andere Personengruppen (bspw. Personen ohne Wohnsitz) nicht in der Grundgesamtheit erfasst. Zudem zeigt sich, „dass auch im SOEP die Anstaltsbevölkerung⁷⁴ [die prinzipiell zur Grundgesamtheit der wohnberechtigten Bevölkerung gehört, TG] aus praktischen Gründen der Stichprobenziehung sowie des Feldzugangs nicht hinreichend repräsentiert werden kann“ (Wagner et al., 1994, S. 74); dies ist generell eine Einschränkung der meisten Bevölkerungsumfragen die die „Wohnbevölkerung“ zur Grundgesamtheit haben (siehe dazu bspw. Schnell et al., 1999, S. 248)⁷⁵.

⁷⁴„Bewohner von Wohnheimen (Schüler, Lehrlings-, Studenten-, Berufstätigen- und Altenwohnheime im Rahmen des ADM-Samples sind in die Stichprobenziehung einbezogen. Anstaltsinsassen im engeren Sinn (Krankenhäuser/Sanatorien, Altenpflegeheime, Kasernen und Strafanstalten) wurden in der Ausgangsstichprobe des Panels nicht erfasst“ (Wagner et al., 1994, S. 75). Schnell (1997) berichtet, dass auch beim ADM-Design die Zugehörigkeit zur Anstaltsbevölkerung mehr oder minder willkürlich vom Interviewer festgestellt wird (S. 25).

⁷⁵Ein gegenteiliges Beispiel der Verzerrung der Grundgesamtheit *durch Hineinwachsen in* den Anstaltsbereich beschreiben Waterton und Liewesley (1987, S. 270) für einen Einstellungssurvey in Großbritannien: „Although people in institutions were not part of the target population [...], if panel members moved into an institution they were interviewed if willing“.

Eine Besonderheit der Abgrenzung der Grundgesamtheit des SOEP liegt in der Erhebungskonzeption des SOEPs begründet: aus „stichproben- und erhebungstechnischen Gründen“ (Wagner et al., 1994, S.75) erfolgte die Stichprobenziehung zu Start des Panels getrennt für die Teilpopulationen „deutsche Haushalte“ und „ausländische Haushalte“. Stichprobe A (deutsche Haushalte) hat alle Personen in Privathaushalten zur Grundgesamtheit, deren Haushaltsvorstand *nicht* türkischer, griechischer, jugoslawischer, spanischer oder italienischer Staatsangehörigkeit ist; die angestrebte Grundgesamtheit von Stichprobe B (ausländische Haushalte) umfasst hingegen alle Privathaushalte in Deutschland, deren Haushaltsvorstand die türkische, griechische, jugoslawische, spanische oder italienische Staatsangehörigkeit besitzt. Damit fallen unter die Grundgesamtheit der Stichprobe A auch alle ausländischen Haushalte, deren Haushaltsvorstand Angehöriger eines anderen, nicht von der Grundgesamtheit von Stichprobe B abgedeckten Staates ist (vgl. SOEP-Gruppe, 2001, S. 8).

4.2 Die Substichproben des SOEP

Unter dem SOEP sind mehrere Substichproben zusammengefasst; neben den beiden Startstichproben im Jahr 1984 „deutsche Haushalte“ und „Ausländer“ ist das SOEP in den vergangenen 22 Jahren um weitere Substichproben ergänzt worden. Mittlerweile sind sieben einzelne Stichproben unter dem SOEP zusammengefasst und im SOEP integriert (siehe Tabelle 4.1).

4.2.1 Stichprobe A (Westdeutschland)

1984 startete das SOEP mit der Stichprobe A „deutscher Haushalte“. Grundgesamtheit sind alle Personen in Privathaushalten (mit den genannten Einschränkungen, s.o.). Das Verfahren der Stichprobenziehung „gleich weitgehend dem Standardverfahren der Arbeitsgemeinschaft Deutscher Marktforschungsinstitute (ADM)“ (Rendtel und Projektgruppe Sozio-Oekonomisches Panel, 1995, S. 31)⁷⁶: Das ADM-Masterband von 1982 diente als Grundlage

⁷⁶Zu mehrstufigen Auswahlverfahren (u.a. auch das ADM-Design) allgemein siehe Schnell et al. (1999, S. 264ff), zum ADM-Design und Änderungen am ADM-Design siehe ausführlich Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute (ADM) und Arbeitsgemeinschaft Media Analyse (AG.MA) (1999, v.a. S. 61ff)

zur Auswahl auf der ersten Auswahlstufe. Insgesamt wurden 3.360 Wahlkreise durch „systematisches Ziehen mit Intervall und zufälliger Startzahl“ ausgewählt (Rendtel und Projektgruppe Sozio-Oekonomisches Panel, 1995, S. 32). Aus diesen Wahlkreisen wurden in einer geschichteten Stichprobe wiederum 563 Sampling Points ausgewählt⁷⁷. Innerhalb der Sampling Points wurden Haushalte nach dem Random-Route-Verfahren ausgewählt. Von einer vorgegebenen Startadresse ausgehend wurden 84 Haushalte entlang der Route notiert und schließlich jeder siebte Haushalt davon für die Befragung ausgewählt⁷⁸. Die Nettostichprobe lag mit 4.554 Haushalten schließlich über dem Sollumfang von 4.500 Haushalten (vgl. Wagner et al., 1994, S. 76).

4.2.2 Stichprobe B (Ausländer)

1984 stand vor allem der Vergleich deutscher Haushalte mit den Gastarbeiterhaushalten der Anwerbeländer Türkei, Jugoslawien, Griechenland, Spanien und Italien im Vordergrund (Rendtel und Projektgruppe Sozio-Oekonomisches Panel, 1995, S. 30). Stichprobe B bestand „genau genommen wiederum aus fünf eigenständigen Teilstichproben der fünf zahlenmäßig größten ausländischen Nationalitäten, die 1984 in der Bundesrepublik lebten“ (Wagner et al., 1994, S.76).

Für diese Teilpopulationen wurde ein disproportionaler Auswahlansatz verfolgt, sodass Haushalte aus diesen Ländern einen höheren Auswahlatz erhalten, als es ihrem Anteil an der Population (alle Haushalte in der Bundesrepublik) entspricht (vgl. Wagner et al., 1994, S. 76). Stichprobe B ist eine mehrstufig geschichtete Auswahl mit einer Registerstichprobe aus den Ausländerregistern der Sampling Points auf der letzten Auswahlstufe. Für jede der Nationalitäten der Stichprobe B wurde zunächst eine Zufallsauswahl von Kreisen und kreisfreien Städten (die Sampling Points) durchgeführt, geschichtet nach der Anzahl der Personen der entsprechenden Nationalität. Für die nichttürkischen Nationalitäten wurde ein dispro-

⁷⁷Die Schichten wurden durch Bundesland, Regierungsbezirk und Gemeindetyp vorgegeben (148 Regionalzellen); Wagner et al. (1994, S. 76) sprechen von 584 Sampling Points

⁷⁸Insgesamt wurden pro Sample Point 10 Adressen notiert, wovon nur acht für die Befragung anzuwerben waren; zwei Adressen wurden als Reserve eingesetzt, wenn weniger als acht Haushalten eines Sample Points zur Teilnahme bereit waren (vgl. Rendtel und Projektgruppe Sozio-Oekonomisches Panel, 1995, S. 33).

proportionaler Auswahlatz angestrebt. Schließlich wurden in den ausgewählten Sample Points aus den Ausländerregistern wiederum durch systematisches Ziehen mit zufälliger Startzahl und Intervall 20 Adressen gezogen (6 davon als Reserveadresse aufgrund eines erwarteten höheren Anteils neutraler Ausfälle (Wagner et al., 1994, S. 77)). Diese 14 Adressen wurden aufgrund „regionaler Gesichtspunkte“ zu zwei Klumpen zusammengefasst und einer der beiden Klumpen (= 7 Adressen) zur Befragung ausgewählt (Rendtel und Projektgruppe Sozio-Oekonomisches Panel, 1995, S. 33). Der Haushalt der so ausgewählten Befragungsperson gehörte dann zur Grundgesamtheit, wenn der Haushaltsvorstand die gleiche Nationalität wie die ausgewählte Person besitzt (Wagner et al., 1994, S. 77). Stichprobe B umfasste 1984 als Nettobestand schließlich 1.415 Haushalte.

4.2.3 Stichprobe C (DDR/Ostdeutschland)

Im Juni 1990 - mit der Wirtschafts- und Währungsunion aber noch vor dem Beitritt der DDR zum Gebiet der Bundesrepublik - wurde das SOEP um eine neue Stichprobe „Deutsche in der DDR“ (Stichprobe C) erweitert. Für die Stichprobenziehung wurde auf das zentrale Einwohnerregister der DDR zurückgegriffen (genauer: Verwendung des Infratest- Master-Samples für die DDR, vgl. Wagner et al. (1994, S. 77) und Rendtel und Projektgruppe Sozio-Oekonomisches Panel (1995, S. 33)). Aus dem Master-Sample wurden zunächst 753 Sample-Points ausgewählt, geschichtet nach Land- und Stadtkreisen und Gemeindegrößenklassen mit der Schichtgröße proportional zur Einwohneranzahl (Rendtel und Projektgruppe Sozio-Oekonomisches Panel, 1995, S. 34). Aus diesen Sample-Points wurden wiederum in einer geschichteten Zufallsauswahl 360 Sample-Points ausgewählt (mit Schichtgröße dieses Mal proportional zur Anzahl der Haushalte). Aus den Adressen der 360 Sample-Points wurden schließlich Startadressen für einen Random-Walk ausgewählt: Ausgehend von dieser Startadresse war jeder dritte Haushalt als Befragungshaushalt zu notieren und zur Teilnahme zu bewegen⁷⁹. Stichprobe C umfasste in der ersten Welle 2.179 Haushalte.

⁷⁹Im Unterschied zu Stichprobe A „musste auf einen getrennten Adressenvorlauf verzichtet werden“ (Wagner et al., 1994, S.78).

4.2.4 Stichprobe D (Zuwanderer)

Zuwanderung aus dem Ausland war bis 1994 nicht durch das SOEP abgedeckt, lediglich der Nachzug von Familienmitgliedern in Gastarbeiterhaushalte der Stichprobe B bzw. in Haushalte mit Haushaltsvorstand anderer Nationalität der Stichprobe A wurde durch das Haushalts- und Weiterverfolgungskonzept erfasst, ebenso Übersiedler aus der DDR in die BRD: „Diese Situation hat sich aber in den letzten Jahren grundlegend gewandelt“ (Rendtel und Projektgruppe Sozio-Oekonomisches Panel, 1995, S. 41): Das SOEP stand vor dem Problem der zahlenmäßig bedeutsamen, im SOEP aber nicht erfassten Zuwanderung deutschstämmiger Personen (Aussiedler und Übersiedler) aus Osteuropa. Aus diesem Grund wurde 1994 eine weitere Stichprobe von Zuwandererhaushalten gezogen (SOEP-Gruppe, 2001; Burkhauser et al., 1997), die zwischen 1984 und 1994 in die Bundesrepublik gezogen sind⁸⁰. Im Gegensatz zur ausländischen Bevölkerung existiert für Aus- und Übersiedler kein Register, beide Gruppen besitzen außerdem die deutsche Staatsangehörigkeit. Für ein Adressscreening von Zuwandereradressen wurde ein mehrstufiges Auswahlverfahren verwendet: Im Rahmen zweier allgemeiner Umfragebusse (Infratest-Bus und Infra-Scope⁸¹) wurden 1992 über eine Screeningfrage Zuwandererhaushalte identifiziert und um Teilnahme am SOEP gebeten⁸². Teilnehmende Haushalte wurden dann gebeten, Adressen von weiteren befreundeten Zuwandererhaushalten anzugeben. Dieses sogenannte Schneeballverfahren wird allgemein bei Stichproben für seltene Populationen verwendet (vgl. Schnell et al., 1999, S. 280) und findet auch in der Netzwerkanalyse Anwendung (Scott, 1991, S. 61). Schneeballverfahren sind allerdings

⁸⁰Genauer: nach Westdeutschland; Einwanderung in die DDR bis 1990 bzw. Ostdeutschland nach 1990 wurde nicht berücksichtigt, da nur wenige Personen in dieser Zeit in die DDR bzw. nach Ostdeutschland zogen (Burkhauser et al., 1997, S. 10).

⁸¹Der Infratest-Bus basiert auf dem ADM-Design (Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute (ADM) und Arbeitsgemeinschaft Media Analyse (AG.MA), 1999), Infra-Scope ist eine Telefonstichprobe (vgl. Burkhauser et al., 1997, S. 12). Zu Telefonstichproben in Deutschland siehe bspw. Gabler (1998), zu Stichproben für Zuwanderer in Deutschland bspw. Salentin (1999) oder Humbert und Schneiderheinze (2000).

⁸²Von 330 Screeninghaushalten erklärten sich 195 zur Teilnahme bereit, 138 wurden schließlich erfolgreich in einer ersten Welle interviewt (Burkhauser et al., 1997, S. 13).

keine zufälligen Auswahlen, sondern Verfahren, die auf einer bewussten Auswahl basieren (vgl. Schnell et al., 1999, S. 280)⁸³. Die Befragung der Screeninghaushalte nach weiteren Zuwandererhaushalten erbrachte 141 weitere Zuwandererhaushalte, 98 davon zu einer Teilnahme bereit. Burkhauser et al. (1997) beschreiben die Probleme, die mit diesem Ansatz verbunden sind (S. 13): Zum einen findet durch die Wahl des Infratest-Busses (Stichprobe nach dem ADM-Design) als Auswahlgrundlage ein Oversampling von Aussiedler und übersiedlern gegenüber sonstigen Einwanderern statt⁸⁴, durch das Schneeballverfahren findet implizit ein weiteres Oversampling von Aussiedlern gegenüber übersiedlern und sonstigen Zuwanderern statt⁸⁵. In einem zweiten Adressscreening 1994 auf Basis des Infratest-Busses wurde aufgrund der Probleme und Selektionseffekte auf ein weiteres Schneeballverfahren verzichtet und lediglich gescreente Haushalte in die Stichprobe übernommen. In diesem, zweiten Screening wurden 725 Zuwandererhaushalte identifiziert, 548 davon waren zur Teilnahme bereit. Um das Oversampling des ersten Screenings nicht weiter zu verstärken, wurden 148 Haushalte von Aus- und übersiedler zufällig gelöscht, 295 Haushalte wurden schließlich erfolgreich interviewt (Burkhauser et al., 1997, S. 13-14). Insgesamt besteht die Zuwandererstichprobe D in der ersten Welle aus 522 erfolgreich interviewten Haushalten (ohne Schneeballhaushalte 427).

⁸³Im Rahmen der Netzwerkanalyse sind Auswirkungen und Probleme dieses Auswahlverfahrens besonders anschaulich: „A snowball sample, of course, is not a random sample: the structure that is discovered is, in fact, ‘built in’ to a snowball sampling method itself“ (Scott, 1991, S. 61).

⁸⁴Das ADM-Design wählt Wahlkreise als Sample Points; die Wahlkreisgröße hängt von der Anzahl der wahlberechtigten Bevölkerung innerhalb des Wahlkreises ab; Aus- und übersiedler gehören zu dieser wahlberechtigten Bevölkerung, Zuwanderer anderer Nationalitäten, abgesehen von EU-Ausländern bei Kommunal- und Europawahlen, in der Regel nicht.

⁸⁵„[...] Ethnic Germans tend to immigrate in networks [...], while übersiedler, asylum seekers, and people from EU-member states are less likely to rely on networks“ (Burkhauser et al., 1997, S. 13). Auch Salentin (1999, S. 125) beschreibt dieses Problem bei Schneeballverfahren und Zuwandererstichproben: zum Einen finde „Migration kettenförmig statt“, zum Anderen migrierten Aussiedler im Gegensatz zu anderen Zuwanderergruppen häufiger im Familienverband, dadurch ergäben sich auch „unterschiedliche Reichweiten von Netzwerken“ (S. 125).

4.2.5 Stichprobe E (Stabilisierung, Querschnitt)

Eine weitere Stichprobe wurde 1998 gezogen, hauptsächlich, um neue Erhebungsmodi⁸⁶ und Paneffekte (Unterschiede und Veränderungen aufgrund der Teilnahme an einem Panel) zu untersuchen. Das Stichprobendesign entspricht dem der Stichprobe A für Westdeutschland (Abschnitt 4.2.1), jedoch ausgeweitet auch auf die neuen Bundesländer. Stichprobe E umfasste in der ersten Welle 1.067 Haushalte. Stichprobe E war unabhängig von den SOEP-Altstichproben mit Privathaushalten in der gesamten Bundesrepublik als Grundgesamtheit (vgl. SOEP-Gruppe, 2001, S. 8) und 1998 ein repräsentativer (im Sinne einer Zufallsstichprobe) Querschnitt der in Deutschland lebenden Wohnbevölkerung⁸⁷.

4.2.6 Stichprobe F (Innovation)

2000 wurde das SOEP um eine weitere Stichprobe erweitert; mit Stichprobe F sollten im Rahmen des SOEP „bestimmte innovative Methoden“ (Infratest Sozialforschung, 2001, S. 3) erprobt werden. Vom Design angelegt war Stichprobe F ähnlich wie der 1998 erstmals befragte Querschnitt der Stichprobe E (Abschnitt 4.2.5). Betont wurde jedoch bei der Stichprobenziehung die Repräsentation der ausländischen Wohnbevölkerung, die ausgehend vom ADM-Design zwar explizit zur Grundgesamtheit gehört, aber ausgehend von Wahlkreisen als Primary Sampling Units (PSU) untererfasst wird⁸⁸. Die Untererfassung ausländischer Haushalte auf der ersten Stufe soll dadurch behoben werden, dass auf der zweiten Auswahlstufe (Auswahl der Haushalte in den Sampling Points durch Random-Route) der Behebungsweg für die Identifizierung der ausländischen Haushalte verdoppelt wird: „Damit wird die Wahrscheinlichkeit, auf Ausländer-Haushalte zu treffen, theoretisch um den Faktor 2 erhöht“ (Infratest Sozialforschung, 2001, S. 11)⁸⁹. In allen 986 Sampling Points der ersten Auswahlstufe

⁸⁶Der größte Unterschied zu den SOEP-Altstichproben besteht im Einsatz von Laptops beim persönlichen Interview (CAPI Computer Assisted Personal Interview).

⁸⁷mit den in Abschnitt 4.1 erwähnten Einschränkungen und Problemen.

⁸⁸Vergleichbar dem Problem bei der Zuwandererstichprobe, siehe Fußnote 84.

⁸⁹Dieses Verfahren wird jedoch auch nur als „second best-Lösung“ beschrieben, da auch „damit die strukturelle Verzerrung in der Auswahl der Sample-Points nicht behoben wird (Infratest Sozialforschung, 2001,

wurden 12 Adressen durch Random-Walk erfasst, die restlichen 12 Adressen des verlängerten Begehungsweges wurden dann erfasst, wenn im Haushalt mindestens eine Person nicht-deutscher Staatsangehörigkeit lebt⁹⁰. Mit 6.060 erfolgreich interviewten Haushalten in der ersten Welle hat Stichprobe F die Fallzahl des SOEP deutlich aufgestockt und die Erfassung der ausländischen wohnberechtigten Bevölkerung verbessert (vgl. SOEP-Gruppe, 2001, S. 8).

4.2.7 Stichprobe G (High Income)

Ziel einer 2002 neu gestarteten Stichprobe war vor allem die „Bereitstellung einer verbesserten Mikrodatenbasis zur Analyse von Haushalten im oberen Einkommensbereich“ (Schupp et al., 2003, S. 1), unter anderem vor dem Hintergrund eines in der Umfrageforschung oft beobachteten „Mittelstandsbiases“, der sich unter anderem durch im Vergleich niedrigeren Teilnahmeraten an den Rändern der Verteilung sozioökonomischer Variablen (bspw. Bildung oder Einkommen) bemerkbar macht (vgl. Rendtel und Projektgruppe Sozio-Oekonomisches Panel, 1995, S. 193). Schließlich sollte die Datenbasis für belastbare Aussagen für die Armuts- und Reichtumsberichterstattung der Bundesregierung verbessert werden (Schupp et al., 2003, S. 9). Dabei ergibt sich ein ähnliches Problem der Auswahlgrundlage wie bei der Zuwandererstichprobe (Abschnitt 4.2.4): „Es gibt kein Verzeichnis aller Haushalte im oberen Einkommensbereich in Deutschland, aus dem man direkt nach Zufallsverfahren eine Stichprobe ziehen könnte“ (Infratest Sozialforschung, 2002a, S. 5). Wie bei der Zuwandererstichprobe wurde auch hier ein Huckepackverfahren über eine Screeningfrage nach dem Haushaltseinkommen im Rahmen eines großen Umfragebusses gewählt. Als Screeningbasis wurde die telefonische Mehrthemenbefragung Infra-Scope von Infratest verwendet. Infra-Scope ist eine Telefonstichprobe nach ADM-Design (vgl. Infratest Sozialforschung, 2002a, S. 8)⁹¹. Als

S. 11): Ausländer, die in Gebieten mit hohem Ausländeranteil leben, erhalten dadurch eine geringere Auswahlwahrscheinlichkeit als Ausländer in Gebieten mit niedrigem Ausländeranteil (S. 11).

⁹⁰Zum Teil wurden Adressen des verlängerten Begehungsweges auch als Ersatzadressen für qualitätsneutrale Ausfälle (bspw. Geschäftsadressen oder unbewohnte Wohnungen und Häuser) unter den ersten 12 Adressen verwendet (Infratest Sozialforschung, 2001, S. 15).

⁹¹Eine Beschreibung des Infratest-Telefon-Master-Samples (ITMS) findet sich bei Infratest Sozialforschung (2002a, S. 9f), zur Möglichkeit und Einschränkungen von Telefonstichproben (u.a. das sog. „Gabler-Häder-

untere Grenze des gescreenten Haushaltsnettoeinkommen wurde ein monatliches Haushaltsnettoeinkommen von 7.500DM und mehr festgelegt. Von 99.175 Haushalten in Infra-Scope 2001 liegen für 83.329 Haushalte Angaben zum Haushaltsnettoeinkommen vor, darunter 6.330 Haushalte mit Haushaltsnettoeinkommen von 7.500DM oder mehr (vgl. Infratest Sozialforschung, 2002a, S. 13ff). Voraussetzung dafür, dass diese Haushalte an der SOEP-Erhebung teilnehmen ist die Zustimmung zur Speicherung ihrer Telefonnummer. Das ist bei knapp 90% der Haushalte der Fall (87% bei allen Infra-Scope-Haushalten). Um eine regionale Klumpung der Befragungshaushalte für die mündlich-persönlich durchzuführende Befragung der 5.663 zur Teilnahme bereiten Haushalte zu erhalten, wurden rund 2.000 „überflüssige“ (S. 16) Haushalte aus der Stichprobe gelöscht; dabei wurde zudem eine disproportional Schichtung der Stichprobe nach Einkommenshöhe (erhöhter Auswahlatz für Haushalte mit Haushaltsnettoeinkommen über 10.000DM) und Region (erhöhter Auswahlatz für die neuen Bundesländer) vorgenommen. Für die verbleibenden 3.672 Haushalte wurden telefonische Kontaktinterviews durchgeführt, um sie zur Teilnahme am SOEP zu gewinnen und Adressinformationen zu erhalten⁹², 2.495 Haushalte erklärten in den telefonischen Kontaktinterviews ihre Teilnahmebereitschaft⁹³ und erhielten erstes Informationsmaterial über das SOEP zugesandt. Insgesamt 300 Haushalte sind für die mündlich-persönliche Erstbefragung ausgefallen, nahezu vollständig durch Ausfälle von Interviewern (S. 20). 488 Haushalte verweigerten dem Interviewer ein Interview, zu 70 Haushalten konnte kein Kontakt hergestellt werden, 6 Haushalte waren aus Krankheits- oder Altersgründen nicht interviewbar. Dadurch reduziert sich die Anzahl der erfolgreichen Erstinterviews auf 1.631. In einem letzten Schritt wurde überprüft, ob der befragte Haushalt tatsächlich zur Zielpopulation der Haushalte mit einem Haushaltsnettoeinkommen von 7.500DM und mehr gehört. Im Infra-Scope wird das Einkommen in groben Gruppen abgefragt, im SOEP als offene Frage, ausserdem unterscheiden sich die Auskunftspersonen zwischen beiden Erhebungen, schließlich können zwischen

Verfahren“) sei auf Gabler (1998) verwiesen.

⁹²Bis zu diesem Zeitpunkt liegt lediglich eine Telefonnummer als Kontaktinformation vor.

⁹³Zu keiner Zeit wurden die kontaktierten Haushalten darüber informiert, dass sie eine Sondergruppe innerhalb des SOEPs darstellen (Infratest Sozialforschung, 2002a, S. 20).

Tabelle 4.1: Überblick über die SOEP-Substichproben

Stichprobe	Startjahr	Umfang	Beschreibung
A	1984	4.554	HH Westdeutschland
B	1984	1.415	ausländische HH Westdeutschland
C	1990	2.179	HH DDR
D	1994/1995	522	Zuwanderer-HH Westdeutschland
E	1998	1.067	HH Deutschland, Querschnitt, CAPI-Test
F	2000	6.060	HH Deutschland, vorwiegend CAPI
G	2002	1.224	HH Deutschland, HHsnettoeinkommen $\geq 7.500DM$
Insgesamt	1984-2002	17.021	HH Deutschland

den beiden Befragungen bis zu 18 Monate vergangen sein. Stichprobe G (High Income) besteht in der ersten Welle schließlich aus 1.224 Haushalten, die zur Zielpopulation „Haushalte mit Haushaltsnettoeinkommen 7.500DM und mehr“ gehören (vgl. Infratest Sozialforschung, 2002a, S. 23).

Tabelle 4.1 zeigt abschließend einen Überblick über die SOEP-Stichproben A-G zusammen mit dem Jahr der ersten Welle und der Anzahl der Haushalte, die in den jeweiligen ersten Wellen befragt wurden. Tabelle 4.2 zeigt zusammenfassend das Design der einzelnen Stichproben, ihre Grundgesamtheit und den durchschnittlichen Auswahlsatz.

Abbildungen 4.1 und 4.2 zeigen die Entwicklung der Fallzahlen der Personen (ohne Kinder bis 16 Jahre) und Haushalte. Deutlich ist der Zuwachs an Haushalten und Personen durch die neuen Substichproben (dicke Punkte auf den Linien) in Abbildung 4.1 zu sehen. Deutlich ist auch, dass das SOEP vor der Änderung des Weiterverfolgungskonzepts (siehe folgender Abschnitt) vergleichsweise schneller schrumpft als in den Jahren nach 1990. Abbildung 4.2 zeigt die Entwicklung der Fallzahlen über alle Wellen des SOEP nach der Teilnahmedauer: 3.470 Personen, haben über alle 22 Wellen des SOEP seit seinem Start 1984 teilgenommen, das sind über 28% des ursprünglichen Samples 1984. Insgesamt sind seit 1984 über 47.000 Personen in Kontakt mit dem SOEP gekommen, davon haben über 43.000 mindestens ein erfolgreiches Interview absolviert. Auffällig in dieser Betrachtung ist der starke Rückgang

Tabelle 4.2: Überblick über die SOEP-Substichproben: Auswahlsatz und Design

Stichprobe	Auswahlsatz*	Design
A	0.0002	ADM-Design
B	0.0008	mehrstufige Registerstichprobe
C	0.0004	mehrstufige Registerstichprobe
D	0.0002	ADM-Design, Telefon, Schneeball
E	0.00003	ADM-Design, CAPI-Test
F ¹	0.00028	abgewandeltes ADM-Design, vorwiegend CAPI
F ²	0.0005	
G	– ³	ADM-Telefon, Abschneideverfahren

¹: deutsche Haushalte; ²: ausländische Haushalte; ³: keine Angabe; *: Quelle: Haisken DeNew und Frick (2005, S. 19ff)

nach 6 bzw. 16 Teilnahmejahren für Personen und Haushalte.

4.3 Weiterverfolgungskonzept und Panelpflege des SOEP

Ziel eines Panels ist die Befragung der gleichen Untersuchungseinheiten in bestimmten Intervallen; das bedeutet, dass die gleichen Untersuchungseinheiten verfolgt werden müssen, damit sie zu einem späteren Zeitpunkt wieder befragt werden können: „Für eine Panelerhebung sind das Konzept der Weiterverfolgung der Stichprobenmitglieder und die Sicherung von deren Teilnahmebereitschaft die zentralen Parameter“ (Wagner et al., 1994, S. 85). Hier zeigt sich nach Rendtel und Projektgruppe Sozio-Oekonomisches Panel (1995, S. 35) auch ein „Dilemma, das die Arbeit mit Paneldaten wie ein roter Faden durchzieht: [...] die Schwierigkeit, die Erhebungseinheiten im Längsschnitt zu definieren“. Diese Erhebungseinheiten sind besonders bei Haushaltspanels selbst Veränderungen unterworfen: Haushalte lösen sich auf, teilen sich (bzw. gründen sich neu), sind als ganze mobil, vergrößern oder verkleinern sich. Daraus zieht Ruspini (2002, S. 31) den Schluss: „It is individuals, not households, who are followed over time: individuals are much more stable in a longitudinal context and so are

Abbildung 4.1: Entwicklung der Fallzahlen im SOEP, Haushalte und Personen, 1985-2005

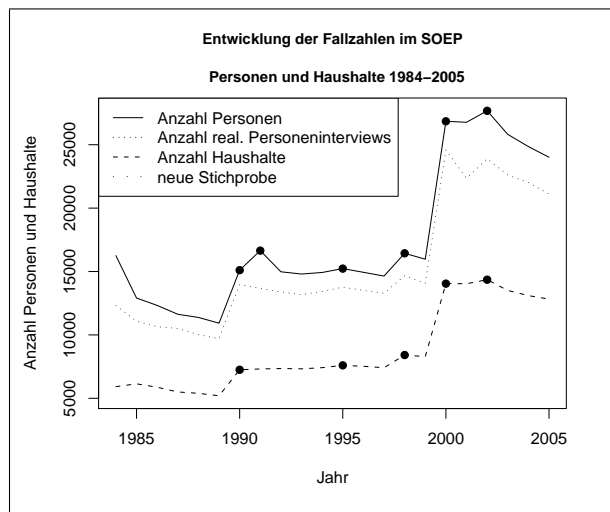
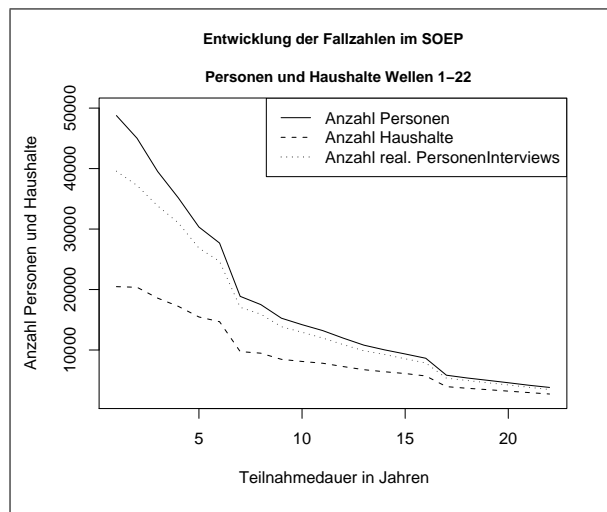


Abbildung 4.2: Entwicklung der Fallzahlen, nach Teilnahmejahren



easier to track and follow“⁹⁴.

4.3.1 Weiterverfolgungsregeln

Für ein Weiterverfolgungskonzept der Personen eines Haushaltspanels müssen Regelungen getroffen werden, welche Personen aus einem Befragungshaushalt in der nächsten Welle wieder befragt werden. Ab hier sind die Bezeichnungen „Stammhaushalt“ und „Stammpersonen“ hilfreich; als Stammhaushalt soll ein Haushalt dann bezeichnet werden, wenn er bei der Stichprobenziehung als Befragungshaushalt ausgewählt worden ist, als Stammpersonen alle Personen des Stammhaushalts. Ein Stammpersonenkonzept verfolgt nur Personen des Stammhaushalts in die nächste Welle, ausser den im Stammhaushalt verbleibenden Stammpersonen werden nur Kinder des Stammhaushalts, die erstmals das Befragungsalter erreichen,

⁹⁴Eine Variante dieser Trackingregel, die keine Personen über die Zeit verfolgt, beschreiben bspw. Thomas et al. (2001, S. 561) für das Weltbank-Panel zur Erfassung des Lebensstandard (oder allgemein für Länder mit sehr hoher Mobilität, wie sie bspw. für Entwicklungsländer charakteristisch ist): „Most panels in developing countries have revisited the original housing structure and interviewed whoever is there, if anyone“. Und weiter heisst es dort: „Following dwellings is the simplest option, because dwellings almost never move“ (zitiert nach Thomas et al. (2001)).

in der nächsten Welle befragt. Zudem müssen auch Personen in die nächste Welle verfolgt werden, die den Stammhaushalt verlassen und einen eigenen Haushalt neu gründen oder in einen bestehenden Nicht-Stammhaushalt ziehen. In diesen Nicht-Stammhaushalten wird aber lediglich die ursprüngliche Stammperson befragt⁹⁵. Dieses Stammpersonenkonzept ist allerdings mit Nachteilen behaftet (vgl. Rendtel und Projektgruppe Sozio-Oekonomisches Panel, 1995, S. 38): zum einen ist es ineffizient, Nicht-Stammpersonen in den neugegründeten Haushalten auszuschließen, ausserdem schrumpft ein Panel auf diese Weise schneller⁹⁶, als bei anderen Weiterverfolgungskonzepten. Dieses Stammpersonenkonzept wird bspw. vom SOEP-Vorbild PSID (Panel Study of Income Dynamics) verwendet (Lillard und Panis, 1998, S. 438); bis zur 6. Welle des SOEP (1989) galt dieses Stammpersonenkonzept auch für das SOEP (Rendtel und Projektgruppe Sozio-Oekonomisches Panel, 1995, S. 39)⁹⁷. Das Britische Haushaltspanel BHPS verwendet ein umfassenderes Weiterverfolgungskonzept⁹⁸ (vgl. Ruspini, 2002, S. 89): Zunächst werden auch hier alle Stammpersonen weiterverfolgt und Jugendliche ab dem 16. Geburtstag in die Stichprobe mit einbezogen; zudem werden abweichend vom reinen Stammpersonenkonzept bei Haushaltsneugründungen⁹⁹ durch eine Stammperson alle Haushaltsmitglieder dieses Haushalts befragt und weiterverfolgt, solange die Stammperson im Nicht-Stammhaushalt lebt (vgl. Laurie et al., 1999, S. 271). Zuzüge von Nicht-Stammpersonen in einen Stamm-Haushalt werden analog dazu befragt und weiterver-

⁹⁵In der amerikanischen PSID werden zusätzlich zu den Stammpersonen nur die (neuen) Ehepartner einer Stammperson mit befragt, allerdings nur solange sie Teil des Stammhaushalts sind (vgl. Lillard und Panis, 1998).

⁹⁶„Bei der praktischen Umsetzung zeigte sich jedoch, dass die demographischen Verluste und die Verluste durch die Teilnehmerverweigerungen größer sind als die Stichprobengewinne“ (Rendtel und Projektgruppe Sozio-Oekonomisches Panel, 1995, S. 39).

⁹⁷Wagner et al. (1994, S. 86) schreiben, dass ab Welle 7 (West) 1990 das Weiterverfolgungskonzept des SOEP geändert wurde, bei Ruspini (2002, S. 90) findet sich wiederum Welle 6 als erste Welle mit geändertem Weiterverfolgungskonzept.

⁹⁸Dieses Weiterverfolgungskonzept wird bei Ruspini (2002) als „Original Sample Members-Konzept“ bezeichnet (S. 89).

⁹⁹Genauer: Haushaltsneugründung durch eine Stammperson oder Zuzug einer Stammperson in einen bestehenden Nicht-Stammhaushalt.

folgt, solange sie Mitglieder im Stammhaushalt sind¹⁰⁰. Obwohl das SOEP ursprünglich das Stammpersonenkonzept verwendet hat, „ist es für Längsschnittanalysen, insbesondere von demographischen Ereignissen, aber sinnvoller, nicht nur Stammpersonen, sondern alle Personen weiterzuverfolgen, die jemals in Kontakt mit dem SOEP gekommen sind“ (Wagner et al., 1994, S. 85); weiter heisst es dort: „Bei der Umsetzung des ursprünglichen SOEP-Konzepts zeigte es sich ohnehin, dass das Weiterverfolgungskonzept von Stammpersonen für die praktische Feldarbeit zu kompliziert war, das heißt, es wurden auch Personen interviewt, die nach den Regeln des Stammpersonenkonzepts nicht hätten interviewt werden dürfen“, zudem vereinfache sich bei einem umfassenderen Konzept die Berechnung der Ziehungswahrscheinlichkeiten (S. 86). Dieses Weiterverfolgungskonzept des SOEP schließt Wechsel von Haushaltsmitgliedern in Anstaltsunterkünfte mit ein (vgl. Rendtel und Projektgruppe Sozio-Oekonomisches Panel, 1995, S. 36; sowie Haisken DeNew und Frick, 2005, S. 19)¹⁰¹: im SOEP „bestand anfänglich die Hoffnung, dass die Panelstichprobe - via Weiterverfolgung - in den Folgewellen gleichsam in den Anstaltsbereich hineinwächst“ (Wagner et al., 1994, S. 94). Zudem müssen für Weiterverfolgungsregeln Regelungen getroffen werden, wie weit (bei Umzug) und wie lange (bei temporärem Ausfall) eine Befragungsperson weiterverfolgt wird; Rendtel und Projektgruppe Sozio-Oekonomisches Panel (1995, S. 34): „In der Regel werden die Einheiten jedoch nur dann weiterverfolgt, solange sie innerhalb des Erhebungsgebietes umziehen“. Im SOEP werden Haushalte und Personen nur innerhalb Deutschlands weiterverfolgt, bei einem Umzug ins Ausland fallen diese Haushalte und Personen aus der Stichprobe¹⁰². Zudem scheiden Haushalte und Personen aus dem Panel aus, die endgültig

¹⁰⁰Für diese nicht weiter zu verfolgenden Personen liegen im Längsschnitt also nur temporär Informationen vor, „providing important contextual information for any longitudinal analyses, while also adding to the overall sample size for cross-sectional analysis“ (Laurie et al., 1999, S. 272).

¹⁰¹Im Anstaltsbereich, bspw. Altenheim, treten dann allerdings andere Probleme auf bspw. Kontaktprobleme (u.a. durch Gatekeeper) oder Teilnahmeschwierigkeiten aufgrund von Krankheit oder Alter.

¹⁰²Dies ist nicht unproblematisch, vor allem in Hinblick auf Stichprobe B der ehemaligen Gastarbeiterhaushalte. Bei großen Erhebungsgebieten wie bspw. beim Europäischen Haushaltspanel ECHP steht einer Weiterverfolgung der Erhebungseinheiten „die Komplexität der Feldorganisation entgegen, die unabhängig in den zwölf [1995, TG] Mitgliedsländern der EU organisiert ist“ (Rendtel und Projektgruppe Sozio-Oekonomisches Panel, 1995, S. 35); vgl. auch Peracchi (2002, S. 65): „Although Eurostat sets general guidelines in order to

„trotz intensiver Bemühungen nicht mehr aufzufinden“ sind (Wagner et al., 1994, S. 87) oder die ausdrücklich eine weitere Teilnahme an der Befragung verweigern. Zudem scheiden Haushalte und Personen aus, bei denen in zwei aufeinanderfolgenden Wellen kein Kontakt hergestellt oder kein Interview durchgeführt werden konnte. Durch die Veränderungen der Haushalte durch Auflösungen, Aufspaltungen, Zu- und Auszüge aus Befragungshaushalten ergeben sich im Längsschnitt Identifikationsprobleme. Grundsätzlich ist jeder Haushalt durch eine eindeutige Haushaltsidentifikationsnummer, jede Person durch eine eindeutige Personenidentifikationsnummer im Längsschnitt identifizierbar. Die Personenidentifikationsnummer ist nicht von Veränderungen der Haushalte betroffen, sie bleibt für alle Personen immer gleich. Durch Veränderungen der Haushalte besteht aber das Problem, dass Haushaltsidentifikationsnummern im Längsschnitt keine eindeutige Zuordnung der Personen zu Haushalten darstellen. Spalten sich Befragungshaushalte bspw. in zwei separate Haushalte auf, behält nur ein Teil des Haushalts die alte Identifikationsnummer, der andere Teil erhält eine neue eindeutige Haushaltsidentifikationsnummer. Rendtel und Projektgruppe Sozio-Oekonomisches Panel (1995, S. 36): „bei Haushaltsteilungen behält ein Teil die Identifikationsnummer (alter Haushalt), während der andere Teil eine neue Identifikationsnummer erhält (neuer Haushalt). Verbleibt ein Haushaltsteil an der alten Adresse, so wird er als alter Haushalt weitergeführt. Verbleibt kein Haushaltsteil an der alten Adresse, so wird der Haushaltsteil, in dem der frühere Haushaltsvorstand lebt als alter Haushalt weitergeführt“¹⁰³.

4.3.2 Panelpflege und Tracking

Neben Regelungen, welche Haushalte und Personen in der nächsten Welle wieder befragt werden sollen, müssen auch Vorkehrungen der Feldorganisation getroffen werden, wie zu befragende Haushalte und Personen wieder aufgefunden werden und deren Teilnahmebereitschaft

ensure comparability of survey results, the NDUs [Erhebungseinheiten der Mitgliedsländer, TG] largely rely on their normal rules and routines [...] making it difficult to assess data quality“.

¹⁰³Damit bleibt ein Problem, wenn zwei Haushaltsvorstände zweier Befragungshaushalte aus ihren ursprünglichen Haushalt ausziehen und einen neuen Haushalt gründen; Nach dieser Regelung erhalten die verbliebenen Ursprungshaushalte eine neue Identifikationsnummer, allerdings erhält auch der neugegründete Haushalt der Haushaltsvorstände eine neue eindeutige Identifikationsnummer.

gesichert werden können¹⁰⁴; diese Maßnahmen werden als Tracking-Techniken (Maßnahmen zum Wiederauffinden) und Panelpflege (Sicherung der Teilnahmebereitschaft) beschrieben. Beide sollen Ausfälle in der folgenden Welle reduzieren und zielen darauf ab, den Kontakt in der nächsten Welle zu ermöglichen und die Teilnahme kontaktierter Haushalte und Personen zu sichern oder erhöhen. Dabei sind Tracking und Panelpflege nicht immer voneinander zu trennen. Ruspini (2002, S. 90f) beschreibt mehrere Maßnahmen, um die Teilnahmebereitschaft in der nächsten Welle zu sichern; darunter bspw. Befragten die Wahl des Erhebungsmodus zu überlassen, so dass sie selbst zwischen persönlich-mündlichem, telefonischem oder schriftlichem Interview wählen können; ausserdem soll durch ein und denselben Interviewer ein Vertrauensverhältnis zwischen Befragtem und Interviewer ermöglicht und gefestigt werden. Auch ein personalisiertes Ankündigungsschreiben vor und Dankeschreiben nach Abschluss des Interviews sollen die Teilnahmebereitschaft in der nächsten Welle erhöhen. Im SOEP erhalten Befragte eine Broschüre über die Befragung unter einem leicht einprägsamen Namen „Leben in Deutschland“ zusammen mit dem Ankündigungsschreiben¹⁰⁵; darin wird

¹⁰⁴Ein besonderes Problem bereitet das Wiederauffinden von Befragungspersonen bei sehr mobilen Populationen; Hill (2004) beschreibt dies vor dem Hintergrund von Panelerhebungen in Entwicklungsländern: „In developed countries attrition is mostly due to refusals [...] whilst in developing countries attrition is mostly due to moves“ (S. 495, vgl. auch Thomas et al. (2001, S: 558)). Eine Empfehlung hier ist das Bekanntnetzwerk mit offiziellem *und Spitznamen* zu erfassen, das bei Umzug Auskunft über den Verbleib der Befragungspersonen geben kann; „it is also useful to record place of birth, ethnicity, date of birth, place of origin, the location of any property owned“ (S.496). Ein weiteres Problem beim Wiederauffinden nicht umgezogener Befragungspersonen bereiten fehlende Adresssysteme oder wenn Befragte ihre Adresse nicht kennen, aber den Weg dorthin beschreiben können; Hill (2004) nennt Beispiele aus Peru und Malaysia, bei denen Befragte Karten mit Orientierungspunkten zeichneten oder GPS-Koordinaten eingesetzt wurden (S. 496) oder auch Stromableser zu umgezogenen Personen und Haushalten befragt wurden (Hill, 2002, S. 36). Eine Beschreibung von Panelerhebungen in Entwicklungsländern und deren Trackingtechniken findet sich ausführlicher bei Hill (2002, S. 30ff). Thomas et al. (2001) ziehen aus den Problemen für Panelstudien in Entwicklungsländern allerdings nicht den Schluss, dass sie zwangsläufig unter höheren Ausfallraten leiden (S. 557f).

¹⁰⁵In der ersten Welle der Stichprobe C konnte aufgrund des fehlenden Adressvorlauf kein Ankündigungsschreiben versandt werden; „dies hat jedoch zu keinen nennenswerten Problemen geführt“ (Wagner et al., 1994, S. 78).

der Besuch des Interviewers angekündigt. Beim Interviewerbesuch bekommen Befragte zudem separate Informationen zum Datenschutz und ein kleines Geschenk ausgehändigt, zudem erhalten alle Teilnehmer ein Monatslos der „Aktion Mensch“. Um Befragungshaushalte und -personen wieder auffinden zu können werden bspw. Angabe einer kostenlosen Telefonnummer der Feldorganisation bzw. eine Antwortpostkarte zum Mitteilen von Adressänderungen empfohlen (das ist bspw. beim britischen Haushaltspanel BHPS der Fall; vgl. Laurie et al., 1999, S. 274); teilweise werden in Panelerhebungen Namen und Adressen von Freunden oder Bekannten erfragt, die im Falle eines Umzugs der Zielperson oder des Zielhaushalts Auskunft über die neue Adresse geben (Ruspini, 2002, S. 91). Im SOEP sind es oftmals die Interviewer selbst, die Adressinformationen über umgezogene Haushalte oder Personen liefern können. Adressrecherchen zu nicht mehr auffindbaren Haushalten werden ansonsten über Postämter oder die Einwohnermeldeämter¹⁰⁶ durchgeführt (Wagner et al., 1994, S. 92).

4.4 Interviewereinsatz beim SOEP

Für die Aufrechterhaltung der Teilnahmebereitschaft in einem Panel ist die Kontinuität der Interviewer-Befragten-Beziehung von großer Bedeutung¹⁰⁷: ein Wechsel des Interviewers gilt als guter Prädiktor für einen folgenden Ausfall (auch wenn, so Zabel (1998, S. 493), die Richtung der Kausalität nicht ganz klar ist); zudem sind es die Interviewer, die einen Ausfall einer Befragungsperson feststellen und erfassen, daher ist es denkbar, dass systematische Zusammenhänge zwischen einzelnen Interviewern und den von ihnen erfassten Ausfällen bestehen; bspw. dass einzelnen Interviewer von einzelnen Ausfallkategorien gehäuft Gebrauch machen¹⁰⁸.

¹⁰⁶Eine „einfache Melderegisterauskunft“ wird „im SOEP regelmäßig eingeholt, wenn Befragungspersonen nicht mehr an der im Vorjahr gültigen Adresse aufzufinden sind“ (Infratest Sozialforschung, 2002b, S. 4).

¹⁰⁷Zabel (1998, S. 488) betont die Bedeutung des Interviewers beim Ausfallprozess: Für den amerikanischen Survey of Income and Program Participation (SIPP) berichtet er zudem, dass erfolgreiche Teilnehmer der Welle im Jahr 1985, gefragt nach dem Grund ihrer wiederholten Teilnahme, als Hauptgrund angaben, den „Interviewer zu mögen“ (S. 492)

¹⁰⁸Generell besteht bei der Untersuchung auf Interviewereffekte das Problem, dass Interviewer nicht zufällig zu Befragten zugeteilt sind, sondern dass Interviewer einem oder mehreren Sampling Points zugeordnet sind;

Abbildung 4.3: Anzahl eingesetzte Interviewer, Anzahl Interviews pro Interviewer, 1985-2005

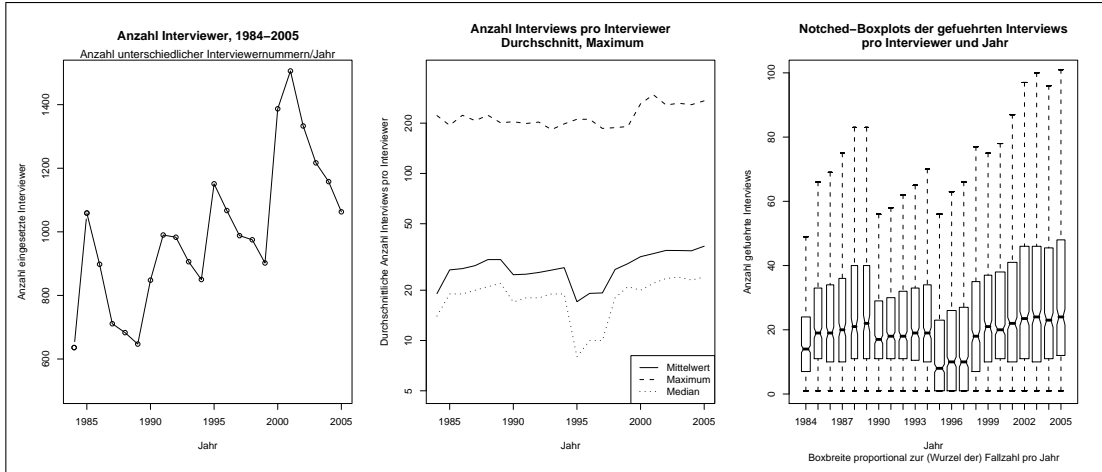
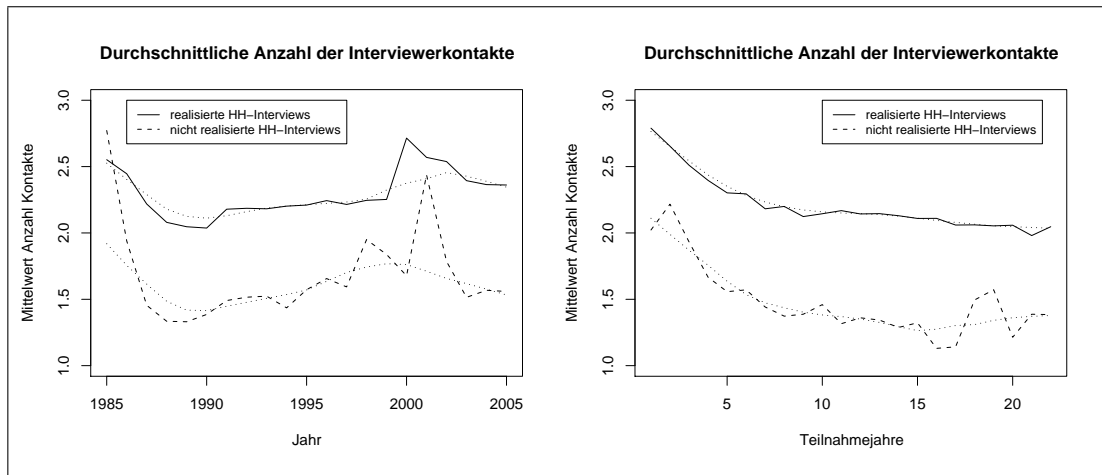


Abbildung 4.3 zeigt die Entwicklung der Anzahl der eingesetzten Interviewer¹⁰⁹ und die Anzahl der geführten Interviews. In der linken Abbildung ist deutlich die Zunahme der eingesetzten Interviewer zu sehen; startete das SOEP mit mehr als 600 Interviewern im Jahr 1984, hat sich die Anzahl im darauffolgenden Jahr auf über 1000 erhöht. Danach sinkt die Anzahl wieder ab, um mit und nach jeder Erweiterung des SOEP um eine Substichprobe wiederum lassen sich ihre Effekte zunächst nicht isoliert von anderen Effekten betrachten (vgl. O’Muircheartaigh und Campanelli, 1999, S. 438). Zudem besteht das Problem der Selbstselektion: besonders fähige Interviewer erhalten bspw. zur Nachbearbeitung die besonders schwer zu kontaktierenden und zur Teilnahme zu bewegendem Fälle, oder besonders schwere Fälle werden zur Nachbearbeitung nur von wenigen einzelnen Interviewer bearbeitet (vgl. auch Zabel, 1998, S. 493).

¹⁰⁹Genau genommen ist in den Abbildungen lediglich die Zahl der unterschiedlichen Interviewernummern erfasst; Interviewernummern sind unglücklicherweise nicht immer im selben Datensatz gespeichert: In der ersten Welle des SOEP 1984 ist die Interviewernummer des Personeninterviews im Datensatz APBRUTTO gespeichert, in allen übrigen Wellen für realisierte Interviews im Netto-Personendatensatz jeder Welle \$P; für Ausfälle ist kein Eintrag im Nettodatensatz, für sie wurde aus dem jeweiligen Bruttodatensatz \$PBRUTTO jeder Welle die Interviewernummer zugespült; war für eine Person mit einem realisierten Personeninterview keine Interviewernummer im Personendatensatz gespeichert, wurde die Interviewernummer des Haushaltsinterview aus dem Haushaltsdatensatz \$HP zugespült. Einige Interviewernummern wurden aus der Berechnung der durchschnittlichen Anzahl der Interviews pro Interviewer ausgeschlossen, da auf sie ungewöhnlich viele (300 und mehr Interviews pro Welle) entfallen. Dies sind die Interviewernummern 999999, 999991, 888888, 777777, 666662, 666661 und fehlende Interviewernummern (0, -1).

Abbildung 4.4: Durchschnittliche Anzahl Interviewerkontakte bis zum Kontakt, 1985-2005 bzw. Welle 1 bis 22



der anzusteigen. Den höchsten Stand erreicht die Anzahl der Interviewer im Jahr 2002 nach einem starken Anstieg auf rund 1500 Interviewer. Die durchschnittliche Anzahl der geführten Interviews pro Interviewer und Jahr (mittlere Abbildung) ist seit 1984 von durchschnittlich 17 auf über 36 Interviews pro Jahr gestiegen (wobei die Hälfte der Interviewer höchstens 8, zuletzt höchstens 24 Interviews führten); das Maximum der geführten Interviews pro Interviewer und Jahr liegt zwischen 183 und 296 Interviews¹¹⁰. Die Notched-Box-Plots rechts verdeutlichen die Streuung der Anzahl der Interviews pro Jahr und Interviewer¹¹¹; besonders auffällig ist hier (wie in der mittleren Abbildung) das deutliche Absinken der durchschnittlichen Interviewanzahl in den Jahren 1995 bis 1997; gleichzeitig hat hier auch die Zahl der eingesetzten Interviewer (die Boxbreite ist proportionale zur Wurzel der eingesetzten Interviewer) zugenommen¹¹². Aufgrund der deutlichen Streuung der Anzahl geführter Interviews

¹¹⁰12 Interviewer führen mehr als 200 Interviews pro Jahr (einige auch in mehreren Jahren): Interviewer 2674, 33111, 53104, 66745, 77127, 89370, 92649, 121835, 180548, 196908, 211362, 278238, fünf von ihnen (2674, 53104, 77127, 92649, 196908) mehr als 250 Interviews in (mindestens) einem Jahr.

¹¹¹Ausreisser ausserhalb der Whisker und der Outer Fence oberhalb der Box sind nicht enthalten.

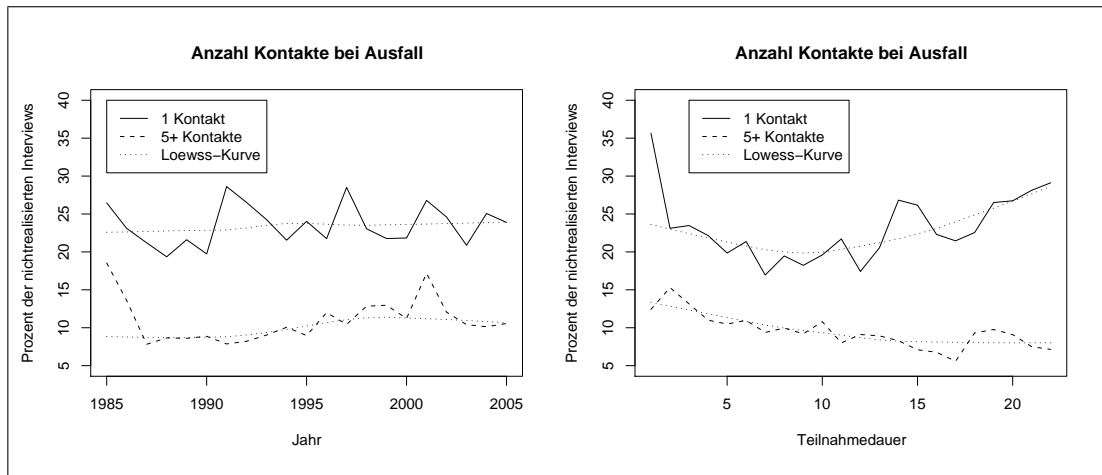
¹¹²In diesen Jahren hat *keine* Erweiterung des SOEP stattgefunden; 1994, also *vor* der Zunahme der Anzahl der Interviewer wurde das SOEP durch die verhältnismäßig kleine Stichprobe der Aus- und übersiedlerhaushalte (vgl. Tabelle 4.1) erweitert.

pro Interviewer wird später auch, zumindest deskriptiv, zu prüfen sein, ob sich Ausfälle und Ausfallarten bei einzelnen Interviewern häufen.

Abbildung 4.4 zeigt den Verlauf der durchschnittlichen Anzahl der Interviewerkontakte mit dem Haushalt¹¹³ bis zu einem realisierten Interview bzw. bis zu einem Ausfall. Deutlich ist in der linken Abbildung der starke Rückgang der durchschnittlichen Anzahl der Kontaktversuche bis zu einem erfolgreichen Kontakt zu sehen; 1985 besteht kaum ein Unterschied zwischen der durchschnittlichen Anzahl Kontaktversuche, die für einen Kontakt nötig sind, der dann allerdings nicht realisiert werden kann und der Anzahl der Kontaktversuche für ein schließlich erfolgreiches Interview: durchschnittlich 2.5 Kontaktversuche sind für erfolgreiche Interviews nötig gewesen, bei nicht realisierten Interviews liegt die durchschnittliche Anzahl Kontaktversuche sogar noch etwas höher. Danach sinkt allerdings die durchschnittliche Anzahl Kontaktversuche für nicht realisierte Interviews stark ab auf rund 1.5 Kontaktversuche, wogegen bei realisierten Interviews die Anzahl der Kontaktversuche bei über zwei liegt. Insgesamt ist in der Abbildung auch zu sehen, dass jeweils zum Zeitpunkt einer neuen SOEP-Substichprobe die Anzahl der Kontaktversuche auch bei nicht realisierten Interviews steigt; hier wiederholt sich der Effekt, der zu Beginn des SOEP zu beobachten war: zu Beginn einer Stichprobe werden mehr Kontaktversuche benötigt, unabhängig davon, ob das Interview dann realisiert werden kann oder nicht; dies bestätigt sich auch, betrachtet man die durchschnittliche Anzahl der Kontaktversuche nach der Teilnahmedauer; beide Kurven im rechten Teil von Abbildung 4.4 fallen mit der Teilnahmedauer deutlich ab. Das ist in Übereinstimmung damit, dass in früheren Wellen Ausfälle noch vermehrt durch Nichterreichbarkeit und Nichtauffinden des Haushalts geschehen; für diese Beobachtungen sind sowohl für dann realisierte als auch nicht realisierte Interviews mehr Kontaktversuche, evtl. auch an mehreren unterschiedlichen Adressen, nötig, in späteren Wellen nach längerer Teilnahme nimmt der Anteil der Verweigerungen und nicht auswert- und durchführbarer Interviews zu; hier sind prinzipiell nicht mehrere Kontaktversuche notwendig, um einen Ausfall festzustellen. Im

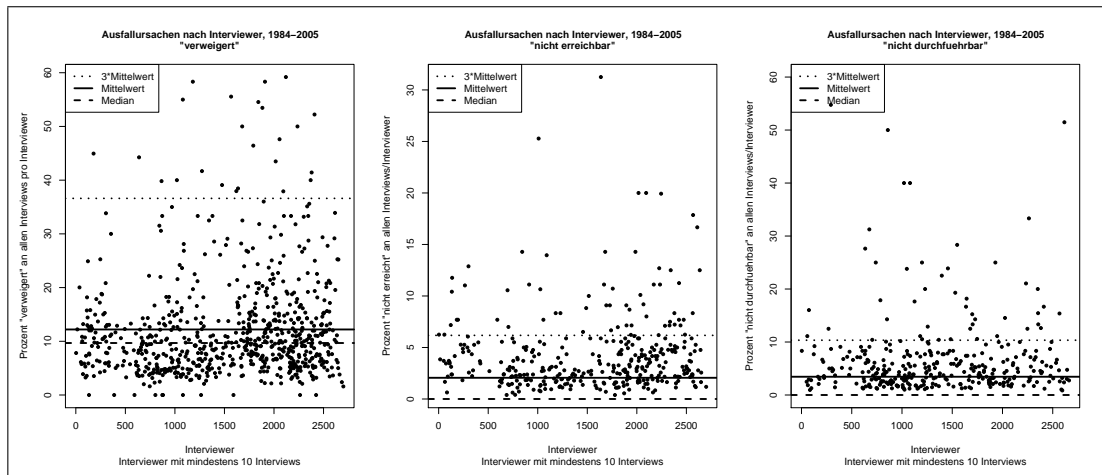
¹¹³Die Anzahl der Interviewerkontakte ist nicht in den Personendatensätzen gespeichert, sondern im Haushaltsdatensatz \$HBRUTTO; die Anzahl der Interviewerkontakte wird dann über die wellenaktuelle Haushaltsnummer den Personen zugespielt.

Abbildung 4.5: Durchschnittliche Anzahl Interviewerkontakte bis zum Ausfall, 1985-2005 bzw. Welle 1 bis 22



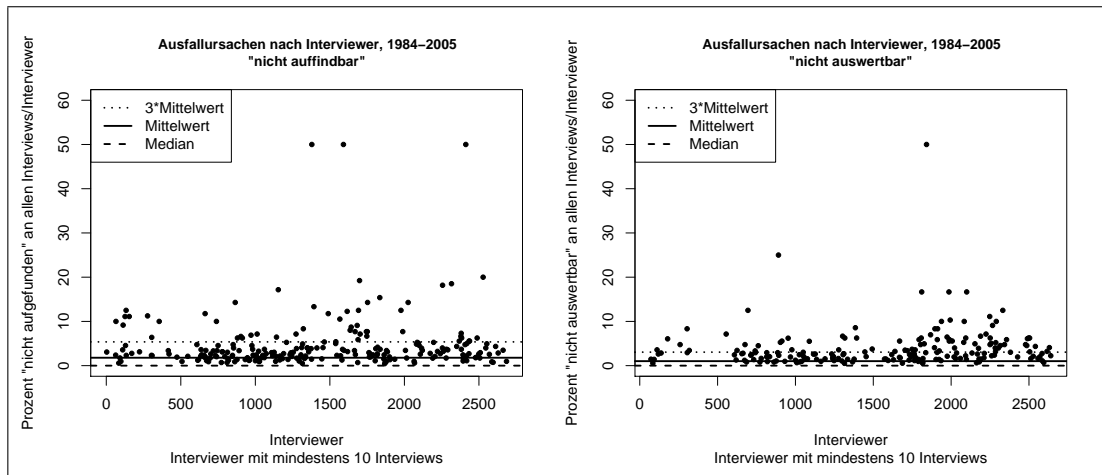
Durchschnitt über alle Jahre werden über 20% der Ausfälle beim ersten Interviewerkontakt festgestellt (siehe linker Teil Abbildung 4.5), allerdings ist auch hier Bewegung in den Jahren nach einer neuen SOEP-Stichprobe feststellbar: hier ist der Anteil der beim ersten Kontakt festgestellten Ausfälle deutlich höher. Der Anteil der Ausfälle, die erst nach 5 oder mehr Interviewerkontakten erfasst werden (können), fällt nach dem Start des SOEP nach der zweiten Welle von über 15% auf unter 10% stark ab; dieser Anteil steigt über die Jahre schließlich auf rund 10% an: über die Jahre ist also von den Interviewern mehr Aufwand betrieben worden, Haushalte zu kontaktieren, besonders auffällig ist der Peak im Jahr 2001. Betrachtet man den Anteil der durchschnittlichen Kontaktversuche für die Ausfälle nach Teilnahmedauer, so sinkt der Anteil der Ausfälle, die erst nach 5 oder mehr Kontaktversuchen festgestellt werden kontinuierlich ab und liegt schließlich bei rd. 10% der Ausfälle; demgegenüber steigt der Anteil der Ausfälle, die schon beim ersten Kontaktversuch durch den Interviewer in der Feldzeit erfasst werden. Auch dies ist in Übereinstimmung mit den Ergebnissen, dass mit zunehmender Teilnahmedauer zunehmend Ausfälle eine Rolle spielen, die auch bei frühen Kontaktversuchen festgestellt werden können (bspw. Verweigerungen, nicht auswertbare Interviews, Teilnahmeunfähigkeit, Ausfall durch Tod) und Ausfälle durch Nichterreichbarkeit und Nichtauffinden des Haushalts zurückgehen. Insgesamt ist also feststellbar, dass die Zahl der Interviewer im SOEP stark gestiegen ist, allerdings auch Fluktuationen unterworfen

Abbildung 4.6: Anteile der Ausfälle an allen Interviews pro Interviewer (verweigert, nicht auffindbar, nicht durchführbar)



ist und nach einer neuen Substichprobe schnell wieder zurückgeht. Auffällig ist die starke Zunahme im Jahr 2000 und 2001 auf über 1400 Interviewer; diese Zunahme hat aber weniger zu einem verringerten Workload für die einzelnen Interviewer geführt, als zu einer vermehrten Anstrengung, auch zu schwer zu kontaktierenden Haushalten schließlich einen Kontakt herzustellen. Betrachtet man den Anteil, den die einzelnen Ausfälle an allen Interviews eines Interviewers ausmachen, zeigt sich, dass sich einzelne Ausfallarten deutlich bei einigen Interviewer konzentrieren. Abbildung 4.6 zeigt den Anteil der Verweigerungen (linke Abbildung), den Anteil Nichterreichbarer (Mitte) und der Ausfälle aufgrund einer Teilnahmeunfähigkeit an allen Interviews eines Interviewers. 22 Interviewer haben einen Anteil Verweigerungen an allen ihren Interviews, der über dem dreifachen des Durchschnitts (ca. 11% aller geführten Interviews) über alle Interviewer liegt; hier handelt es sich allerdings bei allen um Interviewer, die insgesamt nur sehr wenige Interviews führen, deutlich weniger als der Durchschnitt aller Interviewer. Der durchschnittliche Anteil für Ausfälle durch Nichterreichbarkeit pro Interviewer liegt bei rund 2%; 52 Interviewer haben mit rund 7% einen Anteil „nicht erreichbar“ an allen von ihnen geführte Interviews von mehr als dem dreifachen des Durchschnitts. Auch diese Interviewer führen im Schnitt weniger Interviews als im Durchschnitt aller Interviewer. Für 40 Interviewer liegt der Anteil nicht durchführbarer Interviews über dem dreifachen des Durchschnitts aller Interviewer (rechte Abbildung), auch sie

Abbildung 4.7: Anteile der Ausfälle an allen Interviews pro Interviewer (nicht auffindbar, nicht auswertbar)



führen mit durchschnittlich 16 Interviews verhältnismäßig wenige Interviews verglichen mit allen Interviewern. 46 Interviewer haben einen dreifach erhöhten Anteil nicht auffindbarer Befragungspersonen, 32 einen dreifach erhöhten Anteil nicht auswertbarer Interviews¹¹⁴(vgl. Abbildung 4.7); sie führen allerdings mit rund 20 Interviews im Durchschnitt kaum weniger Interviews als der Durchschnitt aller Interviewer. Der hohe Anteil einzelner Ausfallarten gerade für Interviewer mit wenigen Interviews könnte ein Hinweis auf die von O'Brien et al. (2002, S. 2502) geäußerte Vermutung sein, dass unerfahrene Interviewer Ausfälle zu schnell feststellen.

5 Datenstruktur und -aufbereitung

Die Datenstruktur des SOEP ist vergleichsweise komplex: das SOEP verbindet Mikrodaten auf Haushaltsebene mit Mikrodaten auf Individualebene im Haushaltskontext, zudem werden Querschnittsinformationen mit Längsschnittdaten verknüpft; die komplexe Struktur und Trennung von Haushalts- und Individualdaten als auch Querschnitts- und Längsschnittdaten macht eine Aufbereitung der Daten notwendig und aufwendig, besonders auch

¹¹⁴Vor allem 4 Interviewer fallen mit einem Anteil von rd. 15% und mehr nicht auswertbarer Interviews auf: Interviewernummern 105309 (45 Interviews), 201960 (11), 206270 (11) und 229717 (14 Interviews).

Abbildung 5.1: Datenstruktur des SOEP, jährliche Querschnittsdatsätze

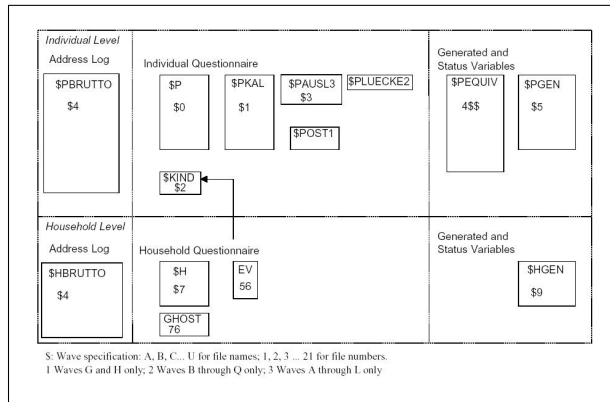
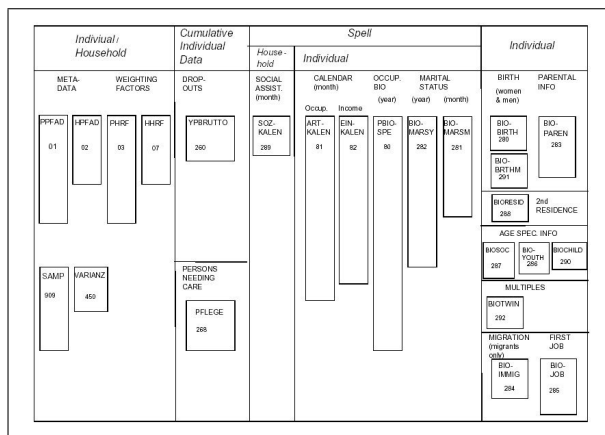


Abbildung 5.2: Datenstruktur des SOEP, Längsschnittsdatsätze



Quelle: Haisken DeNew und Frick (2005, S. 30 u. 32)

die Aufbereitung der Information zu den Paneldatensätzen. Das SOEP besteht aus mehreren Datensätzen, aufgeteilt in jährliche Personen- und Haushaltsdatensätze und Datensätze für Haushalte und Personen im Längsschnitt. Abbildung 5.1 zeigt die Querschnittsstruktur, Abbildung 5.2 die Längsschnittstruktur des SOEP; eine Aufbereitung der Ausfallursachen der Personen aus dem SOEP wird dadurch erschwert, da Informationen zu Ausfällen aus dem SOEP sowohl auf Haushalts- und Personenebene vorliegen und um Informationen aus Längsschnittsdatsätzen ergänzt werden müssen.

5.1 Ausfallursachen im SOEP

Im SOEP werden die Ausfallursachen der Haushalte und Personen im Kontaktprotokoll vom Interviewer erfasst und im Bruttodatsatz der Personen und Haushalte gespeichert. Insgesamt werden 29 verschiedene Zustände des Befragungsstatus erfasst und gespeichert; Tabelle 5.1 zeigt diese Informationen aus der Variable \$pergz der Bruttodatsätze \$pbrutto. Sie enthält den Code, unter dem ein bestimmter Befragungsstatus im Datensatz gespeichert ist, die Bezeichnung des Codes und zusätzlich unterschiedliche alternative Bezeichnungen bzw. Änderungen der Bezeichnung über die Wellen hinweg; schließlich sind Umkodierungen beschrieben, die notwendig wurden, da die Kodierung des Befragungszustandes über die

verschiedenen Wellen hinweg nicht konsistent erfolgt ist; zum Beispiel ändert sich mit der Welle 10 die Kodierung einer endgültigen Verweigerung von Code [40] zu Code [46]; um eine konsistente Kodierung über alle Wellen sicherzustellen wurden daher Umkodierungen vorgenommen. Einige Ausfallursachen werden mehrmals aufgeführt, bspw. die Nummern 11 und 14 („in Feldzeit nicht erreicht“) sowie die Nummern 16 und 21 („sonstige unklarer Fall“). Zudem beschreiben mehrere Codes den gleichen Befragungsstatus, bspw. die Nummern 2 bis 8 (realisierte Interviews), die Nummern 15, 23 und 25 (nicht auswertbar bzw. verwertbar) sowie die Nummern 28 und 29 (nicht auffindbare Adressen bzw. Haushalte). Insgesamt lassen sich die detaillierten Ausfallgründe in die in Abschnitt 1.4 und 1.5 ab Seite 25 beschriebenen Ausfallursachen einordnen: Während die Ausfälle mit Nummern 1 bis 8 eine erfolgreiche Teilnahme in der aktuellen Welle bezeichnen, ergeben Ausfallgründe mit den Nummern 9 bis 21 einen temporären Ausfall und zerstören ein monotones Teilnahmemuster (siehe Abschnitt 1.5.3 ab Seite 38). Ausfälle mit den Nummern 22 bis 29 sind endgültig und absorbierend. Bei Ausfällen durch einen Wegzug ins Ausland liegt zumindest das Jahr des Wegzugs als auswertbare Information vor; bei Ausfällen durch Tod liegt zumindest die Information „Todesjahr“ vor; während Kinder unter 17 Jahren in den Bruttodatensätzen \$PBRUTTO (und PPFAD) vermerkt sind, zählen sie nicht als Befragungsperson und sind dementsprechend nicht in den Nettodatensätzen \$P der eigentlichen Befragung enthalten; sie werden daher aus der weiteren Analyse ausgeschlossen¹¹⁵. Die Zellenbesetzung der 29 erfassten Ausfallursachen zeigt über die Wellen starke Variation, ausserdem kann davon ausgegangen werden, dass Interviewer kaum zwischen den einzelnen, ähnlichen Ausfallursachen unterscheiden; daher

¹¹⁵Dabei zeigen sich Inkonsistenzen im SOEP-Datensatz: Im Bruttodatensatz \$PBRUTTO sind Kinder unter 17 Jahren in der Variable \$NETTO als Kinder gekennzeichnet, bildet man allerdings eine Kontrolle aus dem Geburtsjahr (aus den \$P-Datensätzen bzw. aus dem PPFAD-Datensatz) und dem Befragungsjahr, so zeigt sich, dass 189 Kinder unter 17 Jahren im Kontaktprotokoll als Befragungspersonen geführt werden, als auch 8 Personen als Kinder im Nettodatensatz geführt werden, die zum Zeitpunkt der Erhebung älter als 17 Jahre alt sind: Schließlich können über den Eintrag im Kontaktprotokoll, über das Geburtsjahr und das Jahr der Erhebung 5.598 Personen unterhalb des Befragungsalters identifiziert werden, denkbar sind hier bspw. Übertragungsfehler. Diese Beobachtungen werden von einer weiteren Aufbereitung und Analyse ausgeschlossen. Zum Teil kehren sie jedoch als Befragungspersonen mit dem Überschreiten des Erstbefragungsalters in das SOEP zurück. Für 1008 Beobachtungen liegen zudem keine Informationen zum Geburtsjahr vor.

Tabelle 5.1: Ausfallursachen im SOEP, Variable \$PERGZ („Zweisteller“) aus dem Datensatz \$PBRUTTO

Nr.	Code	Wert	Variante, Bemerkung
1	[0]	HHwechsel	
2	[10]	realisiert	Interview diese Welle realisiert; erstmals realisiert
3	[11]	wiederholt realisiert	zum 2.mal realisiert; Interview dieser und voriger Welle realisiert
4	[13]	realisiert, HH-Lücke Vorjahr bearbeitet	Interview, HH-Lücke Vorjahr geschlossen
5	[14]	realisiert, HH-Lücke Vorjahr NICHT bearbeitet	Interview, HH-Lücke Vorjahr bleibt offen
6	[16]	realisiert, P-Lücke Vorjahr bearbeitet	Interview, Personen-Lücke Vorjahr geschlossen; Lücke Vorjahr nacherhoben
7	[17]	realisiert, P-Lücke Vorjahr NICHT bearbeitet	Interview, Personen-Lücke Vorjahr bleibt offen; Lücke Vorjahr nicht nacherhoben
8	[18]	realisiertes Pi, kein HHi	Personen-Interview, kein HH-Interview
9	[20]	derzeit nicht durchführbar	
10	[21]	alt und krank	
11	[22]	in Feldzeit nicht erreicht	bis Ende Feldarbeit nicht erreicht
12	[23]	Ausländer lang in Heimat	Ausländer lange Zeit in Heimat
13	[24]	in Krankenhaus über Feldzeit	krank bis Feldende
14	[25]	in Feldzeit nicht erreicht	nicht erreichbar; umkodiert auf [22]
15	[26]	nicht auswertbar	
16	[29]	sonstiger unklarer Fall	sonstige unklare Fälle
17	[30]	derzeit nicht teilnahmebereit	
18	[31]	Fragebogen nicht ausgefüllt	kein Fragebogen
19	[32]	vielleicht nächstes Jahr	nur in Welle 2, sonst siehe [33]
20	[33]	keine Zeit - keine Lust	ab Welle 3 mit Code [32], umkodiert von [32] auf [33]
21	[39]	sonstige unklare Fälle	umkodiert auf [29]
22	[40]	endgültig verweigert	endgültige Verweigerung
23	[46]	nicht auswertbar	Frabo nicht auswertbar; umkodiert zu [26]
24	[46]	endgültig verweigert	ab Welle 10; umkodiert zu [40]
25	[46]	endgültig nicht verwertbar	Welle 12; umkodiert auf [26]
26	[50]	ins Ausland verzogen	
27	[60]	verstorben	
28	[80]	in Feldzeit nicht gefunden	in Feldzeit nicht auffindbar
29	[90]	Adresse Haushalt nicht auffindbar	ab W12: endgültig nicht auffindbar

Analog zu Spalten 1-3 sind Ausfallursachen im Längsschnittdatensatz YPBRUTTO Variable YPERGZ aufbereitet und kodiert

ist eine Zusammenfassung der ausführlichen und doppelten Ausfallkategorien notwendig. Im folgenden werden die Nummern 1 bis 8 aus Tabelle 5.1 als „realisiert (0)“ zusammengefasst,

die Nummern 28 und 29 als „nicht auffindbar (1)“, Nummern 11 und 14 als „nicht erreichbar (2)“, Nummern 30 bis 33 sowie Nummer 40 als „verweigert (3)“. Als „nicht auswertbar (4)“ werden die Nummern 15 und 23 zusammengefasst, Nummer 10 wird als Kategorie 5 „alt und krank (5)“ beibehalten, Nummern 9 und 13 werden als „nicht durchführbar (6)“ zusammengefasst. Ausfälle durch Tod werden in Kategorie 7 gefasst, Ausfälle durch Wegzug ins Ausland (Nummern 12 und 26) als achte Kategorie. Neutrale Ausfälle bilden eine eigene, neunte Kategorie. Ausfälle mit unbekannter oder unklarer Ursache (Nummern 16 und 21 sowie fehlende Angaben zum Ausfall) bilden eine zehnte Kategorie.

Grundsätzlich sind Ausfallinformation in den jährlichen Bruttodatensätzen $\$PBRUTTO$ und $\$HBRUTTO$ ¹¹⁶ in den Kontaktprotokollen der Interviewer gespeichert; in diesen Kontaktprotokollen verzeichnet der Interviewer, ob der Haushalt aufgefunden und Kontaktpersonen im Haushalt erreicht wurden; zudem werden hier Ausfälle der Kontaktpersonen verzeichnet, sofern sie erreichbar sind. Zusätzlich sind Kontaktinformation in den Längsschnittdatensätzen $PPFAD$ ¹¹⁷ und $YPBRUTTO$ enthalten. Während $PPFAD$ vor allem zur Identifizierung zu einer bestimmten Welle aktueller Befragungspersonen¹¹⁸ von zentraler Bedeutung ist, müssen Informationen für Ausfälle aus dem Panel aus den Datensätzen $\$PBRUTTO$ und $YPBRUTTO$ kombiniert werden, um vollständige und valide Informationen über Ausfallursachen zu erhalten; während in $\$PBRUTTO$ lediglich wellenaktuelle Ausfallursachen enthält, können aus $YPBRUTTO$ aktuellere Informationen zu einem Ausfall in einer früheren Welle vorliegen; das liegt darin begründet, dass zum Zeitpunkt des Interviewerbesuchs der Ausfall vom Interviewer festgestellt wird, bspw. wenn eine Person (trotz mehrere Kontaktversuche) nicht erreichbar ist und vom Interviewer dementsprechend als „nicht erreichbar“ im Kontaktprotokoll vermerkt

¹¹⁶ $\$$ bezeichnet den Platzhalter für das Jahr der Erhebung bzw. der Welle; 1985 lautet der Datensatzname entsprechend $BPBRUTTO$, 2005 entsprechend $VPBRUTTO$; für die Nettodatensätze $\$P$ und die Haushaltsdatensätze gilt entsprechendes.

¹¹⁷Der Datensatz $PPFAD$ bildet das „Skelett“ des SOEP; in ihm sind Basis-Informationen aller im SOEP erfassten (und jemals im Rahmen des SOEP kontaktierten) Personen erfasst; entsprechend der Basis-Datensatz $HPFAD$ der Haushalte.

¹¹⁸Die Variable $\$netto$ bezeichnet alle Personen entweder als realisiertes Element der aktuellen Welle $\$$, als aktueller Ausfall in Welle $\$$ oder als Kind unterhalb der Altersgrenze zur erstmaligen Befragung im Haushalt.

Tabelle 5.3: Aufbereitung der Ausfallursachen im SOEP: Datensätze und Variablen

PPFAD	\$PBRUTTO	YPBRUTTO
- Längsschnittinformation	- jährliche Information	- Längsschnittinformation
- \$NETTO (Netto- vs. Bruttobestand)	- \$PERG(Z) (Ausfallursachen)	- YPERG(Z) (endgültige Ausfallursache)
- \$HHNR (wellenaktuelle Haushaltsnummer)	- \$HHNR (wellenaktuelle Haushaltsnummer)	- YPZUG (HH-Zugehörigkeit, Tod, Umzug)
- GEBJAHR (Geburtsjahr)	- \$GEBJAHR (Geburtsjahr)	
- TODJAHR, TODINFO (Todesjahr, Informationsquelle)	- \$STISTAT ((Nicht-)Stammperson)	
Ausschluss von:		Aktualisierung Ausfälle durch:
- Kinder vor erster Befragung	- Kinder vor erster Befragung	- Ausfall durch Tod
- Beobachtungen nach Ausfall	- Beobachtungen nach Ausfall	- Ausfall durch Wegzug ins Ausland
		- Auszug Nichtstammpersonen (bis 1990)

wird. Die alleinige Verwendung wellenaktueller Ausfallinformationen ist allerdings problematisch, da sich hinter den vom Interviewer erfassten Ausfallursachen andere Ausfälle verbergen können; bspw. ist ein Krankenhausaufenthalt oder der Tod einer Befragungsperson ohne weitere Information nicht unbedingt vom Interviewer als solcher feststellbar; ohne eine Überprüfung und Aktualisierung der Ausfallursachen sind somit Vermischungen von Ursachen möglich und damit auch eine Vermischung der Effekte verschiedener Ausfallursachen. Im SOEP sind die aktualisierten Ausfallursachen im Längsschnittdatensatz YPBRUTTO in den Variablen YPERG bzw. YPERGZ¹¹⁹ erfasst, die Variable YPZUG enthält Informationen zum Wegzug ins Ausland oder Tod der Befragungsperson. In der Aufbereitung der Ausfallursachen der Personen werden also Informationen aus den Datensätzen PPFAD, \$PBRUTTO und YPBRUTTO verwendet (siehe Tabelle 5.3). Grundsätzlich gilt, dass für Beobachtungen, die in PPFAD als nicht zum Nettobestand einer Welle gekennzeichnet sind Ausfallursachen in \$PBRUTTO vorliegen; diese Angaben in den Kontaktprotokollen werden durch Angaben aus dem Datensatz YPBRUTTO aktualisiert, sofern eine Beobachtung in YPBRUTTO als „verstorben“ erfasst ist; zudem werden Ausfallursachen in \$PBRUTTO durch Angaben in YPBRUTTO aktualisiert,

¹¹⁹Die Variable YPERG enthält zusammengefasste, einstellig codierte Ausfallursachen der zweistellig codierten Ausfallursachen der Variable YPERGZ („Zweisteller“).

Tabelle 5.4: Änderung der Ausfallursachen in \$PBRUTTO durch Angaben in YPBRUTTO

	Ausfall durch	\$PBRUTTO ¹	YPBRUTTO ²	aktualisiert ³	aktualisiert in%
[-99]	keine Information	1.590	1.911	2	0.0
[-9]	Kinder u.16 Jahre			5,598	1.5
[0]	realisiert	341.752	–	337,780	87.7
[1]	nicht erreichbar	1.235	1	1,236	0.3
[2]	nicht auffindbar	1.156	36	1,192	0.3
[3]	verweigert/nicht bereit	31.444	25	31,466	8.2
[4]	nicht auswertbar	483	1	484	0.1
[5]	alt und krank	22	–	22	0.0
[6]	n.durchfuehrbar/krank	2.000	–	2,000	0.5
[7]	verstorben	6	2.446	2,452	0.6
[8]	ins Ausland verzogen	52	1.546	1,637	0.4
[9]	neutraler Ausfall ⁴	–	157	157	0.0
[10]	unbekannt	1.050	5.846	1,050	0.3
	Fehlende Angaben	4.286	373.107	–	0.0
	Insgesamt	385.076	385.076	385.076	100

¹: Variable \$PERGZ in \$PBRUTTO ²:Variable YPERGZ in YPBRUTTO

³: Angaben in \$PPBRUTTO, aktualisiert durch YPBRUTTO

⁴: bis 1990: Nicht-Stammperson verzogen, keine Weiterverfolgung

Kategorien 5 und 6 werden zusammengefasst in Kategorie 5

Beobachtungen mit Code -99 werden Kategorie 10 zugeschrieben

wenn in \$PBRUTTO keine, und in YPBRUTTO Angaben zu einem Ausfall verzeichnet sind¹²⁰. Tabelle 5.4 zeigt die Aufbereitung der Ausfallursachen anhand der verschiedenen Datensätze. Nach Aufbereitung bleiben 2 Beobachtungen, für die keine Angabe zum Ausfall vorliegen. Insgesamt werden durch Angaben zum Ausfall in YPBRUTTO 4.286 von 385.076 Beobachtungen aktualisiert. Ausnahmslos sind das Fälle, in denen in \$PBRUTTO *keine* Information zum Ausfallgrund vorliegen. YPBRUTTO enthält vor allem aktualisierte Informationen zu (auch ins Ausland) verzogenen oder verstorbenen Personen. Die dritte und vierte Spalte in Tabelle 5.4 enthält das Ergebnis dieser Datenaufbereitung, die in der folgenden Analyse als abhängige und unabhängige Variable verwendet wird. Insgesamt liegen im SOEP über die Jahre

¹²⁰Das heisst, eine änderungen der Ausfallursache wird nur für Beobachtungen vorgenommen, die in PPFAD und \$PBRUTTO als Ausfall erfasst sind und für die in YPBRUTTO ein Ausfallergebnis eingetragen ist. Umgekehrt werden Ausfallursachen in \$PBRUTTO nicht durch unbekannte Ausfallursachen in YPBRUTTO ersetzt oder als realisiert gekennzeichnet.

385.076 Beobachtungen vor. 5.598 Beobachtungen werden im Weiteren ausgeschlossen, weil sie noch unter dem Erstbefragungsalter des SOEP von mindestens 16 Jahren liegen (bzw. als solche im Kontaktprotokoll verzeichnet sind). überwiegend handelt es sich um realisierte Personeninterviews, 31.000 Verweigerungen machen mit 8% der Beobachtungen insgesamt die häufigste Ausfallursache aus dem SOEP aus; 2.452 Beobachtungen sind als „tot“ im SOEP nachgewiesen; die Kategorien 5 und 6 werden aufgrund der niedrigen Fallzahl von Kategorie 5 zusammengefasst, sodass insgesamt 2.022 Ausfälle durch eine Teilnahmeunfähigkeit erfasst sind. Insgesamt sind 1.637 Umzüge ins Ausland im SOEP dokumentiert, 1.192 Beobachtungen konnten in der folgenden Welle nicht mehr aufgefunden, 1.236 nicht erreicht werden. 484 Interviews wurden als nicht auswertbar gekennzeichnet.

6 Entwicklung der Ausfälle im SOEP

Panelausfälle sind, wie bereits erwähnt, von der Struktur und vom Ausmaß nicht mit Ausfällen bei Querschnittserhebungen oder in der ersten Welle eines Panels zu vergleichen; in späteren Wellen eines Panels sind Teilnehmeraten höher als in Querschnittserhebungen oder in der ersten Welle eines Panels¹²¹; nach einer erfolgreichen ersten Welle sind Teilnehmer mit Interviewer, Befragungsprozedur, Auftraggeber und Thema der Befragung vertraut, so dass Unsicherheiten auf Seiten der Befragten im Vergleich zu Querschnittserhebungen bei Kontakt in der nächsten Welle reduziert sind; zudem kann in den Folgewellen des Panels eine Kontaktwahrscheinlichkeit erhöht werden, bspw. durch Kontakt zwischen Befragungsinstitut und Zielhaushalten zwischen den Feldzeiten, so dass bspw. Aus- und Umzüge schon vor Beginn der neuen Feldzeit berücksichtigt werden können oder aber auch Kontaktzeiten an die Zielhaushalte angepasst werden können. Daher werden in den Folgewellen eines Panel die Teilnehmeraten deutlich höher sein als in der ersten Welle bzw. in Querschnittserhebungen

¹²¹Atrostic et al. (2001, S. 219f) erwarten bei den folgenden Wellen eines Panels steigende Nonresponseraten und begründen dies allein mit der zunehmenden Belastung der Befragten durch die wiederholte Teilnahme. Weiter bemerken sie, dass sich vor allem die Zusammensetzung des Nonresponse in den folgenden Wellen ändert und zunehmend Ausfälle aufgrund von Verweigerungen stattfinden. Dem steht die Vermutung von Laurie et al. (1999, S. 278) entgegen, dass das Panel mit der Zeit immer kooperativer wird.

und auf diesem Niveau auch vergleichsweise stabil verlaufen.

Abbildung 6.1 zeigt die Entwicklung der Ausfälle der Personen seit Beginn des SOEP 1984; per Definition gibt es im ersten Jahr des SOEP¹²² keine (Panel-) Ausfälle; die Regressionsgerade zeigt einen leichten Anstieg von rund 0.2% pro Jahr. Einige Jahre sind auffällig durch einen vergleichsweise hohen Anteil Ausfälle (1985, 1992/93, 2001) bzw. vergleichsweise niedrigen Anteil Ausfälle (v.a. 1990 und 2000). Letztere sind dadurch zu erklären, dass in den Jahren 1990 und 2000 große, neue Substichproben das SOEP ergänzen (Stichproben C und F); in diesen Substichproben gibt es im Startjahr per Definition keine Panelausfälle, dadurch erhöht sich bei der Berechnung der Nonresponserate lediglich der Nenner deutlich, wogegen der Zähler vergleichsweise unverändert bleibt. Für die Jahre mit einem besonders hohen Anteil von Ausfällen gilt, dass 1985, 1991, 1999, 2001 die Jahre *nach* der ersten Welle einer neuen Substichprobe sind; offensichtlich sind Ausfälle kurz nach Eintritt in das Panel häufiger als in späteren Wellen. Deutlich ist der Unterschied zwischen der Regressionsgerade und dem Lowess-Smoother zu erkennen, die geglättete Linie zeigt eine u-förmige Entwicklung: bis in die Mitte der 90er Jahre sinkt der Anteil der Ausfälle, um dann anschließend langsam wieder auf rund 12% jährlich anzusteigen.

Die höheren Ausfallraten nach Start einer Stichprobe werden deutlicher, betrachtet man nicht die Ausfälle nach (Kalender-) Jahren, sondern plottet den Anteil der Ausfälle gegen die Teilnahmedauer: in Abbildung 6.2 ist deutlich der starke Anstieg der Ausfälle zum zweiten Teilnahmehjahr¹²³ und der folgende Rückgang in den folgenden Wellen auf einen stabilen Anteil von unter 10% Ausfälle. Damit lässt sich festhalten, dass Panelausfälle aus dem SOEP vor allem in den ersten Wellen nach Einstieg stattfinden, mit zunehmender Teilnahmedauer sinkt die Wahrscheinlichkeit, aus dem Panel auszufallen stark ab, um sich dann bei durchschnittlich unter 10% Ausfälle auch für ältere Panelteilnehmer zu stabilisieren. Während sich der jährliche Anteil der Ausfälle im SOEP auf diesem Niveau stabilisiert, berichtet

¹²²Nicht nur im ersten Jahr, sondern genauer in der ersten Welle der Stichprobe A und B im Jahr 1984; grundsätzlich gilt das auch für alle späteren neuen Substichproben des SOEP.

¹²³Bei der Betrachtung nach Teilnahmehjahren gilt nicht, dass es im ersten Jahr keine Ausfälle gibt. Neuzugänge in das Panel (durch das Erreichen des Befragungsalters oder Zuzüge in Haushalte) können im ersten Teilnahmehjahr durch Nonresponse ausfallen.

Abbildung 6.1: Entwicklung der Ausfälle im SOEP, 1985-2005, $\beta = 0.18$

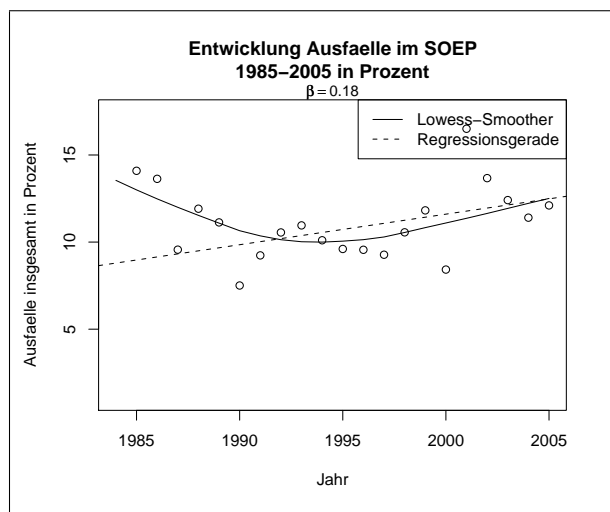
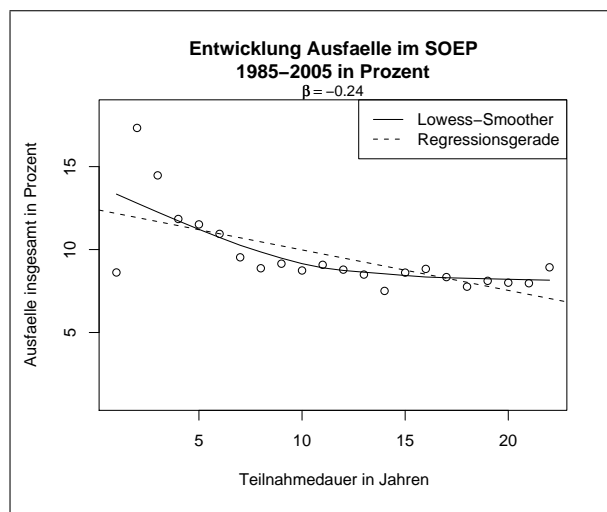


Abbildung 6.2: Entwicklung der Ausfälle im SOEP, nach Teilnahmejahren, $\beta = -0.24$



bspw Zabel (1998) für die amerikanischen SIPP und PSID auch Rückgänge der Anteile der Ausfälle, jedoch verlaufen bei beiden die Anteile der Ausfälle auf deutlich niedrigerem Niveau bei unter 3% pro Jahr¹²⁴. Auch Fitzgerald et al. (1998, S. 253f) berichten für die PSID für die erste Welle eine Nonresponserate von 12% und anschliessend „generally about 2.5 or 3 percent annually“. Auch Jones et al. (2006, S. 549f) berichten diesen Rückgang der Ausfälle nach der ersten Welle für das britische Haushaltspanel BHPS, als auch Nicoletti und Peracchi (2005, S. 768) für das Europäische Haushaltspanel.

Wie in Abschnitt 1 bereits beschrieben, können sich Nichtteilnehmer deutlich voneinander unterscheiden, ebenso kann sich auch die Entwicklung der verschiedenen Ausfallursachen über die Zeit (bzw. die Anzahl der teilgenommenen Wellen) deutlich unterscheiden.

Im Durchschnitt rund 3.2% aller Ausfälle jedes Jahr kommen zustande, weil Interviewer den Haushalt der Zielperson an der alten Adresse nicht mehr auffinden können und eine neue Adresse in der Feldzeit nicht in Erfahrung gebracht werden kann. über die Jahre ist nur ein geringer Rückgang von durchschnittlich -0.2% pro Jahr zu beobachten, deutlich in Abbil-

¹²⁴Während im PSID nach 21 Jahren noch rund die Hälfte der ursprünglichen Stichprobe an der Befragung teilgenommen hat, ist es aufgrund der höheren Ausfallraten im SOEP nach 21 Jahren daher nur noch rund ein Drittel.

Abbildung 6.3: Entwicklung Ausfälle „Nicht auffindbar“, 1985-2005, $\beta = -0.21$

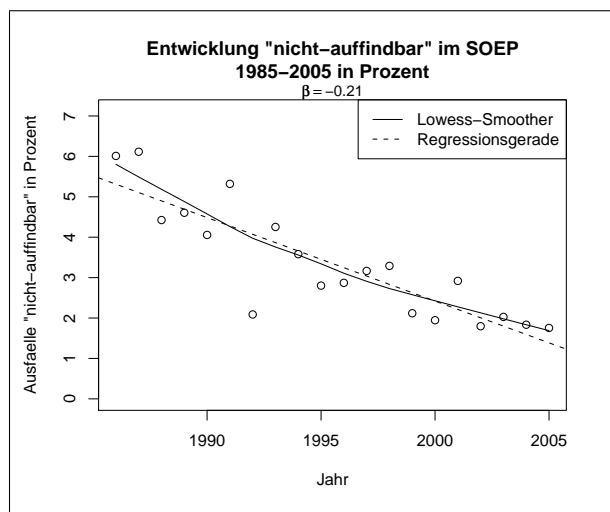
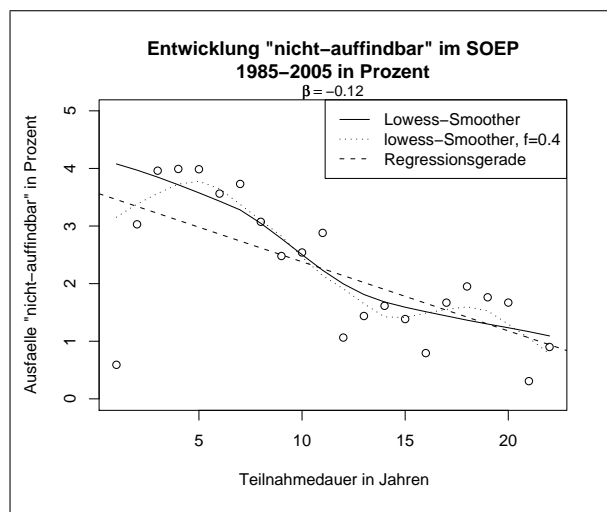


Abbildung 6.4: Ausfälle „Nicht auffindbar“, nach Teilnahmejahren, $\beta = -0.12$



dung 6.3. Während in den ersten Jahren des SOEP der Anteil „Nicht Auffindbar“ noch bei 6% lag, sind es seit dem Jahr 2000 nur noch rund 2%; evtl. ist hier eine veränderte Feldarbeit des Erhebungsinstituts verantwortlich, z.B. mit einer veränderten Nachbearbeitung umgezogener Haushalte und vermehrten Nachforschungen bspw. bei Meldeämtern oder auch einer verlängerten Feldzeit für nicht aufgefundene Haushalte.

Deutlich ist in Abbildung 6.4, dass dieser Rückgang des Anteils an allen Ausfällen auch mit zunehmenden Teilnahmejahren gilt, wenn auch hier der jährliche Rückgang ($\beta = -0.11\%$) nicht so stark ist. Deutlich ist allerdings auch der starke Anstieg zu Beginn auf über 3%. Eine Erklärung für den sinkenden Anteil nicht auffindbarer Zielhaushalte liegt hier im gleichzeitig zunehmenden Alter der Befragungspersonen begründet; während junge Personen besonders mobil sind, nimmt die Mobilität mit dem Alter zunehmend ab, und damit auch die Wahrscheinlichkeit, dass ein Haushalt an der alten Adresse nicht mehr auffindbar ist¹²⁵.

Durchschnittlich 2.1% aller Ausfälle pro Jahr kommen durch Nichterreichbarkeit der Zielhaushalte zustande. Hier ist über die Zeit von 1985 bis 2005 eine Zunahme von 0.1% pro Jahr

¹²⁵Einen Rückgang des Anteils der Ausfälle durch Nichtauffindbarkeit des Haushalts stellen auch Atrostic et al. (2001) für vier große amerikanische Panelstudien fest: „Repeated interviews in panel surveys tend to increase the chance of locating units [...] after the first wave“.

Abbildung 6.5: Entwicklung Ausfälle „Nicht erreichbar“, 1985-2005, $\beta = 0.11$

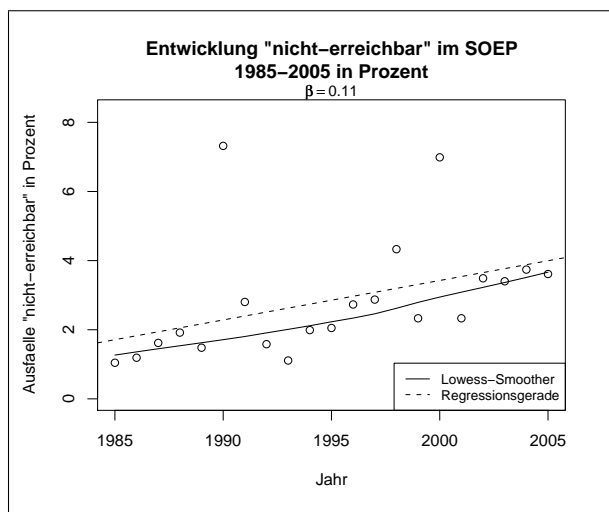
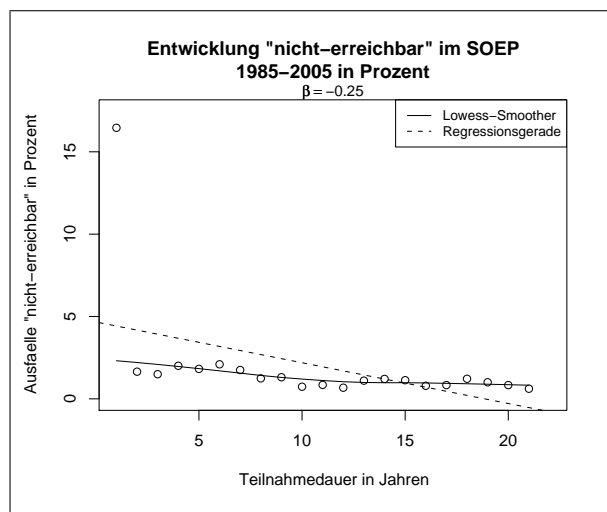


Abbildung 6.6: Ausfälle „Nicht erreichbar“, nach Teilnahmejahren, $\beta = -0.25$



zu beobachten (siehe Abbildung 6.5). Auffällig sind die Jahre 1990, 1998 und 2000 mit einem deutlich höheren Anteil nichterreichbarer Personen. 1990 wurde das SOEP um Stichprobe C in der (damals noch) DDR erweitert, 1998 folgte die Erweiterung um Stichprobe E und 2000 schließlich um Stichprobe F; eventuell ist hier eine höhere Anzahl von zu führenden Interviews pro Interviewer mit einer verringerten Anstrengung und einer verringerten Anzahl von Kontaktversuchen durch die Interviewer erkaufte worden (vgl. Groves, 2006, S. 666).

Betrachtet man die Entwicklung des Anteils nichterreichbarer Personen nach Teilnahmedauer (Abbildung 6.6), so ist hier der Anteil an allen Ausfällen über die Teilnahmedauer vergleichsweise konstant und stabil bei rund 1% aller Ausfälle; lediglich das erste Jahr der Teilnahme ist als deutlicher Ausreißer mit über 18% nichterreichbarer Personen an allen Ausfällen deutlich¹²⁶. Offensichtlich gelingt es in den Folgewellen den Interviewern, ihre Kontaktzeiten mit den Zielhaushalten abzusprechen oder entsprechend ihrer Kenntnis von Merkmalen der Haushalte, die eine erfolgreiche Kontaktzeit versprechen anzupassen. Der steigende Anteil von Ausfällen nicht erreichbarer Personen entspricht der in der Literatur häufig mit Ausfällen verbundenen Vermutung, dass Ausfälle durch Nichterreichbarkeit vor

¹²⁶Deutlich ist zu sehen, wie die Regressionsgerade durch diesen Ausreißer bestimmt ist, während der Lowess-Smoother nahezu waagrecht verläuft.

Abbildung 6.7: Entwicklung Ausfälle „Nicht durchführbar“, 1985-2005, $\beta = -0.57$

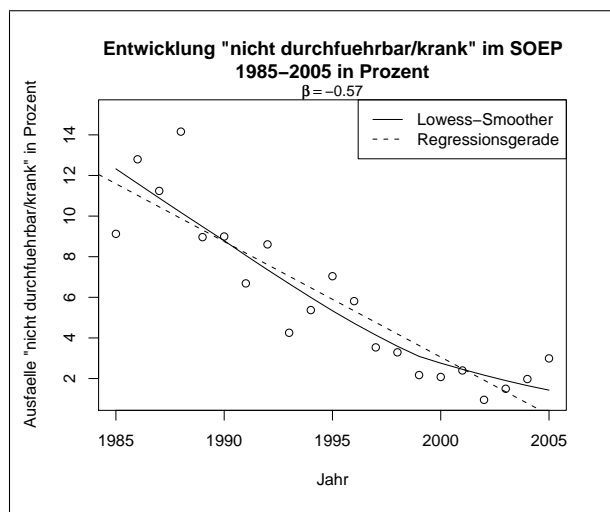
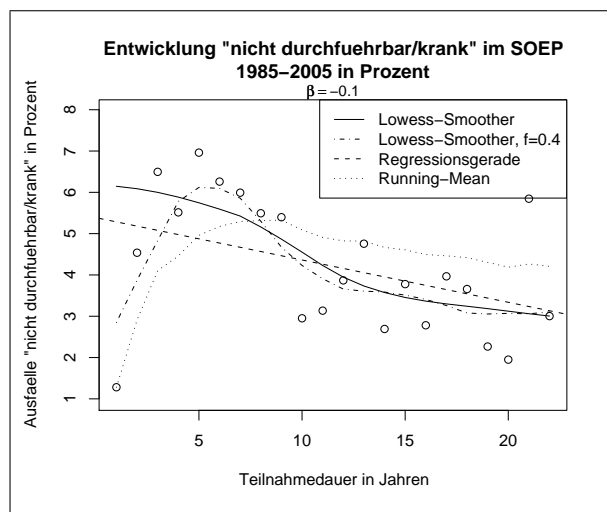


Abbildung 6.8: Ausfälle „Nicht durchführbar“, nach Teilnahmejahren, $\beta = -0.1$



allem mit der Veränderung der Haushaltstruktur und einem erhöhten Anteil vergleichsweise kleinerer Haushalte bzw. 1-Personenhaushalte zusammenhängt¹²⁷. Zudem ändert sich mit zunehmendem Alter auch das Erreichbarkeitsmuster.

Die Entwicklung der Ausfälle aufgrund von Krankheit oder sonstigen Einschränkungen, die eine Teilnahme unmöglich machen, sind in Abbildungen 6.7 und 6.8 dargestellt. Während über die Jahre ein deutlicher Rückgang des Anteils nicht durchführbarer Interviews (rund 0.6% pro Jahr) beobachtbar ist, ergibt sich ein weniger deutliches Bild, betrachtet man die Entwicklung nach Teilnahmejahren; auch hier ist ein Rückgang bemerkbar, der mit ca. -0.1% pro Jahr aber deutlich schwächer ausfällt. Insgesamt ist diese Entwicklung etwas überraschend, da vermutet werden kann, dass mit zunehmender Teilnahmedauer am Panel, und damit auch zunehmend älteren Befragungspersonen, auch der Anteil der nicht mehr teilnahmefähigen Personen ansteigt; allerdings macht sich hier das Weiterverfolgungskonzept des SOEP bemerkbar: da nicht nur Stammpersonen in der nächsten Welle wieder befragt wer-

¹²⁷Auch im SOEP ist die Veränderung hin zu einer kleineren Haushaltsgröße bemerkbar; der Anteil der Ein-Personenhaushalte steigt kontinuierlich auf bis 35%, der Anteil größerer Haushalte mit 3 oder mehr Personen ist dagegen rückläufig, siehe Abbildung A.1 im Anhang auf Seite 143. Die durchschnittliche Haushaltsgröße ist seit 1984 von 2,8 auf unter 2,4 gesunken.

Abbildung 6.9: Entwicklung Ausfälle „verstorben“, 1985-2005, $\beta = -0.08$

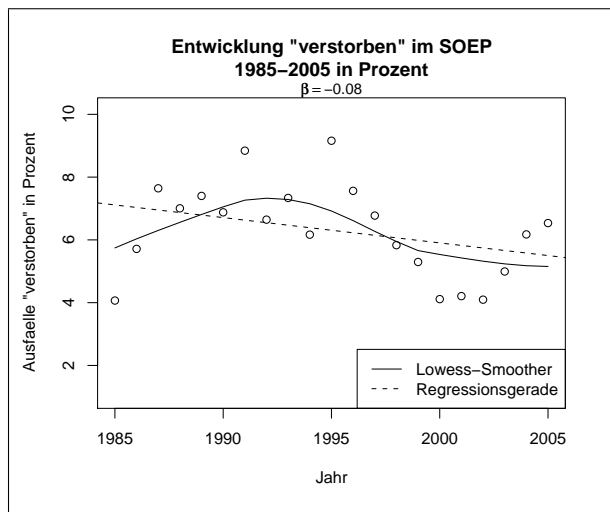
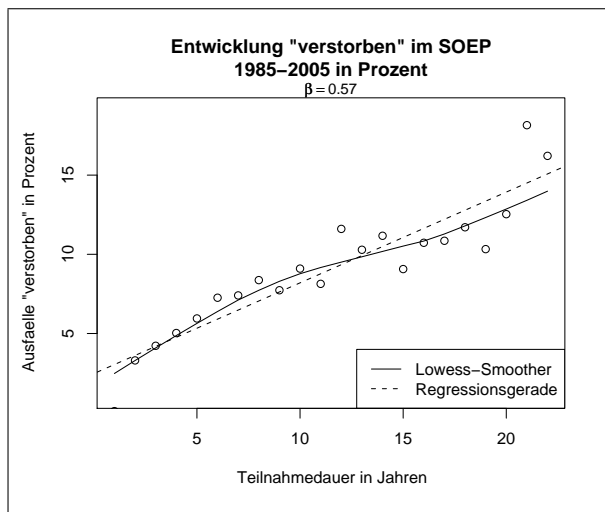


Abbildung 6.10: Ausfälle „verstorben“, nach Teilnahmejahren, $\beta = 0.57$



den, sondern alle Personen, die seitdem in Kontakt mit dem SOEP gekommen sind, verjüngt sich das SOEP auch mit jeder weiteren Welle; zudem kann vermutet werden, dass nicht alle Ausfälle nichtteilnahmefähiger Personen als solche erkannt und vom Interviewer erfasst werden, da hier u.U. Auskunft von Dritten notwendig ist.

Zudem besteht die Möglichkeit, dass Personen, die aufgrund von Alter oder Krankheit aus dem Panel ausfallen, schließlich versterben und so nicht als „nicht teilnahmefähig“ eingeordnet werden, sondern als Ausfälle durch Tod der Befragungsperson. Deutlich wird das im starken Anstieg von 1% bis auf knapp 8% bis zum 5. Teilnahmejahr in Abbildung 6.8 und dem anschließenden Abfallen der Running-Mean-Kurve¹²⁸.

Jährlich fallen rund 8% der Personen durch Tod aus dem Panel heraus, der Lowess-Smoother in Abbildung 6.9 ist deutlich s-förmig und zeigt zunächst einen Anstieg von 4% auf 9% Mitte der 90er Jahre, erreicht einen Tiefpunkt im Jahr 2000 und steigt seitdem wieder an. Der Rückgang seit Mitte der 90er Jahre geht zu einem Teil auch auf die Aufstockung des SOEP

¹²⁸Der Lowess-Smoother reagiert hier nicht auf den starken Anstieg zu Beginn; wählt man einen kleineren Glättungsparameter f (also eine kleinere Umgebung, die zur Berechnung der Plotpositionen verwendet wird, hier $f = 0.4$), reagiert der Smoother stärker auf Unebenheiten (vgl. Schnell, 1994, S. 109f); ab $f = 0.5$ verschwindet hier der starke Anstieg zu Beginn der Kurve.

Abbildung 6.11: Entwicklung Ausfälle „verzogen“, 1985-2005, $\beta = -0.34$

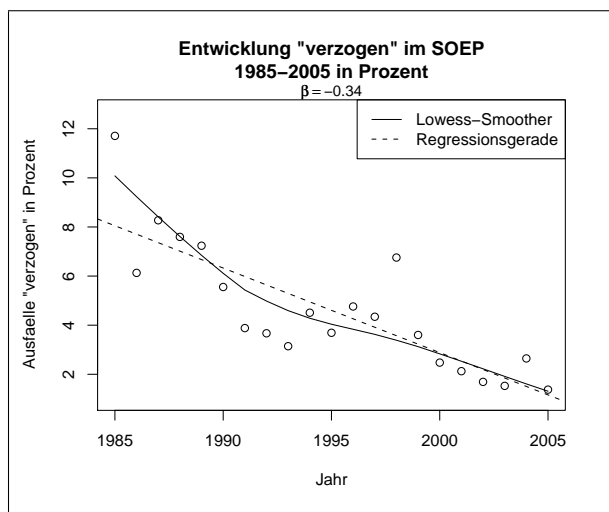
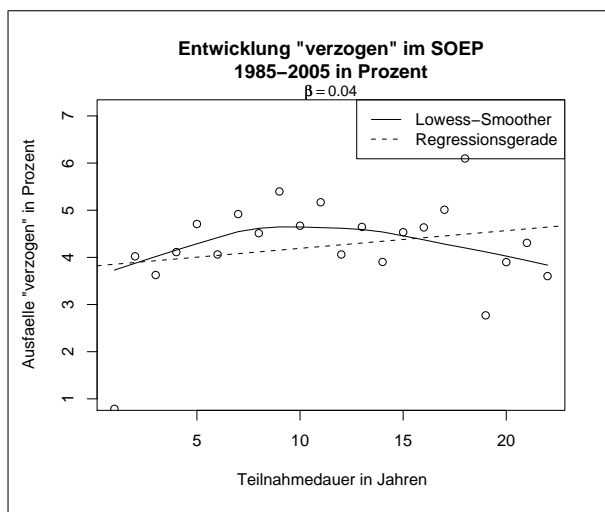


Abbildung 6.12: Ausfälle „verzogen“, nach Teilnahmejahren, $\beta = 0.04$



durch neue Substichproben zurück: 1995 wird das SOEP um die (vergleichsweise) kleine Zuwanderer-Stichprobe D erweitert, 1998 und 2000 ergänzen zwei neue, vergleichsweise große Substichproben das SOEP (Stichproben E und F); diese „Auffrischung“ macht sich auch in einem verringerten Anteil der Ausfälle durch Tod bemerkbar.

Betrachtet man wiederum die Ausfälle nicht nach Kalenderjahren sondern nach der Teilnahmedauer (Abbildung 6.10), zeigt sich deutlich, dass Ausfälle durch Tod als natürliche Ausfallursachen mit dem Alter der Befragungspersonen zunehmen, im Schnitt jedes Teilnahmejahr um 0.6%. Für junge Panelteilnehmer ist Ausfall durch Tod selten (einer von zwanzig Ausfällen), bei Personen mit langer Panelteilnahme (und damit auch höherem Lebensalter) werden Ausfälle durch Tod zu einer häufigeren, natürlichen, Ausfallursache (bis einer von sieben Ausfällen nach 20 Teilnahmejahren); diese Entwicklung stellen auch Fitzgerald et al. (1998, S. 255) für die amerikanische PSID fest¹²⁹.

Zieht ein Haushalt innerhalb Deutschlands um, wird er entsprechend der Weiterverfolgungsregeln des SOEP grundsätzlich in der nächsten Welle wieder kontaktiert; lediglich Umzüge

¹²⁹„[...] there is an slight increase in the percent attrititing because of death and a slight reduction in the percent attrititing because of mobility. Both of these trends are no doubt a result of the increasing age of the 1968 sample“.

Abbildung 6.13: Entwicklung Ausfälle „nicht auswertbar“, 1985-2005, $\beta = 0.06$

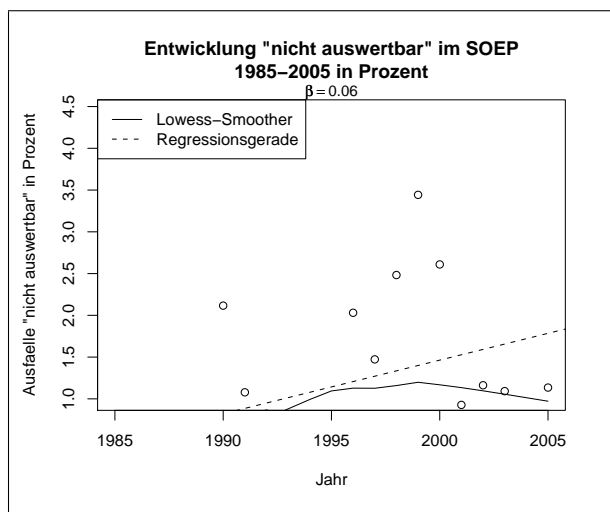
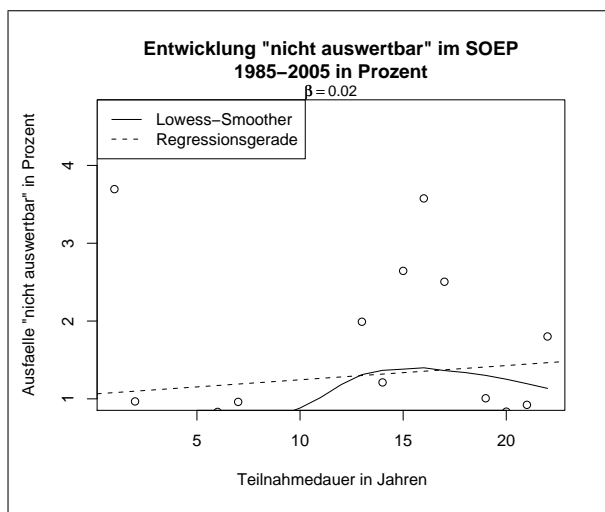


Abbildung 6.14: Ausfälle „nicht auswertbar“, nach Teilnahmejahren, $\beta = 0.02$

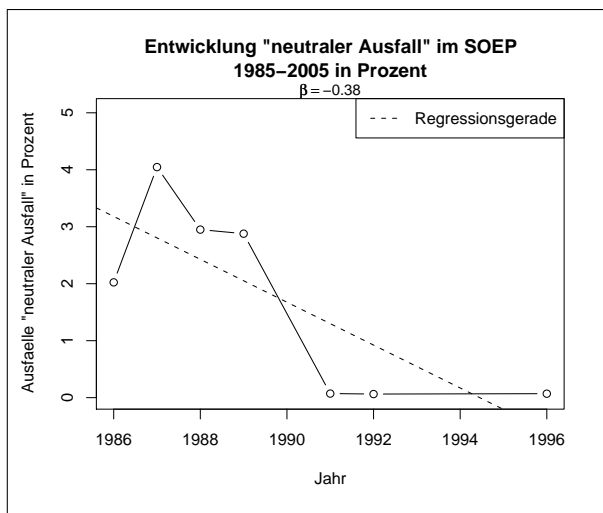


ins Ausland werden nicht weiterverfolgt. Abbildungen 6.11 und 6.12 zeigen die Entwicklung der durch Umzug ins Ausland verlorenen Personen. Insgesamt sinkt der Anteil Ausfälle durch ins Ausland verzogener Haushalte über die Jahre um durchschnittlich mehr als 0.3% pro Jahr.

Deutlich ist in Abbildung 6.11 ein Knick im Jahr 1990 zu erkennen; hier macht sich die geänderte Weiterverfolgungsregel bemerkbar; zudem kann vermutet werden, dass in den früheren Jahren des SOEP auch der Umzug von Personen aus ausländischen Haushalten („Gastarbeiterhaushalte“) eine Rolle spielt, die, 35 Jahre nach Beginn der Anwerbung ausländischer Arbeitskräfte, bei Erreichen des Rentenalters zum Teil Deutschland wieder verlassen und in ihre Heimatländer zurückkehren.

Ab 1991 liegt der Anteil der ins Ausland verzogenen Personen unter 5% jedes Jahr, lediglich 1998 steigt der Anteil erneut auf über 7%, sinkt allerdings anschließend auf unter 2% jedes Jahr. Betrachtet man die Entwicklung der durch Umzug verlorenen Personen nach Teilnahmejahren, zeigt der Lowess-Smoother eine leicht umgekehrt u-förmige Entwicklung mit einem Ausreisser nach 18 Jahren. Nach 5 Jahren steigt der Anteil der ins Ausland verzogenen Personen von 4% auf über 5% und sinkt nach 15 Teilnahmejahren auf unter 4%; auch dies spricht für die Überlegung, dass nach mehr als einem Jahrzehnt der Teilnahme vor

Abbildung 6.15: Entwicklung neutraler Ausfälle, 1985-2005



allen Personen aus Gastarbeiterhaushalten in ihre Heimatländer zurückkehren.

Nicht auswertbare Interviews sind eine seltene Ausfallursache; Abbildungen 6.13 und 6.14 zeigen die Entwicklung über die (Teilnahme-) Jahre. Im Durchschnitt über alle Jahre kommen rund 1.3% aller Personenausfälle pro Jahr durch nicht auswertbare Interviews zustande; lediglich in den Jahren 1995 bis 2000 bzw. nach 15 Teilnahmejahren ist eine Häufung feststellbar. Auffällig ist auch der vergleichsweise hohe Anteil von 4% im ersten Teilnahmejahr. Wie bei Ausfällen durch eine Teilnahmeunfähigkeit kann hier angenommen werden, dass diese Ausfallursache vor allem einen Übergang zu anderen Ausfallursachen darstellt und auch vom Interviewer nur selten als Ausfallursache verwendet wird. Das Gleiche gilt für neutrale Ausfälle; „neutral“ sind Ausfälle im SOEP dann, wenn Personen bspw. aus Design-Gründen ausgeschlossen werden; im SOEP kommen solche Ausfälle nur bis zum Jahr 1989 vor, wenn Nicht-Stammpersonen aus einem Stammhaushalt ausziehen; das Weiterverfolgungskonzept des SOEP sieht erst seit 1990 vor, dass alle Personen, die in Kontakt mit dem SOEP kommen weiterverfolgt werden.

Dazu kommen Personen aus Haushalten in Stichprobe D (Aus- und übersiedler), die nicht zur Zielpopulation gehören bzw. Personen aus Haushalten in Stichprobe G, die unter der Abschneidegrenze des monatlichen Haushaltsnettoeinkommen liegen; allerdings sind sie schon vor der Befragungsphase ausgeschlossen, und nicht mehr im Bruttobestand enthalten, daher

Abbildung 6.16: Entwicklung Ausfälle „verweigert“, 1985-2005, $\beta = 1.81$

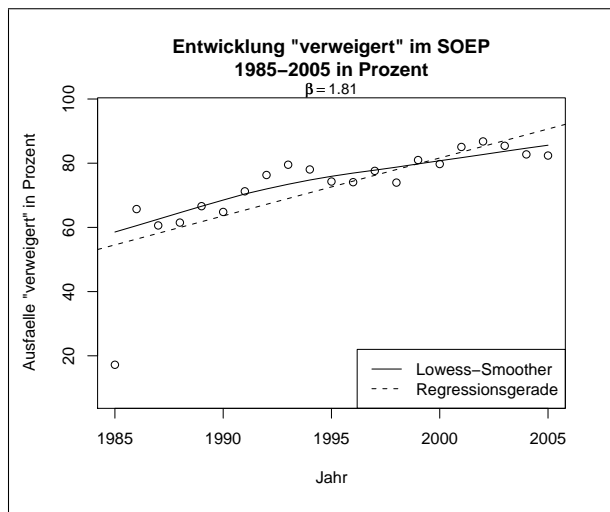
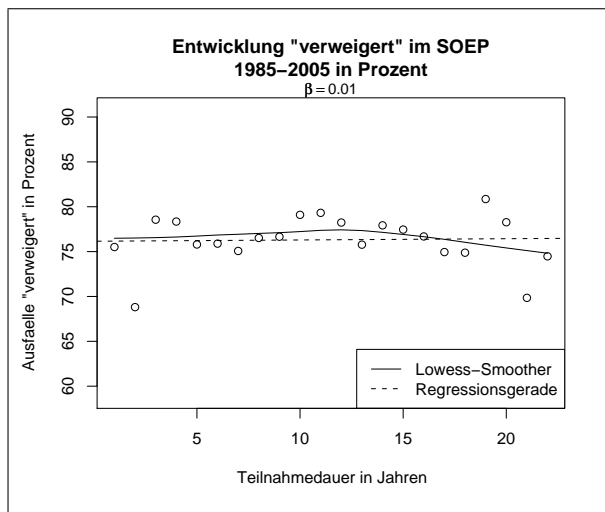


Abbildung 6.17: Ausfälle „verweigert“, nach Teilnahmejahren, $\beta = 0.01$



zeigt Abbildung 6.15 lediglich Ausfälle bis zum Jahr 1989¹³⁰.

Verweigerungen (Abbildungen 6.16 und 6.17) machen den größten Teil der Ausfälle im SOEP aus; im Durchschnitt über alle Jahre kommen über 76% aller Ausfälle eines Jahres durch Verweigerungen zustande; der Anteil der Verweigerungen steigt pro Jahr um durchschnittlich 1.8%; zu Beginn des SOEP lag der Anteil der Verweigerungen noch deutlich niedriger bei rund 60%; hier waren andere Ausfallursachen häufiger (s.o.). Seitdem sind bspw. Ausfälle durch Nichterreichbarkeit und Teilnahmeunfähigkeit zurückgegangen und Ausfälle verstärkt durch Verweigerungen geschehen. Betrachtet man den Anteil der Verweigerungen über die Teilnahmejahre (Abbildung 6.17), zeigt sich, dass der Anteil über die Teilnahmejahre sehr stabil ist, die Regressionsgerade verläuft nahezu parallel zur x-Achse; Befürchtungen, dass das Panel durch Ausfälle mit der Zeit kooperativer wird scheinen damit unbegründet; die Lowess-Kurve zeigt nur eine schwache umgekehrt u-förmige Entwicklung.

¹³⁰Im Datensatz sind für die Jahre 1991, 1992 und 1996 jeweils eine Beobachtung enthalten, die als verzogene Nicht-Stammperson geführt wird und für die keine sonstige Ausfallursache bekannt ist; laut der geänderten Weiterverfolgungsregeln des SOEP hätten diese Personen neuerdings weiterverfolgt werden müssen; hier handelt es sich also eventuell um Fehler bei der Feldarbeit oder fehlerhafte bzw. fehlende Kodierungen des Ausfalls.

Tabelle 7.1: Teilnahmemuster: Vollzeit - monoton - nichtmonoton

Teilnahmemuster	Anzahl Beobachtungen	Anzahl Beobachtungen in %
Vollzeitleilnahme	18,939	40.2
monotones Muster	14,997	31.9
nicht monotones Muster	13,152	27.9
Insgesamt	47,088	100.00

7 Ausfallmuster und Ausfallsequenzen

Betrachtet man die Teilnahme und Nichtteilnahme einer Person über die Zeit nach Teilnahmejahren, so lassen sich zunächst bestimmte Teilnahmemuster (Abschnitt 7.1) identifizieren; betrachtet man zudem die Abfolge von Teilnahme und Ausfallursache über die Teilnahmejahre ergeben sich Teilnahmesequenzen (Abschnitt 7.2).

7.1 Ausfallmuster im SOEP

Wie in Abschnitt 1.5.3 beschrieben, lassen sich Panelteilnehmer über die Zeit in mindestens drei Gruppen einteilen: Vollzeitleilnehmer nehmen über die ganze Zeit seit ihrem Eintritt in das Panel erfolgreich teil und können jede Welle erfolgreich interviewt werden; sie stellen den Ideal- und Normalfall des Panels dar. Ausfälle unterschieden sich dadurch, ob ein Panelteilnehmer bis zu einem endgültigen Ausfall über alle Wellen teilgenommen hat, oder ob der Ausfall lediglich temporär ist, so dass Panelteilnehmer in der nächsten (oder allgemein: in einer späteren Welle) wieder interviewt werden können. Während endgültige Ausfälle nach realisierten Teilnahmesequenzen monotone Ausfallmuster erzeugen (siehe Abschnitt 1.5.3), führen temporäre Ausfälle zu nichtmonotonen Ausfallmuster. Tabelle 7.1 zeigt die Teilnahmemuster im SOEP bis zum Jahr 2005. Insgesamt rund 47.100 Personen sind seit 1984 in Kontakt mit dem SOEP gekommen¹³¹; 40% davon, rd. 19.000 Personen sind Vollzeitleilnehmer, haben also seit Teilnahmebeginn ununterbrochen teilgenommen. Dabei ist es

¹³¹Nicht alle davon seit 1984. Neue Stichproben nach 1984 sowie Zuzüge von Nichtstammpersonen nach 1984 haben ebenfalls das SOEP ergänzt.

irrelevant, wann eine Person zum ersten Mal teilgenommen hat, wichtig ist nur, dass seitdem eine Teilnahme bis 2005 ununterbrochen ist¹³². Rund 15.000 Personen oder 32% der im SOEP seit 1984 erfassten Personen waren bis zu einem endgültigen Ausfall Vollzeitteilnehmer¹³³. 28% der Personen haben nicht-monotone Teilnahmemuster: sie haben mindestens einen temporären Ausfall aufzuweisen, kehren aber nach diesem Ausfall in das SOEP zurück; nach einem temporären Ausfall können sich weitere temporäre Ausfälle anschließen, ein endgültiger Ausfall, oder aber eine Beobachtung kehrt nach einem Ausfall wieder in das SOEP zurück und nimmt an den folgenden Wellen ununterbrochen teil.

Zwei von drei Beobachtungen mit monotonem Teilnahmemuster beenden ihre Teilnahme mit einer Verweigerung, 14% der monotonen Muster kommen durch Tod der Befragungsperson, 8% durch den Wegzug der Befragungsperson ins Ausland zustande. Für knapp 6% der monotonen Muster ist der Ausfallgrund nicht bekannt. Die restlichen Ausfallursachen (Nicht auffindbar, nicht erreichbar, nicht auswertbar, nicht durchführbar, neutrale Ausfälle) spielen mit jeweils um die 1% kaum eine Rolle für monotone Teilnahmemuster.

7.2 Teilnahmesequenzen im SOEP

Die Abfolge der Befragungsergebnisse ergibt über die Zeit (d.h. die Teilnahmejahre) eine Sequenz der Befragungsergebnisse; je nach Teilnahmedauer ergeben sich Sequenzen unterschiedlicher Länge, die entweder mit einer Teilnahme oder einem Ausfall enden. Eine Betrachtung der Sequenzen nach Kalenderjahren ist aufgrund der Struktur des SOEP und der Erweiterung durch neue Substichproben und Neuzugänge aufgrund des Weiterverfolgungskonzeptes nicht sinnvoll, hier müssten jeweils Beobachtungen unterschiedlicher Sequenzlänge

¹³²D.h., eine Person, die 2000 erstmals befragt wurde und seitdem ununterbrochen teilgenommen hat ist ebenso Vollzeitteilnehmer, wie eine Person, die seit 1984 ununterbrochen teilgenommen hat. 3.729 Personen nehmen seit 1984 am SOEP teil (32% des ursprünglichen Samples), 2.947 sind tatsächlich Vollzeitteilnehmer seit 1984 (23%).

¹³³Für Beobachtungen mit einem Ausfall 2005 ist ausser bei Ausfall durch Tod oder Wegzug ins Ausland (noch) nicht sicher, ob es sich um einen endgültigen Ausfall handelt; im Jahr 2005 stellen alle Ausfälle mit bis dahin ununterbrochener Teilnahme monotone Ausfallmuster dar; die Anzahl der monotonen Ausfälle für 2005 wird also überschätzt, die Anzahl der nichtmonotonen Teilnahmemuster für das Jahr 2005 unterschätzt.

miteinander verglichen werden; allerdings ist eine Betrachtung nach Teilnahmejahren (nach Sequenzlänge) aufschlussreich. Abbildung A.2 im Anhang auf Seite 144 zeigt als Beispiel für eine graphische Darstellung der Sequenzen einen Sequenz-Index-Plot aller Teilnehmer mit 20-jähriger Teilnahme (d.h. entweder bis zu einem Ausfall nach 20 Teilnahmejahren oder auch Beobachtungen, die 2005 nach 20 Teilnahmejahren ihre Teilnahme noch nicht beendet haben). Im Sequenz-Index-Plot wird jede Sequenz durch die horizontale Abfolge der Ausfallursachen durch farbige Streifen dargestellt; breite gleichfarbige Streifen zeigen häufige Sequenzen an, deutlich sind zum Beispiel die Beobachtungen sichtbar, die nach 20 Teilnahmejahren noch Vollzeitteilnehmer im Jahr 2005 sind. Auch ist deutlich, dass viele monotone Teilnahmemuster der Sequenzen nach 20 Teilnahmejahren (oder im Jahr 2005) mit einer Verweigerung bzw. durch Tod oder Wegzug ins Ausland enden. Insgesamt ist auch zu erkennen, dass nichtmonotone Teilnahmemuster für die Sequenzen mit 20 Teilnahmejahren nur rund ein Viertel der Teilnahmemuster ausmachen und vor allem durch Verweigerungen und nicht durchführbare Interviews in einer früheren Welle bestimmt sind. Da Sequenz-Index-Plots aber stark von der farblichen Kodierung der Elemente und der Sortierung der x-Achse bestimmt sind, wird auf eine weitere Darstellung der Sequenzen mit Sequenz-Index-Plots verzichtet. Tabelle 7.2 enthält als Beispiele für Teilnahmesequenzen die jeweils 10 häufigsten Sequenzen für Beobachtungen mit (mindestens) einem, fünf zehn, 15 und 20 Teilnahmejahren. Die Sequenzen geben die Abfolge der jährlichen Bearbeitungsergebnisse realisiert (0), nicht auffindbar (1), nicht erreichbar (2), verweigert (3), nicht auswertbar (4), unbekannt (5), nicht durchführbar (6), Ausfall durch Tod (7) und Wegzug ins Ausland (8) sowie neutrale Ausfälle (9) wieder. 47.088 Personen sind seit Beginn des SOEP 1984 in Kontakt mit dem SOEP gekommen (ohne Kinder unterhalb des Befragungsalters), sie haben mindestens eine Sequenz der Länge „1“; aus Tabelle 7.2 ist ersichtlich, dass nach dem Bearbeitungsergebnis „realisiert“ (Symbol 0) das zweithäufigste Element der Sequenzen eine Verweigerung (Symbol 3) und danach als dritthäufigste Sequenz bei einer Sequenzlänge von 1 der Ausfall durch Tod (Symbol 7), gefolgt vom Wegzug der Befragungsperson ins Ausland (Symbol 8) ist. Diese Rangfolge der häufigsten Sequenzen ändert sich nicht, betrachtet man Sequenzen höherer Ordnung: 30.059 Personen nehmen fünf oder mehr Jahre am SOEP teil; bei einer Se-

quenzlänge von 5 (also mindestens fünf Teilnahmejahre) ist eine Vollzeiteilnahme über fünf Wellen das häufigste Muster, gefolgt von einem monotonen Muster mit einer Verweigerung in der letzten Welle. Die dritthäufigste Sequenz bildet ein nichtmonotones Muster mit einem endgültigen Ausfall durch zwei aufeinander folgende Verweigerungen (Sequenz 00033); wiederum folgen auf die Verweigerungen Ausfälle durch Tod und Ausfälle durch Wegzug der Befragungsperson ins Ausland (Sequenzen 00007 und 00008; beide bilden monotone Teilnahmemuster). Die sechsthäufigste Sequenz bei Teilnehmern mit mindestens fünf Teilnahmejahren bilden Ausfälle, die in der vorletzten Welle als „nicht durchführbar - krank“ erfasst sind und in der letzten Welle schließlich durch eine Verweigerung ausfallen. Auch die 9.221 Sequenzen der Länge 10 (mindestens zehn Teilnahmejahre) zeigen dieses Muster: häufigste Sequenz sind Vollzeiteilnahmesequenzen, gefolgt vom monotonen Muster mit Verweigerung in der letzten Welle und dem nichtmonotonen Muster mit zwei aufeinanderfolgenden Verweigerungen in den letzten beiden Wellen. Genau wie bei Panelteilnehmern mit mindestens fünf teilgenommenen Wellen, ist die sechsthäufigste Teilnahmesequenz die durch ein nicht durchführbares Interview angekündigte Verweigerung in der letzten Welle. Bei Teilnahmedauern von 15 oder 20 Jahren spielt diese Sequenz keine Rolle unter den häufigsten zehn Sequenzen. Allerdings unterscheiden sie sich nicht in den häufigsten fünf Teilnahmesequenzen: Vollzeiteilnahme, monotonen Muster mit Teilnahme über alle Wellen bis zur Verweigerung in der letzten Welle bzw. den letzten beiden Wellen für das nichtmonotone Muster, monotonen Teilnahmemuster bis zum Tod oder Wegzug ins Ausland des Teilnehmers. Deutlich ist in Tabelle 7.2, dass, unabhängig von der Sequenzlänge, ein Vollzeitmuster die häufigste Sequenz ist und monotone Muster durch Verweigerungen, Tod und Wegzug ins Ausland entstehen; insgesamt sind monotone Muster häufiger als nicht-monotone Muster¹³⁴; bei den häufigsten Teilnahmesequenzen nicht-monotoner Muster spielen Ausfälle neben Verweigerungen kaum bis gar keine Rolle.

Für Beobachtungen, die endgültig aus dem Panel ausscheiden, sind vor allem die Sequenzen

¹³⁴Dies stellen Nicoletti und Peracchi (2005, S. 769f) auch für Teilnahmemuster für die ersten fünf Wellen des ECHP fest.

Tabelle 7.2: Die 10 häufigsten Sequenzen, Sequenzlänge 20, 15, 10, 5 und 1

	Sequenzlänge=20	Sequenzlänge=15	Sequenzlänge=10	Sequenzlänge=5	Sequenzlänge=1
01	00000000000000000000	0000000000000000	0000000000	00000	0
02	00000000000000000003	0000000000000003	0000000003	00003	3
03	00000000000000000033	000000000000033	0000000033	00033	7
04	00000000000000000007	000000000000007	0000000007	00007	8
05	00000000000000000008	000000000000008	0000000008	00008	5
06	00000000000000300000	3000000000000000	0000000063	00063	2
07	0000000000000030000	000000000030000	0000003033	00303	6
08	000000000000003000	00000000000004	0000000303	03033	4
09	0000000000000000030	0030000000000000	0000003000	30000	9
10	00000000000000000300	0300000000000000	0000030000	03000	1
Anzahl ^a	4,487	9,221	13,963	30,059	47,088

^a: Beobachtungen mit mindestens 20 (15, 10, 5, 1) Teilnahmejahren

der letzten beiden bzw. der letzten drei Wellen von Interesse¹³⁵; eine längere Betrachtung vor einem Ausfall aus dem Panel erscheint nicht plausibel; im Weiteren wird für die Betrachtung der Ausfallsequenzen daher angenommen, dass sich ein endgültiger Ausfall eines Panelteilnehmer nicht über mehr als drei Jahre erstreckt. Tabellen 7.3 und 7.4 zeigen die Sequenzen aller endgültigen Ausfälle aus dem SOEP bis zum Jahr 2004¹³⁶ für die letzten beiden (Tabelle 7.3) bzw. letzten drei Jahre (Tabelle 7.4) zusammen mit ihrer Häufigkeit und dem Anteil der Ausfallsequenz an allen Ausfallsequenzen¹³⁷. Ausfälle durch eine „harte“ Verweigerung nach einem realisierten Interview sind mit über 40% die häufigste Ausfallsequenz, betrachtet man Ausfälle von Personen, die mindestens zwei Jahre am Panel teilgenommen haben (un-

¹³⁵Bakeman und Gottman (1997, S. 103) empfehlen für lange Sequenzen, nur bestimmte Sequenzen (Sequenzteile) zu beschreiben, da für lange Sequenzen die Anzahl der unterschiedlichen Sequenzen (vor allem bei vergleichsweise vielen Sequenzbestandteilen) Häufigkeiten und Wahrscheinlichkeiten einzelner Sequenzen verschwindend gering werden: „Occasionally it may be useful to describe probabilities for particular sequences, no matter whether chains are two-event, three-event, or longer. [...] Thus usually attention focuses on transitional probabilities involving two events“.

¹³⁶Für Ausfälle im Jahr 2005 ist, bis auf Tod und Wegzug ins Ausland, nicht klar, ob es sich dabei um endgültige Ausfälle handelt. Daher sind alle Beobachtungen aus dem Jahr 2005 nicht in den Tabellen 7.3 und 7.4 enthalten.

¹³⁷„The most basic thing to do with event-sequence data is to define particular sequences, count them, and then report frequencies and/or probabilities for those sequences“ (Bakeman und Gottman, 1997, S. 101).

Tabelle 7.3: Sequenzen des Bearbeitungsergebnisses, Jahr vor dem Ausfall, 1984-2004

	Sequenz	Anzahl Ausfälle	Anteil in%	kumulierte Prozent
realisiert, harte Verweigerung	0-3	8,658	40.2	40.2
wiederholte Verweigerung	3-3	6,104	28.3	68.5
realisiert, verstorben	0-7	2,095	9.7	78.2
realisiert, ins Ausland verzogen	0-8	1,274	5.9	84.1
realisiert, unbekannt (2. Welle) ^a	0-5	825	3.8	87.9
nicht durchführbar/verweigert	6-3	626	2.9	90.8
nicht erreichbar, verweigert	1-3	242	1.1	92.0
realisiert, nicht auffindbar	0-2	190	0.9	92.8
nicht auffindbar, verweigert	2-3	190	0.9	93.7
realisiert, nicht auswertbar	0-4	140	0.7	94.4
wiederholt nicht auffindbar	2-2	135	0.6	95.0
realisiert, neutraler Ausfall	0-9	112	0.5	95.5
wiederholt nicht durchführbar	6-6	103	0.5	96.0
verweigert, nicht auffindbar	3-2	93	0.4	96.4

Basis: Anzahl endgültige Ausfälle bis einschließlich 2004: 21.554

enthält nur Ausfallsequenzen ab ca. 100 Beobachtungen

^a: v.a. „keine Angabe“ in Variable **BPERGS(Z)**

geachtet dessen, ob das Teilnahmemuster bis zu den letzten zwei Jahren monoton oder nicht monoton ist, vgl. Abschnitt 7.1). Mit über 28% geschieht mehr als jeder vierte endgültige Ausfall durch zwei aufeinanderfolgende Verweigerungen, dabei ist es irrelevant, ob die zweite, letzte Verweigerung eine „harte“ oder „weiche“ Verweigerung ist: auch nach zwei aufeinanderfolgenden weichen Verweigerungen wird im SOEP in der Regel eine Befragungsperson aus dem Panel ausgeschlossen.

Knapp 10% der Teilnahmesequenzen nach mindestens zwei Teilnahmejahren enden mit dem Tod der Befragungsperson. Betrachtet man die restlichen, selteneren, Sequenzen, fällt auf, dass einem Ausfall durch Tod keine Ausfälle durch „nicht durchführbar“ aufgrund von Krankheit oder Alter oder auch Verweigerungen vorausgehen (das ist lediglich bei 58 Beobachtungen (0.3% der Ausfälle) der Fall; vgl. dazu auch Ausfälle durch Tod in Tabelle 7.4). Ausfälle durch Wegzug der Befragungsperson ins Ausland nach vorhergehendem realisiertem Interview in der Vorwelle ist mit rund 6% die vierthäufigste Ausfallsequenz für endgültige Ausfälle von Teilnehmern mit mindestens 2 aufeinanderfolgenden Teilnahmen am SOEP. Für knapp 4% der endgültigen Ausfällen nach mindestens zwei Teilnahmejahren ist die Ausfallursache

nach einem realisierten Interview unbekannt¹³⁸. Knapp 3% der endgültigen Ausfälle sind Verweigerungen, die sich in der Vorwelle durch ein aufgrund von Alter oder Krankheit nicht durchführbares Interview ankündigen (0.5% der endgültigen Ausfälle geschehen durch Ausschluss aufgrund zwei aufeinanderfolgender Wellen mit Ausfall durch Teilnahmeunfähigkeit). Diese sechs Ausfallsequenzen der endgültigen Ausfälle nach mindestens zwei Teilnahmejahren machen über 90% der Ausfälle nach mindestens zwei Teilnahmejahren aus¹³⁹. Weitere 1,1% der endgültigen Ausfälle sind Verweigerungen nach einem Ausfall durch Nichterreichbarkeit. Knapp 1% der endgültigen Ausfälle mit mindestens zwei Teilnahmen sind Ausfälle nach endgültiger Nichterreichbarkeit; eventuell verbergen sich hierunter zum Teil auch weitere, nicht erfasste Umzüge ins Ausland. Ebenfalls knapp unter 1% liegen Ausfälle, die nach einem Umzug nicht auffindbar sind und im darauffolgenden Jahr zwar aufgefunden und erreicht werden, jedoch nicht zu einer Teilnahme bereit sind; hierunter fallen bspw. Personen nach Haushaltsauflösungen oder Aufspaltungen alter Haushalte, die vor dem Umzug Kontakt mit dem SOEP hatten¹⁴⁰; Ausfälle durch nicht auswertbare Interviews sind eventuell als versteckte Verweigerungen zu werten, sie führen in 0.7% aller Ausfälle zu einem endgültigen Ausschluss vom Panel. Diese Reihenfolge der Ausfallsequenzen für endgültige Ausfälle (Verweigerungen, Ausfall durch Tod und Umzug ins Ausland, wiederholte Verweigerungen und Ausfälle durch Teilnahmeunfähigkeit) verändert sich nicht, zieht man zusätzlich ein weiteres

¹³⁸Das sind vor allem Ausfälle von Beobachtungen der Welle B 1985: per Definition haben sie in der ersten Welle 1984 ein realisiertes Interview; die zweite Welle B des SOEP 1985 ist die einzige Welle, die eine größere Anzahl Ausfälle mit unbekannter Ausfallursache in der Variable `BPERGS(Z)` im Datensatz `BPPRUTTO` ausweist. In Tabelle 7.4 der endgültigen Ausfälle mit mindestens drei Teilnahmejahren kommen Ausfälle mit unbekannter Ursache nicht mehr unter den Ausfallsequenzen mit mehr als 100 Beobachtungen vor.

¹³⁹Bei 9 unterschiedlichen Elementen für die letzten beiden Sequenzen sind insgesamt $9^2 - (3 * 9) = 54$ unterschiedliche Sequenzen möglich (die absorbierenden Zustände sind nur an letzter Stelle der Sequenz möglich, müssen also von der vollständigen Anzahl der Kombinationen abgezogen werden).

¹⁴⁰Vgl. dazu auch Tabelle 7.4: hier haben Ausfälle mit mindestens drei Teilnahmejahren ein realisiertes Interview vor dem Ausfall durch Nichtauffinden des Haushalts durch den Interviewer. Eventuell ist für dieses Muster auch ein Interviewerwechsel (eventuell auch nach einem Umzug) verantwortlich: der (neue) Interviewer findet den Haushalt nicht an der angegebene Adresse; wird der Haushalt in der nächsten Welle vom neuen Interviewer aufgefunden, erfolgt eine Verweigerung.

Tabelle 7.4: Sequenzen des Bearbeitungsergebnisses, 2 Jahre vor dem Ausfall, 1984-2004

	Sequenz	Anzahl Ausfälle	Anteil in%	kumulierte Prozent
realisiert, harte Verweigerung	0-0-3	6,190	36.8	36.8
realisiert, wiederholt verweigert	0-3-3	4,942	29.4	66.2
realisiert, verstorben	0-0-7	1,836	10.9	77.1
realisiert, verzogen	0-0-8	971	5.8	82.9
real., n. durchführbar, verweigert	0-6-3	575	3.4	86.3
wiederholt verweigert	3-0-3	301	1.8	88.1
wiederholt verweigert	3-3-3	263	1.6	89.6
real., n. auffindbar, verweigert	0-2-3	162	1.0	90.6
real., endgültig n. auffindbar	0-0-2	128	0.8	91.4
real., wiederholt n. auffindbar	0-2-2	111	0.7	92.0
real., nicht auswertbar	0-0-4	109	0.7	92.7
real., wiederholt nicht durchführbar	0-6-6	98	0.6	93.3

Basis: Anzahl endgültige Ausfälle bis einschließlich 2004: 16.842
enthält nur Ausfallsequenzen ab ca. 100 Beobachtungen

Jahr vor dem Ausfall mit in Betrachtung¹⁴¹. Tabelle 7.4 zeigt die Teilnahmesequenzen der endgültigen Ausfälle mit mindestens drei Teilnahmejahren (Sequenzlänge von mindestens 3): Insgesamt knapp zwei Drittel aller Ausfälle kommt durch harte und weiche Verweigerungen in der letzten bzw. in den letzten beiden Wellen zustande, 1.8% aller Ausfälle durch eine harte Verweigerung in der letzten Welle, angekündigt durch eine weiche Verweigerung zwei Wellen vor dem Ausfall. Rund 11% aller Ausfälle mit mindestens drei Teilnahmejahren geschehen durch Tod nach wiederholt realisierten Interviews in den Vorwellen. 6% der endgültigen Ausfälle verlassen nach realisierten Wellen das Panel durch Umzug ins Ausland. 4% fallen durch Krankheit oder sonstiger Umstände, die eine Teilnahme unmöglich machen in der Vorwelle aus und verweigern schließlich endgültig (3.4%) oder werden aufgrund anhaltender Teilnahmeunfähigkeit vom Panel ausgeschlossen (0.6%). Insgesamt 1.5% aller Ausfälle nach mindestens drei Teilnahmejahren geschehen durch endgültiges Nichtauffinden des Haushalts an der alten Adresse. Wieder, wie schon in Tabelle 7.3 gesehen, bildet für 0.7% aller Ausfälle ein nicht auswertbares Interview nach vorangegangenen realisierten Interviews eine Ursache

¹⁴¹bei einer Sequenzlänge von 3 sind für die 9 Ausfallursachen insgesamt $9^3 - (3 \cdot 9^2) = 729 - 243 = 486$ Kombinationen möglich; wieder sind Sequenzen mit absorbierenden Zuständen vor dem letzten Sequenzelement abzuziehen.

für einen endgültigen Ausfall.

Aus dieser Betrachtung der Ausfallsequenzen ist deutlich zu erkennen, dass sich Ausfälle auf wenige typische Ausfallsequenzen beschränken: auf acht Ausfallsequenzen der Länge 3 konzentrieren sich über 90% aller Ausfälle. Typische Ausfallsequenzen sind Ausfälle durch Verweigerungen nach mindestens einem vorhergehendem realisiertem Interview sowie Ausfälle durch Tod oder Wegzug der Befragungsperson ins Ausland nach wiederholt realisierten Interviews; auch ein nicht durchführbares Interview gefolgt von einer endgültigen Verweigerung ist typisch für einen Ausfall aus dem SOEP; zudem gehören Ausfälle durch (wiederholtes) Nichtauffinden des Zielhaushalts, zum Teil gefolgt von einer Verweigerung zu den typischen, wenn auch selteneren Ausfallsequenzen.

8 Multinomiales Logit-Modell der Bearbeitungsergebnisse

Eng verbunden mit der Darstellung der Sequenzen im letzten Kapitel ist die Modellierung der Bearbeitungsergebnisse und Panelausfälle als Markov-Kette (vgl. Gottman und Roy, 1990, S. 9f) eines Prozesses mit diskreten¹⁴² Zeit- und Abfolgeschritten n und begrenzten und definierten Zuständen i der Befragungsergebnisse S . Während im letzten Kapitel die letzten beiden (bzw. letzten drei) Sequenzen betrachtet wurden, sind für eine Markov-Kette nicht nur die letzten Sequenzen von Interesse, sondern alle (im Falle einer stationären Markov-Kette) Zweiersequenzen: über alle vollständigen Sequenzen wird ein Fenster der (Sequenz-) Länge 2 geschoben („Moving time window“; vgl. Gottman und Roy, 1990, S. 16f). Die Sequenzfolge der Bearbeitungsergebnisse bis zu einem endgültigen Ausfall ist dann als Markov-Kette beschreibbar, wenn das Bearbeitungsergebnis zum Zeitpunkt n nur vom Bearbeitungsergebnis zum Zeitpunkt $n - 1$ abhängt, aber nicht von früheren Bearbeitungsergebnissen. Ergebnis eines Markov-Modells sind die Übergangswahrscheinlichkeiten zwischen den einzelnen Zuständen; da es sich bei den Bearbeitungsergebnissen und Ausfallursachen um ungeordnete

¹⁴²Davon zu unterscheiden sind Markov-Ketten mit kontinuierlicher Zeiteinteilung: $t = [0, \text{inf})$. n bezeichnet ganze Zahlen, wogegen t reelle Zahlen umfasst (vgl. Norris, 1997, S. 1).

te Kategorien handelt, ist ein multinomiales Logitmodell eine Möglichkeit, diese Übergangswahrscheinlichkeiten des Markov-Modells zu schätzen.

8.1 Abfolge der Bearbeitungsergebnisse als Markov-Kette

Für Markov-Modelle mit diskreter Zeiteinteilung ist nur die Abfolge der Zustände von Belang, aber weder der tatsächliche Zeitpunkt, noch die Dauer, die in einem Zustand verbracht wird (vgl. Csenki, 1994, S. 1). Isaacson und Madsen (1976, S. 13f) beschreiben diese Markov-Eigenschaft eines Prozesses so: „If we know that the particle is in state i at time $n - 1$, we might ask where it will be at time n . Generally a precise answer to this question can not be given; rather we must give a probability distribution over the states S “. Diese Wahrscheinlichkeitsverteilung P ist also gegeben durch die bedingte Wahrscheinlichkeit zum Zeitpunkt n in Zustand $X = j$ zu sein, gegeben Zustand $X = i$ zum Zeitpunkt $n - 1$:

$$P[X_n = j | X_{n-1} = i] \quad (8.1)$$

Die Markov-Bedingung ist dann

$$P[X_n = i | X_{n-1} = i_{n-1}, X_{n-2} = i_{n-2}, \dots, X_1 = i] = P[X_n = i | X_{n-1} = i_{n-1}], \quad (8.2)$$

d.h., Gleichung 8.1 gilt für alle n und alle i unterschiedlichen Zustände: „We might also ask whether or not knowledge of where the particle was *before* time $n - 1$ would alter this conditional probability [Gleichung 8.1, TG]. If such knowledge does not alter the conditional probability distribution, the Markov property is satisfied“ (Isaacson und Madsen, 1976, S. 13-14, Hervorhebung im Original). Betrachtet man die häufigsten Ausfallsequenzen in Tabellen 7.2, 7.3 und 7.4 (Seiten 108, 109 und 111), so zeigen sich hier keine Unterschiede der Reihenfolge der häufigsten Sequenzen, die gemeinsam rd. 90% der Ausfälle ausmachen; Kenntnis von früheren Teilnahmesequenzen verändert also nicht die Reihenfolge der Bedeutung der Ausfallsequenzen und die Bedeutung der einzelnen Ausfallursachen. Die Markov-Annahme (Gleichung 8.2) scheint somit zu halten.

Eine weitere wichtige Eigenschaft von Markov-Ketten ist die Unterscheidung *homogener* bzw. *stationärer* Markov-Ketten und *nichtstationärer* Markov-Ketten (Isaacson und Ma-

dsen, 1976, S. 15). Eine Markov-Kette ist stationär (oder homogen), wenn die Übergangswahrscheinlichkeit von einem Zustand i zu einem Zustand j nicht von einem Zeitpunkt des Übergangs abhängt (Isaacson und Madsen, 1976, S. 15; Bakeman und Gottman, 1997, S. 138f):

$$P[X_n = j | X_{n-1} = i] = P[X_{n+k} = j | X_{n+k-1} = i] \quad (8.3)$$

mit $k = -1, 0, 1, 2, \dots$. In Bezug auf die Abfolge der Panelausfälle bedeutet dies, dass die Wahrscheinlichkeit für einen Übergang des Befragungsergebnisses nur vom letzten Befragungsergebnis abhängt, nicht aber, zu welchem Zeitpunkt, bzw. an welcher Stelle der Teilnahmesequenz (vgl. auch Gottman und Roy, 1990, S. 20). Demnach ist bspw. ein Übergang von einem realisierten Interview in Welle 3 zu einer Verweigerung in Welle 4 gleich wahrscheinlich, wie ein Übergang vom realisierten Interview in Welle 5 zu einer Verweigerung in Welle 6¹⁴³.

Ausgehend vom begrenzten Zustandsraum S der definierten Befragungszustände können nun die Übergangswahrscheinlichkeiten und eine Matrix der Übergangswahrscheinlichkeiten zwischen den Befragungszuständen definiert werden: p_{ij} bezeichnet die Übergangswahrscheinlichkeit von Zustand i zu Zustand j , im Falle nichtstationärer Markov-Ketten (d.h. der Zeitpunkt des Übergangs ändert die Übergangswahrscheinlichkeit) lautet die Übergangswahrscheinlichkeit zwischen den Zuständen i und j $p_{ij}^{n-1,n}$. (vgl. Isaacson und Madsen, 1976, S. 15). Für i Zustände des Zustandsraumes S (Befragungsergebnisse) existieren im stationären Modell i^2 Übergangswahrscheinlichkeiten: eine quadratische Matrix P der Übergangswahrscheinlichkeiten. Einige der Zellen der Häufigkeitsmatrix sind nicht besetzt, d.h. die Wahrscheinlichkeit eines Übergangs in andere Zustände ist gleich 0, da einige Zustände (Befragungsergebnisse bzw. Ausfälle) absorbierend sind und ein Wechsel aus ihnen in andere Zustände zum nächsten Zeitpunkt nicht möglich ist¹⁴⁴. Die einzigen Bedingungen für eine

¹⁴³Dieser Umstand steht dann der Befürchtung von Laurie et al. (1999, S. 278) entgegen; sie nehmen an, dass das Sample mit zunehmender Teilnahmedauer immer kooperativer wird, d.h. die Wahrscheinlichkeit für eine Verweigerung sinkt mit jedem Teilnahmejahr.

¹⁴⁴Fällt eine Person durch Tod aus, ist ein Übergang in einen anderen Befragungszustand in der folgenden Welle ausgeschlossen; gleiches gilt bspw. auch für Wegzug ins Ausland, neutrale Ausfälle und endgültige

Übergangsmatrix sind, dass keine Zelle negative Zahlen (Wahrscheinlichkeiten) enthält und sich Zellen über jede Zeile zu 1 addieren (d.h. ein Übergang findet mit der Wahrscheinlichkeit von 1 in einen die Spalten aufspannenden Zustände statt). Isaacson und Madsen (1976, S. 16-17) bemerken dazu, dass damit zunächst kein Unterschied zwischen einer stochastischen Matrix und einer Übergangsmatrix besteht; eine Übergangsmatrix stellt lediglich ein Spezial- oder Anwendungsfall einer stochastischen Matrix dar: „The first question faced by someone using the theory of Markov chains is whether or not the process is Markov. If he believes he is working with a discrete-time stationary Markov chain, the next step is to find the transition matrix“. Im Falle der Befragungsergebnisse bzw. Ausfälle im SOEP ergibt sich eine 9x9-Matrix der Übergangswahrscheinlichkeiten eines stationären Markov-Modells erster Ordnung

$$P = \begin{pmatrix} p_{1,1} & p_{1,2} & p_{1,3} & p_{1,4} & p_{1,5} & p_{1,6} & p_{1,7} & p_{1,8} & p_{1,9} \\ p_{2,1} & p_{2,2} & p_{2,3} & p_{2,4} & p_{2,5} & p_{2,6} & p_{2,7} & p_{2,8} & p_{2,9} \\ p_{3,1} & p_{3,2} & p_{3,3} & p_{3,4} & p_{3,5} & p_{3,6} & p_{3,7} & p_{3,8} & p_{3,9} \\ p_{4,1} & p_{4,2} & p_{4,3} & p_{4,4} & p_{4,5} & p_{4,6} & p_{4,7} & p_{4,8} & p_{4,9} \\ p_{5,1} & p_{5,2} & p_{5,3} & p_{5,4} & p_{5,5} & p_{5,6} & p_{5,7} & p_{5,8} & p_{5,9} \\ p_{6,1} & p_{6,2} & p_{6,3} & p_{6,4} & p_{6,5} & p_{6,6} & p_{6,7} & p_{6,8} & p_{6,9} \\ p_{7,1} & p_{7,2} & p_{7,3} & p_{7,4} & p_{7,5} & p_{7,6} & p_{7,7} & p_{7,8} & p_{7,9} \\ p_{8,1} & p_{8,2} & p_{8,3} & p_{8,4} & p_{8,5} & p_{8,6} & p_{8,7} & p_{8,8} & p_{8,9} \\ p_{9,1} & p_{9,2} & p_{9,3} & p_{9,4} & p_{9,5} & p_{9,6} & p_{9,7} & p_{9,8} & p_{9,9} \end{pmatrix}. \quad (8.4)$$

Die Zeilen bezeichnen die Befragungsergebnisse der Vorwelle, die Spalten die Befragungsergebnisse der aktuellen Welle¹⁴⁵.

Verweigerungen; zu den absorbierenden Zuständen im SOEP siehe Kapitel 1.5.1.

¹⁴⁵Beispielsweise mit den Zeilen und Spalten realisiert (1), nicht erreichbar (2), nicht auffindbar (3), verweigert (4), nicht auswertbar (5), nicht durchführbar (6), verstorben (7), verzogen (8), unbekannt (9). Im Falle der absorbierenden Zustände (7) und (8) sind dann die Übergangswahrscheinlichkeiten in andere Zustände alle 0.

8.2 Schätzung der Übergangswahrscheinlichkeiten - Multinomiales Logitmodell

Der Zustandsraum der Bearbeitungsergebnisse besteht aus den einzelnen Ausfallursachen $S = (\text{„realisiert“}, \text{„nicht erreichbar“}, \text{„nicht auffindbar“}, \text{„nicht auswertbar“}, \text{„nicht durchführbar“}, \text{„verstorben“}, \text{„verzogen“}, \text{„neutraler Ausfall“} \text{ und } \text{„unbekannte Ursachen“})$; diese Bearbeitungsergebnisse bzw. Ausfallursachen liegen als ungeordnete Kategorien vor. Die Übergangswahrscheinlichkeiten zwischen diesen ungeordneten Kategorien können mit multinomialen Logit-Modellen geschätzt werden¹⁴⁶. Formal ist das multinomiale Logit-Modell definiert als

$$\Pr(y = m|x) = \frac{e^{\beta_{m|b}x}}{\sum_{j=1}^J e^{\beta_{j|b}x}} \quad (8.5)$$

$$\ln \Omega_{m|b}(x) = \ln \frac{\Pr(y = m|x)}{\Pr(y = b|x)} = x\beta_{m|b} \quad (8.6)$$

für alle Kategorien der abhängigen Variable $m = 1$ bis J (Long und Freese, 2001, S. 175). b bezeichnet die gewählte Basiskategorie der abhängigen Variable, dergegenüber alle anderen Kategorien m verglichen werden. Im Falle eines Markov-Prozesses ist $x = y_{t-1}$, d.h. das Bearbeitungsergebnis der Vorwelle. Die vorhergesagten Wahrscheinlichkeiten aufgrund des multinomialen Logitmodells sind unabhängig von der gewählten Basiskategorie, lediglich die Parametrisierung der Koeffizienten ändert sich bei unterschiedlichen Basiskategorien (Long und Freese, 2001, S. 175). Die Wahrscheinlichkeit für Kategorie m zur Basiskategorie k ist also

$$\Pr(y = m|x) = \frac{e^{x\beta_{m|k}}}{\sum_{j=1}^J e^{x\beta_{j|m}}}, \quad (8.7)$$

¹⁴⁶Zur Schätzung der Übergangswahrscheinlichkeiten könnten auch $m - 1$ binäre logistische Regressionen verwendet werden (für $J = 1$ in Gleichung 8.5 entspricht das multinomiale Logitmodell auch einem binären Logitmodell (Greene, 2003, S. 721f)). Allerdings besteht bei der Verwendung der logistischen Regressionen das Problem, dass jeder Vergleich auf einer unterschiedlichen Basiskategorie basiert (Long und Freese, 2001, S. 172).

(vgl. Gleichung 8.5) und das Verhältnis der logarithmierten Odds für Kategorie m zur Basiskategorie k (vgl. Greene, 2003, S. 721)

$$\ln \left[\frac{\text{Pr}_{mj}}{\text{Pr}_{mk}} \right] = x_i(\beta_m - \beta_k) \quad (8.8)$$

(vgl. Gleichung 8.6).

Tabelle 8.1 zeigt die Ergebnisse des multinomialen Logitmodells der Ausfallursachen als Markov-Prozess: abhängige Variable ist das Bearbeitungsergebnis bzw. die Ausfallursache, unabhängige Variable sind Dummies der Bearbeitungsergebnisse der Vorwelle. Als Basiskategorie wurde „realisiert“ gewählt, d.h. die Koeffizienten geben die Verhältnisse der logarithmierten Odds des Bearbeitungsergebnisses zur Basiskategorie „realisiert“ an, für Gleichung 1 („nicht erreichbar“) bspw:

$$\ln \Omega_{\text{nicht erreichbar}|\text{realisiert}} = \ln \frac{\text{Pr}(y = \text{nicht erreichbar}|x)}{\text{Pr}(y = \text{realisiert}|x)} \quad (8.9)$$

$$= x\beta_{\text{nicht erreichbar}|\text{realisiert}} \quad (8.10)$$

Ein Koeffizient von bspw. $\beta = -2.288$ gibt dann an, dass sich die logarithmierte Wahrscheinlichkeit, in der Vorwelle ein realisiertes Interview zu haben und in der nächsten Welle durch Nichterreichbarkeit auszufallen um 2.288 verringert, bzw. das $e^\beta = e^{-2.288} = 0.101$ -fache der logarithmierten Wahrscheinlichkeit für ein wiederholt realisiertes Interview beträgt: das relative Risiko für einen Ausfall durch Nichterreichbarkeit nach einem realisiertem Interview beträgt 0.10.

Tabelle 8.1: Multinomiales Logit-Modell: Bearbeitungsergebnis - Bearbeitungsergebnis Vorwelle

Variable	Koeffizient		Standardfehler (β)
	β	e^β	
Gleichung 1: nicht erreichbar			
realisiert Vorwelle	-2.288***	0.10	0.061
nicht erreichbar Vorwelle	1.989***	7.31	0.138
nicht auffindbar Vorwelle	-0.152	0.86	0.582
verweigert Vorwelle	-0.600***	0.55	0.172
nicht auswertbar Vorwelle ^a	-44.606	0.00	0.000
nicht durchführbar Vorwelle	-1.318**	0.27	0.580
unbekannt Vorwelle	-0.191	0.83	1.007
Konstante	-4.165***	0.02	0.039

Fortsetzung auf folgender Seite...

... Fortsetzung Tabelle 8.1

Variable	Koeffizient		Standardfehler (β)
	β	e^β	
Gleichung 2: nicht auffindbar			
realisiert Vorwelle	1.541***	4.67	0.207
nicht erreichbar Vorwelle	4.524***	92.20	0.279
nicht auffindbar Vorwelle	7.024***	1123.27	0.231
verweigert Vorwelle	4.146***	63.18	0.221
nicht auswertbar Vorwelle ^a	-41.342	0.00	0.000
nicht durchführbar Vorwelle	5.020***	151.41	0.244
unbekannt Vorwelle	6.718***	827.16	0.287
Konstante	-7.492***	0.00	0.204
Gleichung 3: verweigert			
realisiert Vorwelle	-0.052***	0.95	0.020
nicht erreichbar Vorwelle	2.271***	9.69	0.069
nicht auffindbar Vorwelle	2.601***	13.48	0.097
verweigert Vorwelle	3.214***	24.88	0.027
nicht auswertbar Vorwelle	1.964***	7.13	0.211
nicht durchführbar Vorwelle	2.605***	13.53	0.056
unbekannt Vorwelle	2.157***	8.65	0.184
Konstante	-2.642***	0.071	0.019
Gleichung 4: nicht auswertbar			
realisiert Vorwelle	-1.507***	0.22	0.106
nicht erreichbar Vorwelle	1.116***	3.05	0.418
nicht auffindbar Vorwelle ^a	-44.203	0.00	0.000
verweigert Vorwelle	1.138***	3.12	0.169
nicht auswertbar Vorwelle	5.065***	158.38	0.221
nicht durchführbar Vorwelle	2.168***	8.74	0.231
unbekannt Vorwelle ^a	-43.115	0.00	0.000
Konstante	-5.659***	0.00	0.082
Gleichung 5: nicht durchführbar			
realisiert Vorwelle	1.451***	4.27	0.133
nicht erreichbar Vorwelle	2.742***	15.52	0.319
nicht auffindbar Vorwelle	3.815***	45.38	0.305
verweigert Vorwelle	2.979***	19.67	0.161
nicht auswertbar Vorwelle	3.997***	54.43	0.482
nicht durchführbar Vorwelle	4.720***	112.17	0.165
unbekannt Vorwelle	4.943***	140.19	0.311
Konstante	-6.592***	0.00	0.130
Gleichung 6: verstorben			
realisiert Vorwelle	4.439***	84.69	0.500
nicht erreichbar Vorwelle	5.346***	209.77	0.585
nicht auffindbar Vorwelle	5.477***	239.13	0.674
verweigert Vorwelle	5.316***	203.57	0.513

Fortsetzung auf folgender Seite...

... Fortsetzung Tabelle 8.1

Variable	Koeffizient		Standardfehler
	β	e^β	(β)
nicht auswertbar Vorwelle	7.912***	2729.84	0.569
nicht durchführbar Vorwelle	6.796***	894.26	0.518
unbekannt Vorwelle	4.927***	137.97	1.124
Konstante	-9.283***	0.00	0.500
Gleichung 7: verzogen			
realisiert Vorwelle	1.875***	6.52	0.179
nicht erreichbar Vorwelle	4.506***	90.56	0.244
nicht auffindbar Vorwelle	3.985***	53.79	0.383
verweigert Vorwelle	3.332***	27.99	0.207
nicht auswertbar Vorwelle	4.098***	60.22	0.616
nicht durchführbar Vorwelle	4.534***	93.13	0.229
unbekannt Vorwelle	2.847***	17.24	1.022
Konstante	-7.204***	0.00	0.177
Gleichung 8: neutral			
realisiert Vorwelle	2.215***	9.16	0.713
nicht erreichbar Vorwelle	4.335***	76.32	1.001
nicht auffindbar Vorwelle ^a	-41.716	0.00	0.000
verweigert Vorwelle	4.763***	117.10	0.737
nicht auswertbar Vorwelle ^a	-38.974	0.00	0.000
nicht durchführbar Vorwelle	4.781***	119.22	0.867
unbekannt Vorwelle	5.620***	275.89	1.230
Konstante	-9.976***	0.00	0.707
Gleichung 9: keineAngabe.unbekannt			
realisiert Vorwelle	0.971***	2.64	0.137
nicht erreichbar Vorwelle ^a	-44.245	0.00	0.000
nicht auffindbar Vorwelle ^a	-43.456	0.00	0.000
verweigert Vorwelle	-1.012	0.36	0.720
nicht auswertbar Vorwelle	3.133***	22.95	0.730
nicht durchführbar Vorwelle ^a	-43.983	0.00	0.000
unbekannt Vorwelle ^a	-42.353	0.00	0.000
Konstante	-6.644***	0.00	0.134
N			
		379347	
Log-likelihood			
		-157000.50	
$\chi^2_{(63)}$			
		33835.16	
Pseudo-R ² (McFadden)			
		0.097	

Signifikanzniveau : * : 10% ** : 5% *** : 1%

^a: Zellen enthalten nur sehr wenige Beobachtungen

Basiskategorie: „realisiert“

Aus Gleichung 1 in Tabelle 8.1 („Nichterreichbar“ gegen „Realisiert“ in der nächsten Welle) zeigt sich, dass die logarithmierte Wahrscheinlichkeit, nach einer realisierten Welle durch Nichterreichbarkeit auszufallen um 2.28 niedriger ist, als auch in der nächsten Welle wieder realisiert zu werden, das relative Risiko nach einem erfolgreichen Interview in der nächsten Welle nicht erreichbar zu sein beträgt das 0.1-fache der logarithmierten Wahrscheinlichkeit, in der nächsten Welle erneut erfolgreich interviewt zu werden. Die logarithmierte Wahrscheinlichkeit, nach einem Ausfall durch Nichterreichbarkeit in der letzten Welle auch in der nächsten Welle nicht erreichbar zu sein erhöht sich um 1.9, d.h. das relative Risiko, erneut nicht erreichbar zu sein ist 7 mal höher als in der nächsten Welle erfolgreich interviewt zu werden. Demgegenüber ist die logarithmierte Wahrscheinlichkeit nach einem Ausfall durch Nichtauffinden an der alten Adresse in der nächsten Welle (an der neuen Adresse) nicht erreichbar zu sein 0.86 mal die logarithmierte Wahrscheinlichkeit, in der folgenden Welle erfolgreich interviewt zu werden; auch nach einer Verweigerung in der Vorwelle ist die logarithmierte Wahrscheinlichkeit, nicht erreichbar zu sein etwa halb so hoch, wie für ein erfolgreiches Interview. Die logarithmierte Wahrscheinlichkeit, nach einem nicht durchführbarem Interview nicht erreichbar zu sein beträgt nur rund ein Viertel der logarithmierten Wahrscheinlichkeit, erfolgreich interviewt zu werden. Insgesamt muss bei diesen Ergebnissen beachtet werden, dass die logarithmierte Wahrscheinlichkeit für einen Ausfall durch Nichterreichbarkeit gegenüber einem realisierten Interview sehr gering ist (die Konstante beträgt -4.165 , das Risiko für einen Ausfall durch nicht Erreichbarkeit liegt beim 0.02-fachen der logarithmierten Wahrscheinlichkeit für ein realisiertes Interview). Dasselbe gilt auch für alle anderen Ausfallursachen (siehe dazu auch noch einmal Tabelle 7.2 auf Seite 108; viele Zweiersequenzen sind auch bei der großen Fallzahl sehr selten. Die Koeffizienten sind daher mit Vorsicht zu betrachten.).

Gleiches gilt also für eine Ausfall durch Nichtauffinden an der alten Adresse (Gleichung 2 in Tabelle 8.1): Ausfälle durch Nichtauffinden sind insgesamt selten; das relative Risiko nach Ausfall durch Nichtauffinden an der alten Adresse in der Vorwelle auch in der folgenden Welle an der alten (oder einer neu ermittelten) Adresse nicht aufgefunden zu werden beträgt jedoch mehr als das 1000-fache des Risikos, nach dem Nichtauffinden noch ein realisiertes Interview

zu erhalten; insgesamt zeigt sich dies für alle anderen Bearbeitungsergebnisse ebenfalls: das Risiko *nach* einem Ausfall in der Folgewelle nicht auffindbar zu sein scheint in der Folgewelle immer höher zu sein, als ein realisiertes Interview zu erhalten. Obwohl „nicht aufgefunden“ per Definition in der Weiterverfolgungsregel kein absorbierender Zustand ist, scheint die Wahrscheinlichkeit hoch, spätestens in der Folgewelle endgültig aus dem Panel auszufallen. Wie oben gezeigt (siehe Abbildung 6.3 Seite 6.3), machen Ausfälle durch Nichtauffinden des Haushalts jährlich lediglich zwischen 2 und 5% aller Ausfälle aus.

Ein ähnliches Ergebnis zeigt sich, betrachtet man Ausfälle durch Verweigerungen (Gleichung 3 in Tabelle 8.1). Ausfälle durch Verweigerungen machen mit 75% den größten Teil der Ausfälle aus dem SOEP aus (siehe Abbildung 6.16, Seite 96). Nach einem Ausfall in der letzten Welle (ungeachtet der genauen Ausfallursache) ist die Wahrscheinlichkeit für eine Verweigerung in der folgenden Welle deutlich höher, als für ein realisiertes Interview. Die logarithmierte Wahrscheinlichkeit für eine erneute Verweigerung ist rund 25 mal so hoch wie für ein realisiertes Interview. Auch das Risiko einer Verweigerung nach einem Ausfall durch Nichtauffinden ist um das 13-fache höher, als für ein erfolgreiches Interview nach einem Ausfall durch Verweigerung. Besteht eine Lücke durch einen Ausfall, sind Teilnehmer offenbar weniger bereit, im nächsten Jahr wieder am SOEP teilzunehmen; lediglich das Risiko für eine Verweigerung nach einem erfolgreichen Interview in der vorherigen Welle ist geringer als eine erneute Realisierung des Interviews.

Das Risiko eines nicht auswertbaren Interviews nach einer erfolgreichen Teilnahme beträgt 0.22. Für alle Ausfallursachen ist das Risiko, in der nächste Welle durch ein nicht auswertbares Interview auszufallen höher als ein realisiertes Interview: Die logarithmierte Wahrscheinlichkeit nach einem Ausfall durch Nichterreichbarkeit für ein nicht auswertbares Interview ist drei mal so hoch wie für ein realisiertes Interview, ebenso wie nach einem verweiger-ten Interview. Das relative Risiko für ein nicht auswertbares Interview nach einem Ausfall durch Teilnahmeunfähigkeit liegt um das 8-fache höher, für eine erneut nicht auswertbares Interview um das 159-fache.

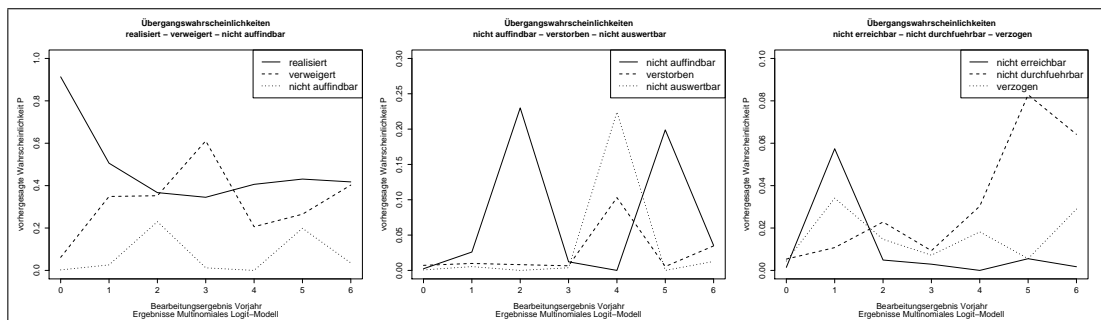
Auch für nicht durchführbare Interviews liegt das Risiko höher nach einem Ausfall in der Vorwelle: die logarithmierte Wahrscheinlichkeit, erneut durch ein nicht durchführbares Inter-

view auszufallen ist um das 112-fache höher als für ein anschließendes realisiertes Interview; nach einem nicht auswertbaren Interview ist das Risiko um das 54-fache erhöht, ähnlich, wenn Personen in der letzten Welle nicht an der alten Adresse aufgefunden wurden, in der nächsten Welle aber (an der alten oder einer neuen Adresse) wieder aufgefunden werden können: hier ist die logarithmierte Wahrscheinlichkeit um das 45-fache höher. Nach einer Verweigerung ist die logarithmierte Wahrscheinlichkeit für ein nicht durchführbares Interview um das 20-fache höher als für ein realisiertes Interview. Vergleichsweise niedrig ist das Risiko, dass auf ein realisiertes Interview ein Ausfall aufgrund von Alter oder Krankheit folgt; die logarithmierte Wahrscheinlichkeit für einen Ausfall ist „lediglich“ um das 4.2-fache höher als die logarithmierte Wahrscheinlichkeit für ein realisiertes Interview.

Ausfälle durch Tod deuten sich vor allem durch nicht auswertbare Interviews (das relative Risiko ist um das 2700-fache höher) nach einem Ausfall aufgrund von Alter oder Krankheit (die logarithmierte Wahrscheinlichkeit für Ausfall durch Tod ist um mehr als das 890-fache höher). Die logarithmierte Wahrscheinlichkeit, dass nach einer realisierten Welle ein Ausfall durch Tod folgt ist um das 85-fache höher als ein erfolgreiches Interview in der folgenden Welle. Für Ausfälle durch Nichterreichbarkeit, Verweigerungen oder nicht aufgefundene Teilnehmer in der Vorwelle beträgt die Wahrscheinlichkeit für einen Ausfall durch Tod jeweils mehr als das 200-fache der logarithmierten Wahrscheinlichkeit eines realisierten Interviews. Ist eine Beobachtung in der Vorwelle nicht erreichbar, ist das relative Risiko, in der nächsten Welle durch einen Umzug ins Ausland auszufallen das 90-fache, bei einem Ausfall durch Teilnahmeunfähigkeit das 93-fache der logarithmierten Wahrscheinlichkeit in der nächsten Welle erfolgreich interviewt zu werden, für in der Vorwelle nicht auffindbare Personen rund das 54-fache. Für Verweigerungen beträgt das relative Risiko gegenüber der Wahrscheinlichkeit eines realisierten Interviews ins Ausland zu verziehen das 28-fache. Die Wahrscheinlichkeit, nach einem nicht auswertbaren Interview durch einen Umzug das Panel endgültig zu verlassen, ist um das 60-fache höher als für ein realisiertes Interview.

Neutrale Ausfälle und Ausfälle mit unbekannter Ursache sollen hier nicht weiter beschrieben werden. Neutrale Ausfälle sind streng genommen nicht als Markov-Prozess beschreibbar; sie sind per Definition unabhängig vom Bearbeitungsergebnis im Vorjahr; Ausfälle mit unbe-

Abbildung 8.1: Vorhergesagte Übergangswahrscheinlichkeiten zwischen den Befragungsergebnissen



kanter Ursache stammen vor allem aus der zweiten Welle des SOEP im Jahr 1985; hier wird vermutet, dass sich dahinter eine Mischkategorie der verschiedenen Ausfallursachen verbirgt.

8.3 Vorhergesagte Übergangswahrscheinlichkeiten des Markov-Modells

Während die Betrachtung der Koeffizienten des multinomialen Logitmodells stark von der Wahl der Basiskategorie abhängt (und die großen Koeffizienten über die geringen Fallzahlen und Bedeutungen für einzelne Ausfallursachen hinwegtäuschen), sind die vorhergesagten Wahrscheinlichkeiten für die Kategorien der abhängigen Variable unabhängig von der Wahl der Basiskategorie. Der folgende Abschnitt zeigt daher die auf Basis des multinomialen Logitmodells der Bearbeitungsergebnisse bzw. Ausfallursachen vorhergesagten Übergangswahrscheinlichkeiten zwischen den einzelnen Kategorien, gegeben das Bearbeitungsergebnis des Vorjahres. Abbildung 8.1 zeigt die vorhergesagten Wahrscheinlichkeiten der einzelnen Ausfallursachen/Bearbeitungsergebnisse.

Im linken Teil der Abbildung 8.1 sind die Übergangswahrscheinlichkeiten für einen Übergang zu einem realisierten Interview, zu einer Verweigerung oder zu einem Ausfall durch Nichtauffindbarkeit, gegeben das Bearbeitungsergebnis der Vorwelle abgebildet. Die X-Achse ist jeweils nach dem Bearbeitungsergebnis im Vorjahr sortiert: realisiert in Vorwelle (0), nicht erreichbar in Vorwelle (1), nicht auffindbar in Vorwelle (2), verweigert in Vorwelle (3), nicht auswertbar in Vorwelle (4), unbekannte Ausfallursache in Vorwelle (5), sowie nicht

durchführbar in Vorwelle (6). Die vorhergesagte Wahrscheinlichkeit, in der nächsten Welle erneut erfolgreich interviewt zu werden liegt bei über 90%, bei einem Ausfall in der Vorwelle durch Nichterreichbarkeit bei über 50%. Ist eine Person an der alten Adress nicht auffindbar, ist die vorhergesagte Wahrscheinlichkeit für ein erfolgreiches Interview in der folgende Welle rund 40%. Auch für die restlichen Ausfallursachen in der Vorwelle (nicht auswertbar, unbekannter Ausfall, nicht durchführbar) beträgt die vorhergesagte Wahrscheinlichkeit für ein realisiertes Interview in der nächsten Welle über 40%.

Nach einem realisierten Interview ist die vorhergesagte Wahrscheinlichkeit für einen Ausfall durch Verweigerung unter 5%, beträgt aber für in der Vorwelle nicht Erreichbare schon deutlich mehr als 30% und knapp 40% nach einem Ausfall durch Nichtauffinden an der alten Adresse in der Vorwelle; dieses Ergebnis ist plausibel bspw. für mobile Personen und Personen mit knappen Zeitressourcen: werden sie nach einem Ausfall (aufgrund ihrer knappen Zeitressourcen oder Mobilität) in der nächsten Welle wieder angetroffen, ist eine Verweigerung (bspw. aufgrund knapper Zeitressourcen) vergleichsweise wahrscheinlich. Mit 60% noch höher, auch höher als die Übergangswahrscheinlichkeit zu einem realisierten Interview, ist die vorhergesagte Übergangswahrscheinlichkeit nach einer Verweigerung in der Vorwelle zu einer erneuten Verweigerung (und damit zu einem endgültigen Ausfall). Nach einem nicht auswertbaren Interview beträgt die Wahrscheinlichkeit rund 20% für einen Übergang zu einer Verweigerung in der nächsten Welle; dass auf die Feststellung einer Teilnahmeunfähigkeit (bspw. aufgrund von Alter oder Krankheit) eine Verweigerung mit einer vergleichsweise hohen Wahrscheinlichkeit folgt ist ebenfalls plausibel.

Die einzige Ausfallursache, die neben einer Verweigerung eine größere Rolle spielt ist ein Ausfall durch Nichtauffinden in der folgenden Welle, auch wenn hier die Wahrscheinlichkeit für einen Übergang maximal 20% für wiederholtes Nichtauffinden oder zu einem unbekanntem Ausfall beträgt¹⁴⁷.

Der mittlere Teil der Abbildung 8.1 zeigt die vorhergesagten Wahrscheinlichkeiten für einen Übergang nach „nicht auffindbar“, „verstorben“ und „nicht auswertbar“. Nach einem reali-

¹⁴⁷Unbekannte Ausfälle stellen offensichtlich eine Mischkategorie dar, bestehend aus Verweigerern, nicht erreichbaren Personen oder nicht auffindbaren Personen.

sierten Interview ist der Übergang zu einer dieser Ausfallursachen mit ca. 1% äusserst gering, genauso der Übergang nach einem Ausfall aufgrund von Nichterreichbarkeit oder auch nach einer Verweigerung. Die vorhergesagte Übergangswahrscheinlichkeit für einen Ausfall durch Tod ist interessanterweise nach einem vorhergehenden Ausfall durch ein nicht auswertbares Interview mit knapp 10% höher als nach einem Ausfall durch eine festgestellte Teilnahmeunfähigkeit (rund 4%); hier hätte vermutet werden können, dass nach Ausfällen aufgrund von Alter oder Krankheit auch eine höhere Wahrscheinlichkeit für einen Ausfall durch Tod folgt. Der Übergang für ein erneut nicht auswertbares Interview beträgt mehr als 20%, die Wahrscheinlichkeit für einen Übergang von einem unbekanntem Ausfall zu einem Ausfall durch Nichtauffinden knapp 20%.

Die restlichen im SOEP erfassten Ausfallursachen (nicht erreichbar, nicht durchführbar, verzogen) sind im rechten Teil von Abbildung 8.1 dargestellt; die Wahrscheinlichkeit für einen Übergang in diese Zustände nach einem realisierten Interview ist mit unter 0.1% äusserst gering. Die Wahrscheinlichkeit, durch einen Wegzug ins Ausland in der nächsten Welle endgültig auszufallen beträgt maximal rund 3% nach einem Ausfall durch nicht erreichen bzw. 2% nach einem Ausfall durch Nichtauffinden an der alten Adresse. Lediglich die Übergangswahrscheinlichkeit wiederholt nicht auffindbar zu sein liegt über 5%, und die Übergangswahrscheinlichkeit für ein nicht durchführbares Interview nach einem unbekanntem Ausfall (über 8%) spielt noch eine vergleichsweise größere Rolle.

Diese vorhergesagten Wahrscheinlichkeiten des multinomialen Logitmodells bilden nun also die Matrix der geschätzten Übergangswahrscheinlichkeiten \hat{P} :

$$\hat{P} = \begin{pmatrix} & \mathbf{0} & \mathbf{1} & \mathbf{2} & \mathbf{3} & \mathbf{4} & \mathbf{5} & \mathbf{6} & \mathbf{7} & \mathbf{8} & \mathbf{9} \\ \mathbf{0}_{n-1} & 0.91 & 0.00 & 0.00 & 0.06 & 0.00 & 0.00 & 0.01 & 0.01 & 0.00 & 0.00 \\ \mathbf{1}_{n-1} & 0.51 & 0.03 & 0.06 & 0.35 & 0.01 & 0.00 & 0.01 & 0.01 & 0.03 & 0.00 \\ \mathbf{2}_{n-1} & 0.37 & 0.23 & 0.00 & 0.35 & 0.00 & 0.00 & 0.02 & 0.01 & 0.01 & 0.00 \\ \mathbf{3}_{n-1} & 0.35 & 0.01 & 0.00 & 0.61 & 0.00 & 0.00 & 0.01 & 0.01 & 0.01 & 0.00 \\ \mathbf{4}_{n-1} & 0.41 & 0.00 & 0.00 & 0.21 & 0.22 & 0.01 & 0.03 & 0.10 & 0.02 & 0.00 \\ \mathbf{5}_{n-1} & 0.43 & 0.20 & 0.01 & 0.27 & 0.00 & 0.00 & 0.08 & 0.01 & 0.01 & 0.01 \\ \mathbf{6}_{n-1} & 0.42 & 0.04 & 0.00 & 0.40 & 0.01 & 0.00 & 0.06 & 0.03 & 0.03 & 0.00 \\ \mathbf{7}_{n-1} & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 1.00 & 0.00 & 0.00 \\ \mathbf{8}_{n-1} & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 1.00 & 0.00 \\ \mathbf{9}_{n-1} & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 1.00 \end{pmatrix} \quad (8.11)$$

Die Spalten der Matrix bezeichnen die (aktuellen) Bearbeitungsergebnisse (0: realisiert, 1: nicht auffindbar, 2: nicht erreichbar, 3: verweigert, 4: nicht auswertbar, 5: unbekannte Ausfallursache, 6: nicht durchführbar, 7: verstorben, 8: verzogen, 9: neutraler Ausfall), die Zeilen die entsprechenden Bearbeitungsergebnisse der Vorwelle; betrachtet man die Zeilen 7, 8 und 9, führen sie mit einer Wahrscheinlichkeit von 1 in denselben Zustand; dies sind die absorbierenden (endgültigen) Zustände „verstorben“, „ins Ausland verzogen“ und die „neutralen Ausfälle“. Somit erfüllt diese Matrix der Übergangswahrscheinlichkeiten alle Bedingungen einer stochastischen Matrix: alle Zellen sind nicht negativ und die Summe über jede Zeile ergibt 1 (die Wahrscheinlichkeit für irgendeinen Übergang beträgt 1). Die erste Spalte der Matrix enthält also die Übergangswahrscheinlichkeiten in ein realisiertes Interview in der aktuellen Welle, die zweite entsprechend in einen Ausfall durch Nichtauffinden des Haushalts, die dritte in einen Ausfall durch Nichterreichbarkeit trotz mehrfachen Kontaktversuche, usw. .

8.4 Bemerkungen zur Bewertung des Modellfits und Regressionsdiagnostik

Das Pseudo- R^2 (nach McFadden) des Gesamtmodells beträgt lediglich 0.097; allerdings ist dabei zu beachten, dass bei der logistischen Regression das R^2 nicht so intuitiv beurteilbar ist, wie bei der linearen Regression (vgl. bspw. Schnell, 1994, S. 253; oder Long und Freese, 2001, S. 80f, 191).

Aufgrund der großen Fallzahl wird als Alternative zum Likelihood-Ratio-Test der Koeffizienten der Wald-Test zur Beurteilung des Einflusses der unabhängigen Variable auf alle Kategorien der abhängigen Variable verwendet (vgl. Long und Freese, 2001, S. 180): alle Koeffizienten der unabhängigen Variablen des multinomialen Logitmodells sind signifikant von 0 verschieden. Kritischste und problematischste Annahme im multinomialen Logitmodell ist zudem die Annahme der Unabhängigkeit von irrelevanten alternativen Kategorien („independence from irrelevant alternatives“, IIA; vgl. Hausman und McFadden, 1984, S. 1219): Die Wahrscheinlichkeit für Kategorie j hängt nicht von den sonstigen Kategorien der abhängigen Variable ab (Greene, 2003, S. 724): „Adding or deleting outcomes does not affect the odds among the the remaining outcomes“ (Long und Freese, 2001, S. 188; vgl. auch Hausman und McFadden, 1984, S. 1220). Für den Hausman-Test werden daher das vollständige Modell und eingeschränkte Modelle ohne jeweils eine Kategorie der abhängigen Variable geschätzt¹⁴⁸. Alle Ergebnisse des (klassischen) Hausman-Tests sind nicht signifikant, d.h. alle Kategorien sind unabhängig von den restlichen Kategorien der abhängigen Variable¹⁴⁹. Spezielle diagnostische Verfahren für die multinomiale logistische Regression sind kaum oder gar nicht vorhanden, spezielle Verfahren beziehen sich hauptsächlich auf den Fit des Gesamtmodells¹⁵⁰ (vgl. Goeman und le Cessie, 2006, S. 980). Die meisten Empfehlungen für

¹⁴⁸Small und Hsiao (1985) schlagen eine Alternative zum klassischen Hausman-Test der IIA-Annahme vor, die auf Stichproben aus dem vollständigen Sample basiert; siehe auch Long und Freese (2001, S. 189f).

¹⁴⁹Eine Alternative für das multinomiale (oder auch für das conditional) Logitmodell bei Hinweisen auf Verletzung der IIA-Annahme ist das multinomiale Probitmodell (vgl. Greene, 2003, S. 728).

¹⁵⁰Goeman und le Cessie (2006) schlagen beispielsweise einen Goodness-of-Fit-Test speziell für das multinomiale Logitmodell auf Basis der Summe geglätteter quadrierter Residuen aller Kategorien vor.

eine Beurteilung des Fits für Beobachtungen unter einem multinomialen Regressionsmodell beziehen sich darauf, einzelne logistische Regressionen der einzelnen Kategorien des multinomialen Modells zu berechnen und diese zur Diagnostik heranzuziehen (Goeman und le Cessie, 2006, S. 980): „you can estimate the series of binary logits implied by the multinomial logit model and use the established methods of examining the fit of observations to binary logit estimates“ (Long und Freese, 2001, S. 191). Eine Diagnostik des verwendeten multinomialen Regressionsmodells wird dadurch weiter erschwert, da nicht nur die abhängige Variable lediglich binäre Ausprägungen besitzt, sondern auch die unabhängigen Variablen. Hinzu kommt, dass Scatterplots binärer Variablen vor allem erst mit Scatterplotsmoothern (bspw. Lowess- oder Loess-Smoothern) sinnvoll eingesetzt werden können (vgl. Schnell, 1994, S. 250f); allerdings verhindert die große Fallzahl (rund 380.000 Beobachtungen) den Einsatz von Scatterplotsmoothern wie Lowess, da dafür unter hohem Rechenaufwand¹⁵¹ dann auch 380.000 lokale Regressionen berechnet werden müssten. Eine Schwäche und Einschränkung des Modells stellen die vergleichsweise geringen Fallzahlen der einzelnen Ausfallursachen (siehe Tabelle 5.4 auf Seite 92) dar.

Tabelle A.1 im Anhang ab Seite 145 zeigt die Ergebnisse eines nichtstationären Markov-Prozesses: zusätzlich zu den Dummies der Bearbeitungsergebnisse der Vorwelle wurde die Teilnahmedauer als Kontrollvariable mit aufgenommen. Insgesamt zeigen die Ergebnisse des nichtstationären Markov-Modells, dass eine zunehmende Teilnahmedauer die Wahrscheinlichkeit für einen Ausfall verringert, lediglich die Wahrscheinlichkeit für einen Ausfall durch ein nicht auswertbares Interview und für Ausfälle durch Tod steigt mit zunehmender Teilnahmedauer. Auch der Modellfit (McFadden- R^2) verbessert sich im Vergleich zum stationären Modell auf 0.11. Allerdings ist beim nichtstationären Modell die IIA-Annahme nicht erfüllt, der Hausman-Test ist nicht signifikant.

¹⁵¹Die stata-Hilfe enthält unter dem Eintrag `lowess`: „Warning: lowess is computationally intensive and may therefore take a long time to run on a slow computer. Lowess calculations on 1,000 observations, for instance, require estimating 1,000 regressions“.

9 Diskussion der Ergebnisse und Schlussfolgerungen

Die vorhergesagten Wahrscheinlichkeiten des multinomialen Logitmodells der Ausfallursachen bestätigen die Eindrücke der Bedeutung der Ausfallursachen auf Basis der Ausfallsequenzen; aus ihnen und den geschätzten Übergangswahrscheinlichkeiten lassen sich Implikationen für die Organisation der Feldarbeit ziehen: Generell gilt, dass temporäre Ausfälle vermieden werden müssen, da sie die Wahrscheinlichkeit für ein erfolgreiches Interview in der nächsten Welle deutlich reduzieren.

Während eine erfolgreiche interviewte Person mit einer sehr hohen Wahrscheinlichkeit auch in der nächsten Welle wieder erfolgreich interviewt werden wird, besteht nach einem temporären Ausfall (im besten Fall) nur noch eine Wahrscheinlichkeit von rund 0.5, dass diese Person in der nächsten Welle erfolgreich aufgefunden, erreicht und interviewt wird; vor allem nach einer Verweigerung ist die Wahrscheinlichkeit für eine erneute Verweigerung hoch, höher als für ein realisiertes Interview; eine Verweigerung ist also als letztes Alarmsignal zu verstehen, und es müssen vom Erhebungsinstitut alle Anstrengungen unternommen werden, um Verweigerungen überhaupt, als auch erneute Verweigerungen zu vermeiden; eventuell sind hier veränderte, differenzierte Anreize in Form von zusätzlichen Incentives oder zusätzlichem Kontakt zwischen den Feldzeiten der eigentlichen Erhebung für Verweigerer in der Vorwelle möglich; eventuell verringert auch ein Einsatz speziell für den Umgang mit Verweigerern geschulter Interviewer die Wahrscheinlichkeiten für eine erneute Verweigerung.

ähnliche Probleme zeigen sich bei nicht mehr auffindbaren Befragungspersonen bzw. nicht erreichbaren Personen: für beide Personengruppen besteht eine vergleichsweise hohe Wahrscheinlichkeit von 0.3 zu verweigern, sollten sie in der nächsten Welle erreicht bzw. aufgefunden und erreicht werden. Vermehrter und verbesserter Kontakt auch zwischen den Feldzeiten verhindert eventuell Ausfälle durch einen Adressverlust, wenn Adressänderungen rechtzeitig bekannt werden, und ein Haushalt in der nächsten Welle an der neuen Adresse kontaktiert werden kann. Im Falle nicht erreichbarer Haushalte besteht zudem eine hohe Wahrscheinlichkeit, dass auch erneut kein Kontakt hergestellt werden kann. Hier müssen Anstrengungen verstärkt werden, auch schwer erreichbare Haushalte zu kontaktieren, bspw. durch mehr Kontaktversuche und stärker variierte Kontaktzeiten; hier könnte ein vermehrter und ver-

besserer Kontakt zwischen Befragten und Erhebungsinstitut die Kontaktwahrscheinlichkeit vergrößern, wenn Terminabsprachen vor der eigentlichen Feldzeit getroffen werden können. Auch wenn kein Zusammenhang zwischen der Anzahl der eingesetzten Interviewer und der Entwicklung der Anteile nicht erreichbarer oder nicht auffindbarer Personen erkennbar ist, ermöglicht eine geringerer Workload der Interviewer eine erhöhte Anstrengung, auch schwer erreichbare Befragungshaushalte zu kontaktieren.

Sind die vorhergesagten Wahrscheinlichkeiten zwar von der Idee der Propensity-Gewichte geleitet, stellen sie jedoch sicher keine adäquate Form der Ausfallgewichte dar, da das hier verwendete Markov-Modell den eigentlichen Ausfallmechanismus nicht abbildet. Das hier verwendete Modell zeigt aber, dass für unterschiedliche Arten von Ausfällen unterschiedliche Wahrscheinlichkeiten für eine weitere Teilnahme bestehen. Werden diese unterschiedlichen Ausfallursachen bei einer Korrektur von Nonresponse nicht unterschieden, vermischen sich diese unterschiedlichen Teilnahmewahrscheinlichkeiten. Eine Verzerrung durch Nonresponse besteht dann möglicherweise weiter, wenn bestimmte Gruppen eine erhöhte Wahrscheinlichkeit für bestimmte Ausfallarten haben und eine Korrektur lediglich auf einer einfachen Unterscheidung zwischen Respondenten und Nonrespondenten basiert.

Literatur

- AHERN, KATHY und LE BROCCQUE, ROBYNE (2005): Methodological Issues in the Effects of Attrition: Simple Solutions for Social Scientists. *Field Methods* 17(1): 53–69.
- ALDERMAN, HAROLD; BEHRMAN, JERE R.; KOHLER, HANS-PETER; MALUCCIO, JOHN A. und COTTS WATKINS, SUSAN (2001): Attrition in Longitudinal Household Survey Data. *Demographic Research* 5(4): 79–124.
- ARBEITSKREIS DEUTSCHER MARKT- UND SOZIALFORSCHUNGSINSTITUTE (ADM) und ARBEITSGEMEINSCHAFT MEDIA ANALYSE (AG.MA) (Hg.) (1999): *Stichproben-Verfahren in der Umfrageforschung. Eine Darstellung für die Praxis*. Opladen: Leske und Budrich.
- ATROSTIC, B. K.; BATES, NANCY; BURT, GERALDINE und SILBERSTEIN, ADRIANA (2001): Nonresponse in U.S. Government Household Surveys: Consistent Measures, Recent Trends, and New Insights. *Journal of Official Statistics* 17(2): 209–226.
- BAKEMAN, ROGER und GOTTMAN, JOHN MORDECHAI (1997): *Observing Interaction: An Introduction to Sequential Analysis*. Cambridge: Cambridge University Press.
- BEHR, ANDREAS; BELLGARDT, EGON und RENDTEL, ULLRICH (2005): Extent and Determinants of Panel Attrition in the European Community Household Panel. *European Sociological Review* 21(5): 489–512.
- BOSE, JONAKI und WEST, JERRY (2002): Examining Additional Nonresponse Bias introduced through Attrition. In: *Proceedings of the Survey Research Methods Section*. Alexandria: American Statistical Association, 278–283. Internetressource: <http://www.amstat.org/sections/SRMS/Proceedings/y2002/Files/JSM2002-000825.pdf> (12. Januar 2007).
- BURKAM, DAVID T. und LEE, VALERIE E. (1998): Effects of Monotone and Nonmonotone Attrition on Parameter Estimates in Regression Models with Educational Data: Demographic Effects on Achievement, Aspirations and Attitudes. *The Journal of Human Resources* 33(2): 555–574.

- BURKHAUSER, RICHARD V.; KREYENFELD, MICHAELA und WAGNER, GERT G. (1997): The Immigrant Sample of the German Socio-economic Panel. *Maxwell Center for Demography and Economics of Aging. Aging Studies Program Paper* (7). Online verfügbar unter <http://www-cpr.maxwell.syr.edu/agpapper/pdf/age7.pdf> (13. März 2007).
- COCHRAN, WILLIAM G. (1983): Historical Perspective. In: MADOW, WILLIAM G.; OLKIN, INGRAM und RUBIN, DONALD B. (Hg.), *Incomplete Data in Sample Surveys. Theory and Bibliographies*, New York: Academic Press, Band 2. 11–25.
- COHEN, G. und DUFFY, J.C. (2002): Are Nonrespondents to Health Surveys Less Healthy Than Respondents. *Journal of Official Statistics* 18(1): 13–23.
- CSENKI, ATTILA (1994): *Dependability for Systems with a Partitioned State Space*, Band 90 von *Lecture Notes in Statistics*. New York: Springer.
- CURTIN, RICHARD; PRESSER, STANLEY und SINGER, ELEANOR (2005): Changes in Telephone Survey Nonresponse Over the Past Quarter Century. *Public Opinion Quarterly* 69(1): 87–98.
- DE HEER, WIM (1999): International Response Trends: Results of an International Survey. *Journal of Official Statistics* 15(2): 129–142.
- DE LEEUW, EDITH und DE HEER, WIM (2002): Trends in Household Survey Nonresponse: A Longitudinal and International Comparison. In: GROVES, ROBERT M.; DILLMAN, DON A.; ELTINGE, JOHN L. und LITTLE, RODERICK J.A. (Hg.), *Survey Nonresponse*, New York: Wiley, Wiley Series in Probability and Statistics. 41–54.
- DEMAIO, THERESA J. (1980): Refusals: Who, Where, Why? *Public Opinion Quarterly* 44(2): 223–233.
- DIGGLE, PETER J.; LIANG, KUNG-YEE und ZEGER, SCOTT L. (1994): *Analysis of Longitudinal Data*. Oxford Statistical Science Series. Oxford: Clarendon Press.
- DILLMAN, DON A.; ELTINGE, JOHN L.; GROVES, ROBERT M. und LITTLE, RODERICK J.A. (2002): Survey Nonresponse in Design, Data Collection and Analysis. In:

- GROVES, ROBERT M.; DILLMAN, DON A.; ELTINGE, JON L. und LITTLE, J.A., RODERICK (Hg.), *Survey Nonresponse*, New York: Wiley, Wiley Series in Probability and Statistics. 3–26.
- DJERF, KARI (2004): Nonresponse in Time: A Time Series Analysis of the Finnish Labor Force Survey. *Journal of Official Statistics* 20(1): 39–54.
- FAY, ROBERT E. (1986): Causal Models for Patterns of Nonresponse. *Journal of the American Statistical Association* 81(394): 354–365.
- FITZGERALD, JOHN; GOTTSCHALK, PETER und MOFFITT, ROBERT (1998): An Analysis of Sample Attrition in Panel Data: The Michigan Panel Study of Income Dynamics. *The Journal of Human Resources* 33(2): 251–299.
- GABLER, SIEGFRIED (Hg.) (1998): *Telefonstichproben in Deutschland*. Opladen: Westdeutscher Verlag.
- GALLAGHER, PATRICIA M.; FOWLER, FLOYD JACKSON und STRINGFELLOW, VICKIE L. (2005): The Nature of Nonresponse in a Medicaid Survey: Causes and Consequences. *Journal of Official Statistics* 21(1): 73–87.
- GALVIN, LIEU N.; SAE-UNG, SAM und KING, KAREN (2000): Effect of Interview Length and Proxy Interviews on Attrition to the Survey of Income and Program Participation. In: *Proceedings of the Survey Research Methods Section*. Alexandria: American Statistical Association, 636–640. Internetressource: http://www.amstat.org/sections/SRMS/Proceedings/papers/2000_106.pdf (12. Januar 2007)
- GOEMAN, JELLE J. und LE CESSIE, SASKIA (2006): A Goodness-of-Fit Test for Multinomial Logistic Regression. *Biometrics* 62: 980–985.
- GOODMAN, JODI S. und BLUM, TERRY C. (1996): Assessing the Non-Random Sampling Effects of Subject Attrition in Longitudinal Research. *Journal of Management* 22(4): 627–652.

- GOTTMAN, JOHN MORDECHAI und ROY, ANUP KUMAR (1990): *Sequential Analysis: a Guide for Behavioral Researchers*. New York: Cambridge University Press.
- GOYDER, JOHN; WARRINER, KEITH und MILLER, SUSAN (2002): Evaluating Socio-Economic Status (SES) Bias in Survey Nonresponse. *Journal of Official Statistics* 18(1): 1–11.
- GRAY, REBECCA; CAMPANELLI, PAMELA; DEEPCHAND, KAVITA und PRESCOTT-CLARKE, PATRICIA (1996): Exploring Survey Non-Response: The Effect of Attrition on a Follow-Up of the 1984-85 Health and Life Style Survey. *The Statistician* 45(2): 163–183.
- GREENE, WILLIAM H. (2003): *Econometric Analysis. International Edition*. Upper Saddle River: Pearson Education.
- GROVES, ROBERT M. (2006): Nonresponse Rates and Nonresponse Bias in Household Surveys. *Public Opinion Quarterly* 70(5): 646–675.
- GROVES, ROBERT M.; CIALDINI, ROBERT B. und COUPER, MICK P. (1992): Understanding the Decision to Participate in a Survey. *Public Opinion Quarterly* 56(4): 475–495.
- GROVES, ROBERT M. und COUPER, MICK P. (1996): Contact-Level Influences on Cooperation in Face-to-Face Surveys. *Journal of Official Statistics* 12(1): 63–83.
- GROVES, ROBERT M. und COUPER, MICK P. (1998): *Nonresponse in Household Interview Surveys*. Wiley Series in Probability and Statistics. New York: Wiley.
- GROVES, ROBERT M.; FOWLER, FLOYD J.; COUPER, MICK P.; LEPOWSKI, JAMES M.; SINGER, ELEANOR und TOURANGEAU, ROGER (2004): *Survey Methodology*. Wiley Series in Survey Methodology. Hoboken: Wiley.
- GROVES, ROBERT M.; SINGER, ELEANOR; CORNING, AMY D. und BOWERS, ASHLEY (1999): A Laboratory Approach to Measuring the Effects on Survey Participation of Interview Length, Incentives, Differential Incentives, and Refusal Conversion. *Journal of Official Statistics* 15(2): 251–268.

- HAIKEN DENNEW, JOHN und FRICK, JOACHIM (2005): *DTC Desktop Companion to the German Socio-Economic Panel (SOEP). Version 8.0 Updated to Wave 21*. Berlin: DIW. Internetressource: <http://www.diw-berlin.de/deutsch/sop/service/dtc/dtc.pdf> (26. November 2006).
- HAUSMAN, JERRY und MCFADDEN, DANIEL (1984): Specification Tests for the Multinomial Logit Model. *Econometrica* 52(5): 1219–1240.
- HAWKES, DENISE und PLEWIS, IAN (2006): Modelling non-response in the National Child Development Study. *Journal of the Royal Statistical Society (Series A)* 169(3): 479–491.
- HECKMAN, JAMES J. (1979): Sample Selection Bias as Specification Error. *Econometrica* 47(1): 153–161.
- HILL, DANIEL HENRY und WILLIS, ROBERT J. (2001): Reducing Panel Attrition: A Search for Effective Policy Instruments. *The Journal of Human Resources* 36(3): 416–438.
- HILL, ZELEE (2002): Reducing Panel Attrition in Panel Studies in Developing Countries. *Young Lives Workingpaper (Department of International Development, Oxford University)* 5: 23–44. Internetressource: <http://younglives.qeh.ox.ac.uk/pdf/wp5.pdf> (10. März 2007).
- HILL, ZELEE (2004): Reducing Attrition in Panel Studies in Developing Countries. *International Journal of Epidemiology* 33(3): 493–498.
- HONGGAO, CAO und HILL, DANIEL H. (2005): *Active versus Passive Sample Attrition: The Health Retirement Study*. Ann Arbor: Institute for Social Research. Universität Michigan. Internetressource: <http://ideas.repec.org/p/wpa/wuwpem/0505006.html> (16. Mai 2007).
- HUMBERT, ANDREAS und SCHNEIDERHEINZE, KLAUS (2000): Stichprobenziehung für telefonische Zuwandererumfragen. Einsatzmöglichkeiten der Namenforschung (Onomastik). *ZUMA-Nachrichten* 24(47): 36–64. Online verfügbar unter http://www.gesis.org/Publikationen/Zeitschriften/ZUMA_Nachrichten/documents/pdfs/47/zn47_07-schneiderheinze.pdf (13. März 2007).

- INFRATEST SOZIALFORSCHUNG (2001): *Erprobung innovativer Erhebungskonzepte für Haushalts-Panel-Stichproben. Erstbefragung 2000 der SOEP-Stichprobe F. Methodenbericht*. München: Infratest Sozialforschung. Internetressource: http://www.diw.de/deutsch/sop/service/doku/meth_2000_stpr_f.pdf (13. März 2007).
- INFRATEST SOZIALFORSCHUNG (2002a): *Lebenslage und Vermögensbildung von Haushalten im oberen Einkommensbereich. Sondererhebung im Rahmen des SOEP 2002. Methodenbericht*. München: Infratest Sozialforschung.
- INFRATEST SOZIALFORSCHUNG (2002b): *Verbesserung der Datengrundlagen für Mortalitäts- und Mobilitätsanalysen: Verbleibstudie bei Panellausfällen im SOEP*. München: Infratest Sozialforschung. Internetressource: <http://diw.de/deutsch/sop/service/doku/verbleibstudie486.pdf> (25. November 2006).
- ISAACSON, DEAN L. und MADSEN, RICHARD W. (1976): *Markov Chains. Theory and Applications*. Wiley Series in Probability and Mathematical Statistics. New York: Wiley.
- JONES, ANREW M.; KOOLMAN, XANDER und RICE, NIGEL (2006): Health-related non-response in the British Household Panel Survey and European Community Household Panel: using inverse-probability-weighted estimators in non-linear models. *Journal of the Royal Statistical Society (Series A)* 169(3): 543–569.
- KALTON, GRAHAM (1986): Handling Wave Nonresponse in Panel Surveys. *Journal of Official Statistics* 2(3): 303–314.
- KALTON, GRAHAM und FLORES-CERVANTES, ISMAEL (2003): Weighting Methods. *Journal of Official Statistics* 19(2): 81–97.
- KAUTTER, JOHN; KHATUTSKY, GALINA; POPE, GREGORY C.; CHROMY, JAMES R. und ADLER, GERALD S. (2006): Impact of Nonresponse on Medicare Current Beneficiary Survey Estimates. *Health Care Financing Review* 27(4): 71–93.
- KENNICHELL, ARTHUR B. (1999): Analysis of Nonresponse Effects in the 1995 Survey of Consumer Finance. *Journal of Official Statistics* 15(2): 283–303.

- LAURIE, HEATHER; SMITH, RACHEL und SCOTT, LYANNE (1999): Strategies for Reducing Nonresponse in an Longitudinal Panel Survey. *Journal of Official Statistics* 15(2): 269–282.
- LEPKOWSKI, JAMES M. und COUPER, MICK P. (2002): Nonresponse in the Second Wave of Longitudinal Household Surveys. In: GROVES, ROBERT M.; DILLMAN, DON A.; ELTINGE, JON L. und LITTLE, J.A., RODERICK (Hg.), *Survey Nonresponse*, Wiley, Wiley Series in Probability and Statistics. 259–272.
- LILLARD, LEE A. und PANIS, CONSTANTIJN W. A. (1998): Panel Attrition from the Panel Study of Income Dynamics: Household Income, Marital Status, and Mortality. *The Journal of Human Resources* 33(2): 437–457.
- LIPPS, OLIVER (2006): Attrition in the Swiss Household Panel: Wave 2 Through Wave 7. *Swiss Household Panel Working Paper* 4/06. Internetressource: http://www.swisspanel.ch/file/working_papers/WP4_06.pdf (27. November 2006)
- LITTLE, RODERICK J.A. (1995): Modelling the Drop-Out Mechanism in Repeated-Measures Studies. *Journal of the American Statistical Association* 90(431): 1112–1121.
- LITTLE, RODERICK J.A. und RUBIN, DONALD B. (1987): *Statistical Analysis with Missing Data*. New York: Wiley.
- LITTLE, TODD D.; LINDENBERGER, ULMAN und MAIER, HEINER (2000): Selectivity and Generalizability in Longitudinal Research: On the Effects of Continuers and Dropouts. In: LITTLE, TODD D.; SCHNABEL, KAI U. und BAUMER, JÜRGEN (Hg.), *Modeling Longitudinal and Multilevel Data. Practical Issues, Applied Approaches and Specific Examples*, London: Erlbaum, Kapitel 10. 177–200.
- LONG, JOHN S. und FREESE, JEREMY (2001): *Regression Models for Categorical Dependent Variables Using stata*. College Station: Stata Press.
- LONGFORD, NICHOLAS T. (2005): *Missing Data and Small-Area Estimation. Modern Analytical Equipment for the Survey Statistician*. New York: Springer.

- LONGFORD, NICHOLAS T.; TYRER, P.; NUR, U.A.M. und SEIVEWRIGHT, H. (2006): Analysis of a long-term study of neurotic disorder, with insights into the process of non-response. *Journal of the Royal Statistical Society (Series A)* 169(3): 507–523.
- LYBERG, INGRID und LYBERG, LARS (1991): Nonresponse Research at Statistics Sweden. In: *Proceedings of the Survey Research Methods Section*. Alexandria: American Statistical Association, 78–87. Internetressource: http://www.amstat.org/Sections/Srms/Proceedings/papers/1991_012.pdf (14. März 2007).
- LYNN, PETER (2006): Editorial: Attrition and non-response. *Journal of the Royal Statistical Society (Series A)* 169(3): 393–394.
- MADOW, WILLIAM G. und SINGH, BAHADUR (1983): Nonresponse and Double Sampling. Introduction. In: MADOW, WILLIAM G.; OLKIN, INGRAM und RUBIN, DONALD B. (Hg.), *Incomplete Data in Sample Surveys. Theory and Bibliographies*, New York: Academic Press, Band 2. 93–95.
- MARINI, MARGARET MOONEY; OLSEN, ANTHONY R. und RUBIN, DONALD B. (1980): Maximum-Likelihood Estimation in Panel Studies with Missing Data. *Sociological Methodology* 11: 314–357.
- MCGUIGAN, KIMBERLY A.; ELLICKSON, PHYLLIS L.; HAYS, RONALD D. und BELL, ROBERT M. (1995): Tracking, Weighting, and Sample Selection Modeling to Correct for Attrition. In: *Proceedings of the Survey Research Methods Section*. Alexandria: American Statistical Association, 402–407. Internetressource: http://www.amstat.org/Sections/Srms/Proceedings/papers/1995_068.pdf (26. November 2006)
- NICOLETTI, CHETI und PERACCHI, FRANCO (2005): Survey Response and Survey Characteristics: Microlevel Evidence from the European Community Household Panel. *Journal of the Royal Statistical Society (Series A)* 168(4): 763–781.
- NORRIS, JAMES R. (1997): *Markov Chains*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

- O'BRIEN, EILEEN M.; MAYER, THOMAS S.; GROVES, ROBERT M. und O'NEILL, GRACE E. (2002): Interviewer Training to Increase Survey Participation. In: *Proceedings of the Survey Research Methods Section*. Alexandria: American Statistical Association, 2502–2507. Internetressource: <http://www.amstat.org/Sections/Srms/Proceedings/y2002/Files/JSM2002-000530.pdf> (26. November 2006).
- OH, H. LOCK und SCHEUREN, FREDERICK (1983): Weighting Adjustment for Unit Non-response. In: MADOW, WILLIAM G.; OLKIN, INGRAM und RUBIN, DONALD B. (Hg.), *Incomplete Data in Sample Surveys. Theory and Bibliographies*, New York: Academic Press, Band 2. 143–184.
- O'MUIRCHEARTAIGH, COLM und CAMPANELLI, PAMELA (1999): A Multilevel Exploration of the Role of Interviewers in Survey Non-Response. *Journal of the Royal Statistical Society. Series A* 162(3): 437–446.
- PERACCHI, FRANCO (2002): The European Community Household Panel: A Review. *Empirical Economics* 27(1): 63–90.
- RAZAFINDRATSIMA, NICOLAS und KISHIMBA, NGOY (2004): Attrition in the COCON Cohort Between 2000 and 2002. *Population* 59(3-4): 357–385.
- RENDTEL, ULRICH und PROJEKTGRUPPE SOZIO-OEKONOMISCHES PANEL (1995): *Lebenslagen im Wandel: Panellausfälle und Panelrepräsentativität*. Frankfurt: Campus.
- RUBIN, DONALD B. (1987): *Multiple Imputation for Nonresponse in Surveys*. New York: Wiley.
- RUSPINI, ELISABETTA (2002): *Introduction to Longitudinal Research*. London: Routledge.
- SALENTIN, KURT (1999): Die Stichprobenziehung bei Zuwandererbefragungen. *ZUMA-Nachrichten* 23(45): 115–135. Online verfügbar unter http://www.social-science-geis.de/Publikationen/Zeitschriften/ZUMA_Nachrichten/documents/pdfs/45/zn45_09-salentin.pdf (13. März 2007).

- SÄRNDAL, CARL-ERIK und LUNDSTRÖM, SIXTEN (2005): *Estimation in Surverys with Non-response*. Wiley Series in Survey Methodology. Hoboken: Wiley.
- SCHAFFER, J.L. (1997): *Analysis of Incomplete Multivariate Data*. London: Chapman and Hall.
- SCHNELL, RAINER (1994): *Graphisch gestützte Datenanalyse*. München: Oldenbourg.
- SCHNELL, RAINER (1997): *Nonresponse in Bevölkerungsumfragen. Ausmaß, Entwicklung und Ursachen*. Opladen: Leske und Budrich.
- SCHNELL, RAINER; HILL, PAUL B. und ESSER, ELKE (1999): *Methoden der empirischen Sozialforschung*. München: Oldenbourg.
- SCHNELL, RAINER und TRAPPMANN, MARK (2006): Konsequenzen der Panelmortalität im SOEP für Schätzungen der Lebenserwartung. *Arbeitspapiere des Zentrums für Quantitative Methoden und Surveyforschung* 2/2006. Internetressource: http://www.uni-konstanz.de/FuF/Verwiss/Schnell/schnell_tote_100306.pdf (10. Februar 2007).
- SCHUPP, JÜRGEN; GRAMLICH, TOBIAS; ISENGARD, BETTINA; PISCHNER, RAINER; WAGNER, GERT G. und VON ROSENBLADT, BERNHARDT (2003): Repräsentative Analyse der Lebenslagen einkommensstarker Haushalte. In: BUNDESMINISTERIUM FÜR GESUNDHEIT UND SOZIALE SICHERUNG (BMGS) (Hg.), *Lebenslagen in Deutschland. Die Armuts- und Reichtumsberichterstattung der Bundesregierung*, Berlin: DIW. Online verfügbar unter: <http://www.bmas.bund.de/BMAS/Redaktion/Pdf/Publikationen/forschungsprojekt-repraesentative-analyse-der-lebenslagen,property=pdf,bereich=bmas,sprache=de,rwb=true.pdf> (27. Oktober 2006).
- SCOTT, JOHN (1991): *Social Network Analysis. A Handbook*. Newbury Park: Sage.
- SINGER, ELEANOR (2002): The Use of Incentives to Reduce Nonresponse in Household Surveys. In: GROVES, ROBERT M.; DILLMAN, DON A.; ELTINGE, JON L. und LITTLE, J.A., RODERICK (Hg.), *Survey Nonresponse*, New York: Wiley, Wiley Series in Probability and Statistics. 163–177.

- SINGER, ELEANOR (2006): Introduction. Nonresponse Bias in Household Surveys. *Public Opinion Quarterly* 70(5): 637–645.
- SINGH, CHARANJIT (1995): A Comparative Analysis of Attrition in Household Panels Studies. *Pa(nel) Co(mparability) Workingpaper (CEPS/INSTEAD Luxemburg)* 10. Internetressource: <http://www.ceps.lu/pdf/7/art930.pdf> (10. März 2007).
- SMALL, KENNETH A. und HSIAO, CHENG (1985): Multinomial Logit Specification Tests. *International Economic Review* 26(3): 619–627.
- SOEP-GRUPPE (2001): The German Socio-Economic Panel (GSOEP) after more than 15 Years - Overview. *Vierteljahreshefte zur Wirtschaftsforschung* 70(1): 7–14. Online verfügbar unter http://www.diw.de/deutsch/produkte/publikationen/vierteljahrshefte/docs/papers/v_01_1_2.pdf (13. März 2007).
- STEEH, CHARLOTTE G. (1981): Trends in Nonresponse Rates 1952-1979. *Public Opinion Quarterly* 45(1): 40–57.
- TARIS, TOON (2000): *A Primer in Longitudinal Data Analysis*. London: Sage.
- THOMAS, DUNCAN; FRANKENBERG, ELIZABETH und SMITH, JAMES P. (2001): Lost but Not Forgotten: Attrition and Follow-up in the Indonesia Family Life Survey. *The Journal of Human Resources* 36(3): 556–592.
- VAN DEN BERG, GERARD J.; LINDEBOOM, MAARTEN und DOLTON, PETER J. (2006): Survey non-response and the duration of unemployment. *Journal of the Royal Statistical Society (Series A)* 169(3): 585–604.
- WAGNER, GERT; SCHUPP, JÜRGEN und RENDTEL, ULLRICH (1994): Das Sozio-oekonomische Panel (SOEP) - Methoden der Datenproduktion und -aufbereitung im Längsschnitt. In: HAUSER, RICHARD; OTT, NOTTBURGA und WAGNER, GERT (Hg.), *Mikroanalytische Grundlagen der Gesellschaftspolitik. Erhebungsverfahren, Analysemethoden und Mikrosimulation*, Berlin: Akademie-Verlag, Band 2. 70–112.

WATERTON, JENNIFER und LIEVESLEY, DENISE (1987): Attrition in a Panel Study of Attitudes. *Journal of Official Statistics* 3(3): 267–282.

ZABEL, JEFFREY E. (1998): An Analysis of Attrition in the Panel Study of Income Dynamics and the Survey of Income and Program Participation with an Application to a Model of Labor Market Behavior. *The Journal of Human Resources* 33(2): 479–506.

Datumsangaben in Klammern: Datum des letzten Zugriffs

A Anhang

A.1 Abbildungen

Abbildung A.1: Entwicklung und Veränderung der Haushaltsgröße, 1985-2005

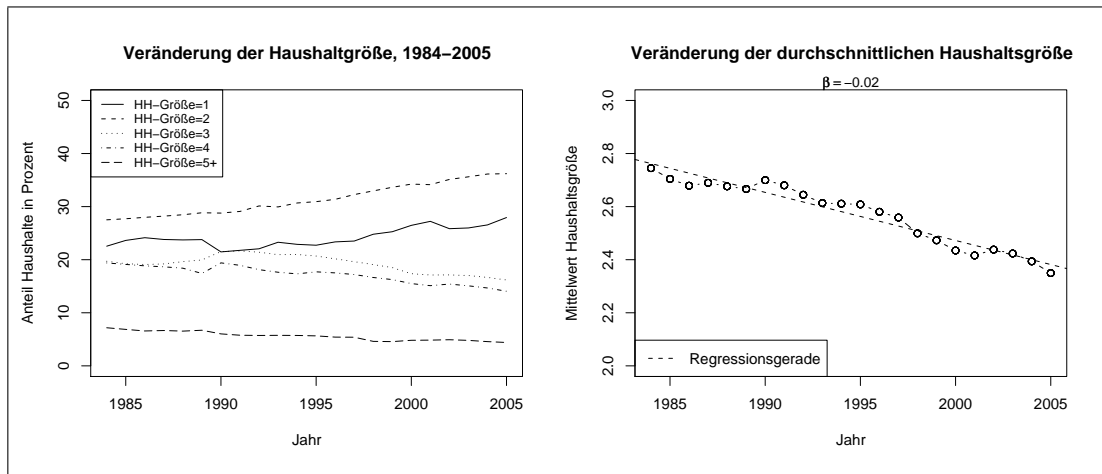
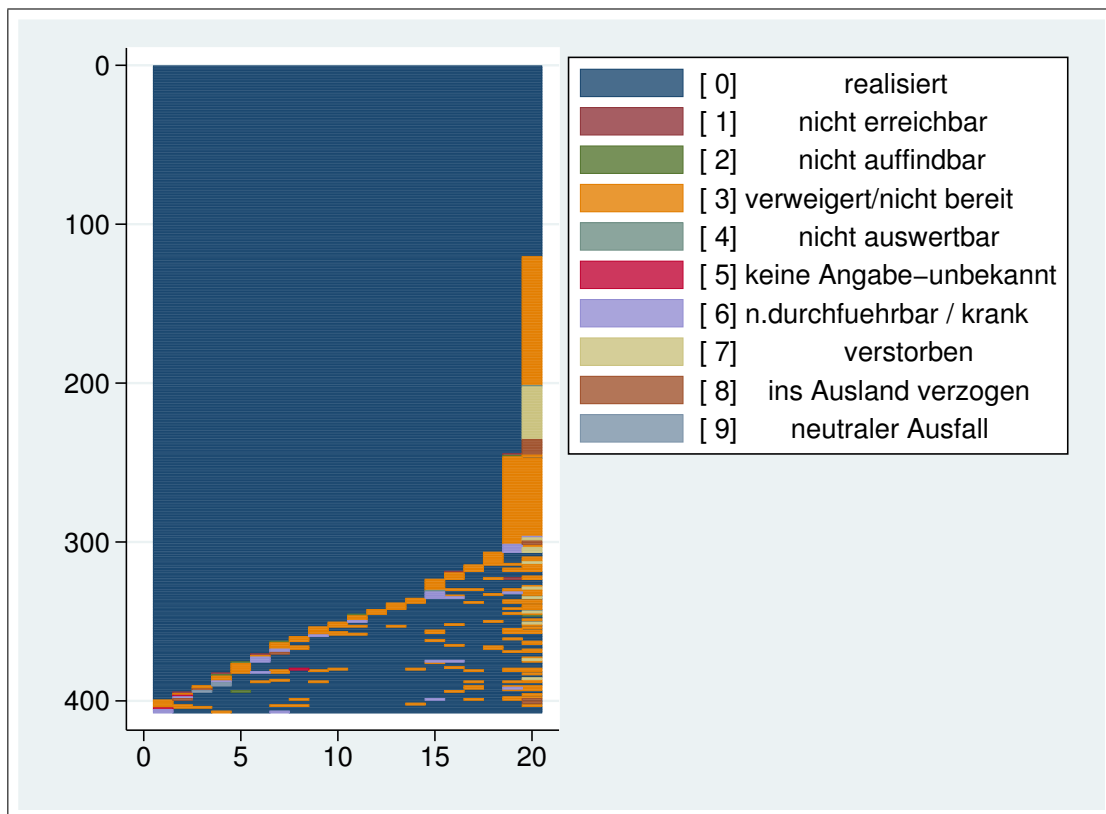


Abbildung A.2: Sequenz-Index-Plot der Teilnahmesequenzen mit Sequenzlänge 20



A.2 Tabellen

Tabelle A.1: Multinomiales Logit-Modell: Bearbeitungsergebnis - Bearbeitungsergebnis Vorwelle/Teilnahmejahr

Variable	Koeffizient	Standardfehler
Gleichung 1: nicht erreichbar		
realisiert Vorwelle	-1.784***	0.080
nicht erreichbar Vorwelle	2.241***	0.140
nicht auffindbar Vorwelle	0.296	0.585
verweigert Vorwelle	-0.124	0.179
nicht auswertbar Vorwelle ^a	-52.493	0.000
nicht durchführbar Vorwelle	-0.847	0.582
unbekannt Vorwelle	-0.019	1.007
Laufendnummer, Anzahl Jahre	-0.085***	0.010
Konstante	-4.073***	0.041
Gleichung 2: nicht auffindbar		
realisiert Vorwelle	2.128***	0.211
nicht erreichbar Vorwelle	4.818***	0.280
nicht auffindbar Vorwelle	7.552***	0.234
verweigert Vorwelle	4.701***	0.223
nicht auswertbar Vorwelle ^a	-50.945	0.000
nicht durchführbar Vorwelle	5.573***	0.247
unbekannt Vorwelle	6.924***	0.287
Laufendnummer, Anzahl Jahre	-0.103***	0.008
Konstante	-7.381***	0.204
Gleichung 3: verweigert		
realisiert Vorwelle	0.211***	0.022
nicht erreichbar Vorwelle	2.403***	0.069
nicht auffindbar Vorwelle	2.829***	0.098
verweigert Vorwelle	3.464***	0.028
nicht auswertbar Vorwelle	2.114***	0.212
nicht durchführbar Vorwelle	2.848***	0.057
unbekannt Vorwelle	2.241***	0.184
Laufendnummer, Anzahl Jahre	-0.041***	0.001
Konstante	-2.598***	0.019
Gleichung 4: nicht auswertbar		
realisiert Vorwelle	-1.671***	0.131
nicht erreichbar Vorwelle	1.032**	0.420
nicht auffindbar Vorwelle ^a	-44.237	0.000
verweigert Vorwelle	0.982***	0.185
nicht auswertbar Vorwelle	4.967***	0.225
nicht durchführbar Vorwelle	2.021***	0.241

Fortsetzung auf folgender Seite...

... Fortsetzung Tabelle A.1

Variable	Koeffizient	Standardfehler
unbekannt Vorwelle ^a	-43.271	0.000
Laufendnummer, Anzahl Jahre	0.023**	0.010
Konstante	-5.683***	0.083
Gleichung 5: nicht durchführbar		
realisiert Vorwelle	1.949***	0.135
nicht erreichbar Vorwelle	2.991***	0.320
nicht auffindbar Vorwelle	4.257***	0.306
verweigert Vorwelle	3.449***	0.163
nicht auswertbar Vorwelle	4.274***	0.482
nicht durchführbar Vorwelle	5.184***	0.168
unbekannt Vorwelle	5.113***	0.311
Laufendnummer, Anzahl Jahre	-0.084***	0.005
Konstante	-6.501***	0.130
Gleichung 6: verstorben		
realisiert Vorwelle	4.230***	0.501
nicht erreichbar Vorwelle	5.239***	0.586
nicht auffindbar Vorwelle	5.303***	0.675
verweigert Vorwelle	5.116***	0.513
nicht auswertbar Vorwelle	7.785***	0.569
nicht durchführbar Vorwelle	6.608***	0.518
unbekannt Vorwelle	4.866***	1.124
Laufendnummer, Anzahl Jahre	0.028***	0.004
Konstante	-9.314***	0.500
Gleichung 7: verzogen		
realisiert Vorwelle	2.099***	0.182
nicht erreichbar Vorwelle	4.619***	0.244
nicht auffindbar Vorwelle	4.178***	0.384
verweigert Vorwelle	3.545***	0.209
nicht auswertbar Vorwelle	4.226***	0.616
nicht durchführbar Vorwelle	4.739***	0.231
unbekannt Vorwelle	2.918***	1.022
Laufendnummer, Anzahl Jahre	-0.035***	0.005
Konstante	-7.166***	0.177
Gleichung 8: neutraler Ausfall		
realisiert Vorwelle	4.560***	0.725
nicht erreichbar Vorwelle	5.652***	1.005
nicht auffindbar Vorwelle ^a	-37.110	0.000
verweigert Vorwelle	6.989***	0.749
nicht auswertbar Vorwelle ^a	-50.684	0.000
nicht durchführbar Vorwelle	7.447***	0.886
unbekannt Vorwelle	7.122***	1.238
Laufendnummer, Anzahl Jahre	-0.797***	0.078

Fortsetzung auf folgender Seite...

... Fortsetzung Tabelle A.1

Variable	Koeffizient	Standardfehler
Konstante	-9.132***	0.712
Gleichung 9: keine Angabe, unbekannt		
realisiert Vorwelle	5.676***	0.186
nicht erreichbar Vorwelle ^a	-34.945	0.000
nicht auffindbar Vorwelle ^a	-36.669	0.000
verweigert Vorwelle	3.602***	0.731
nicht auswertbar Vorwelle	6.520***	0.744
nicht durchführbar Vorwelle ^a	-31.630	0.000
unbekannt Vorwelle ^a	-37.649	0.000
Laufendnummer, Anzahl Jahre	-2.707***	0.117
Konstante	-3.857***	0.178
N	379347	
Log-likelihood	-154234.96	
$\chi^2_{(72)}$	39366.23	
Pseudo-R ² (McFadden)	0.113	
Signifikanzniveau : * : 10% ** : 5% *** : 1%		