

SOEP 2006
Erstbefragung der
Ergänzungsstichprobe H
Methodenbericht

SOEP 2006
Erstbefragung der Ergänzungsstichprobe H

Methodenbericht

Bernhard von Rosenblatt
Nico A. Siegel
Salma Stimmel
Fritz Stutz

TNS Infratest Sozialforschung

München, Juni 2007
MB20.49191

Inhalt	Seite
1. Ziele und Anlage der Erhebung	4
2. Stichprobenanlage	6
2.1 Grundgesamtheit und Auswahlverfahren	6
2.2 Auswahl der Haushalte über getrennte Adressenerhebung	9
2.3 Die realisierte Nettostichprobe	11
3. Durchführung der Erhebung	12
3.1 Interviewereinsatz und Interviewmethode	12
3.2 Zeitlicher Verlauf der Feldarbeit	13
3.3 Stichprobenausschöpfung auf der Haushalts- und Personenebene	15
4. Stichprobenstrukturen und Non-Response Informationen	19
5. Datenprüfung und übergebener Datenbestand	29

1. Ziele und Anlage der Erhebung

Das Sozio-ökonomischen Panel (SOEP) ist eine repräsentative Langzeitbefragung privater Haushalte in Deutschland, durchgeführt seit 1984, deren Datenbestand den Wirtschafts- und Sozialwissenschaften in Deutschland und weltweit für empirische Analysen zu einem breiten Themenspektrum zur Verfügung steht. Die Verantwortung für das Projekt liegt bei der SOEP-Gruppe am DIW unter Leitung von Prof. Gert G. Wagner. Kooperationsinstitut für die Datenerhebung ist seit der ersten Erhebung im Jahr 1984 die Infratest Sozialforschung, München.¹

Die Ausgangsstichproben des Jahres 1984 (Stichprobe A: Deutsche Haushalte, Stichprobe B: Ausländer-Haushalte) wurden im Laufe der Jahre sukzessive durch neue Stichproben ergänzt. Diese können unterschiedliche Funktionen haben:

- Sie bilden entweder neu hinzukommende oder spezielle Grundgesamtheiten ab (Stichprobe C: Ostdeutsche Haushalte, seit 1990; Stichprobe D: Zuwanderer der Dekade 1984-1994, seit 1995; Stichprobe G: Einkommensstarke Haushalte, seit 2002)
- oder sie dienen der Aufstockung des Stichprobenumfangs, wobei dies jeweils mit spezifischen methodischen Innovationen verknüpft war (Stichprobe E, seit 1998, Einführung der CAPI-Technik als Interviewmethode; Stichprobe F, seit 2000, als Basis für die Erprobung innovativer Erhebungskonzepte in Haushaltspanel-Stichproben).

Im Jahr 2006 kommt als weitere Ergänzungsstichprobe die neue **Stichprobe H** hinzu. Ihrer Funktion nach gehört sie zur zweitgenannten Gruppe. Es handelt sich um einen repräsentativen Querschnitt der privaten Haushalte in Deutschland im Jahr 2006, mit dem die Stichprobenbasis des SOEP um rd. 1.500 Befragungshaushalte erweitert wird. Die methodische Funktion liegt darin, eine breitere Basis für spezifische neue Erhebungsinstrumente zu bieten, die an die Technik der computerunterstützten Befragung gebunden sind. Dabei handelt es sich um kognitive Tests, die – nach einer Erprobung im Jahr 2005 – in der SOEP-Haupterhebung erstmals im Jahr 2006 zum Einsatz kommen.² Sie gehören zu einem innovativen Schwerpunkt des SOEP, in dem die Methodik repräsentativer Umfragen mit Messtechniken unterschiedlicher Art verbunden wird.³

In der SOEP-Haupterhebung kommt ein flexibler Mix verschiedener Interviewmethoden zum Einsatz, um die Teilnahmebereitschaft und damit die Panelstabilität so hoch wie möglich zu halten. Interviews mit CAPI-Technik haben dabei einen Anteil von rd. 25%. Die Auswahl der Haushalte, die mit CAPI interviewt werden, ist nicht zufallsgesteuert, sondern wird situativ von Interviewern und Befragungspersonen bestimmt. Die neue Stichprobe H bietet hier nun eine

¹ Heutiger Name des Instituts: TNS Infratest Sozialforschung.

² Lang 2005

³ Vgl. Überblick im Methodenbericht zur Testerhebung 2005.
TNS Infratest Sozialforschung: SOEP 2005 – Methodenbericht zur Testerhebung. München 2005.

wichtige Ergänzung, weil die Erstbefragung im Jahr 2006 – dies ist ein Novum in der Geschichte des Projekts – vollständig und ausschließlich mit CAPI-Interviews durchgeführt wird.

Damit werden die Analysegrundlagen für Daten, die an die CAPI-Technik gebunden sind – in erster Linie also die kognitiven Tests, für methodische Analysen aber auch alle anderen Daten – in doppelter Hinsicht verbessert:

- Zum einen wird die verfügbare Fallzahl der Befragten erhöht (um rd. 1.500 auf der Haushaltsebene und rd. 2.500-3.000 auf der Personenebene)
- zum anderen handelt es sich hier um eine zufallsbasierte Teilstichprobe aller Haushalte bzw. Personen, die die Grundgesamtheit vollständig repräsentiert.

Die verwendeten Erhebungsinstrumente sind dieselben wie in den Panelstichproben A-G. Eine Modifikation erfolgt nur insofern, als bestimmte Basisinformationen, die in der Längsschnittbefragung aus früheren Befragungswellen vorhanden sind, in der Erstbefragung zusätzlich erhoben werden (Wohnsituation und Wohnumfeld im Haushaltsfragebogen, Schulabschluss und berufliche Ausbildung im Personenfragebogen). Der entstehende Datenbestand kann mit dem der übrigen SOEP-Stichproben zu einem Gesamtdatenbestand integriert werden.

Die Stichproben A – D des SOEP sind repräsentative Stichproben bestimmter Teilgruppen der Wohnbevölkerung Deutschlands:

Der hier vorgelegte Methodenbericht konzentriert sich auf die spezifischen Aspekte einer Erstbefragung im SOEP. Die methodischen Verfahren eines Haushaltspanels im Allgemeinen und des deutschen Sozio-ökonomischen Panels im Besonderen werden als bekannt vorausgesetzt, bzw. es wird auf die Literatur zum SOEP und die vorliegenden jährlichen Methodenberichte von TNS Infratest zum SOEP verwiesen.

Den Befragungspersonen gegenüber werden alle Stichproben des Sozio-ökonomischen Panels einheitlich unter dem Projekttitel LEBEN IN DEUTSCHLAND vorgestellt. Das gilt auch für die neue Stichprobe H. Da es sich um eine Erstbefragung handelt, unterscheiden sich die begleitenden Materialien wie Intervieweranleitung, Anschreiben und Broschüren, Unterlagen zur Adressenerhebung usw. von denen der Wiederholungsbefragungen der Stichproben A-G. Die wesentlichen Teile der begleitenden Materialien (Anschreiben, Broschüre) sind in ANHANG zu diesem Bericht wiedergegeben. Für die Erhebungsinstrumente wird im übrigen auf den *Anlagenband* zum Methodenbericht für das SOEP 2006 verwiesen.

2. Stichprobenanlage

Ziel ist eine repräsentative Stichprobe von 1.500 befragten Haushalten. In diesen Haushalten sollten alle Haushaltsmitglieder ab 16 Jahren mit individuellen Interviews an der Befragung teilnehmen. Wenn dies in gleichem Umfang gelingt wie in den SOEP-Stichproben A-G, ist mit durchschnittlich 1,8 teilnehmenden Personen pro Haushalt zu rechnen. Die angestrebte Personenstichprobe hat damit einen Umfang von rd. 2.700 befragten Personen.

Wir beschreiben in diesem Kapitel die statistische Anlage der Stichprobe, also die Verfahren zur Auswahl der Sample-Points und der Befragungshaushalte. Informationen zur realisierten Nettostichprobe enthält das folgende Kapitel zur Durchführung der Befragung.

2.1 Grundgesamtheit und Auswahlverfahren

1. *Grundgesamtheit*

Die Grundgesamtheit der Stichprobe, die über das Auswahlverfahren abgebildet werden soll, umfasst alle Privathaushalte in Deutschland.

Zugleich umfasst die Grundgesamtheit auf der Personenebene die Wohnbevölkerung Deutschlands, soweit diese in Privathaushalten lebt.

Der Anspruch der Repräsentativität auf der Haushalts- ebenso wie auf der Personenebene wird dadurch ermöglicht, dass im Zuge der Befragung alle in den Befragungshaushalten lebenden Personen erfasst werden. Als Befragungspersonen sind alle Haushaltsmitglieder ab 16 Jahren definiert (genauer: Geburtsjahrgänge 1989 und älter). Über die in den Haushalten lebenden Kinder werden Basisinformationen erhoben.

2. *Auswahlverfahren*

Die Auswahl der Befragungshaushalte erfolgt durch eine mehrfach geschichtete, mehrstufige Zufallsstichprobe.

Schichtung

Sämtliche Gemeinden des Untersuchungsgebietes werden nach regionalen Kriterien in Schichten eingeteilt.

Schichtungsmerkmale

Als Schichtungsmerkmale dienen

Bundesländer
Regierungsbezirke
BIK-Gemeindetypen⁴

Aufteilung auf die Schichten

Die Aufteilung der Auswahleinheiten der ersten Auswahlstufe (Sample-Points) erfolgt proportional zur Zahl der Haushalte in den Schichten (Zellen), und zwar so, dass auch die jeweiligen Randbesetzungen der Schichtungsmerkmale optimal die Haushaltsverteilungen der entsprechenden Randkategorien abbilden.

Auswahlstufen

Für die erste Auswahlstufe wird die BRD in rund 53.000 Flächen aufgeteilt anhand der kommunalen, statistischen Bezirke und unter Zuhilfenahme eines geografischen Informationssystems (GIS) zur Straßeneinteilung der BRD. Diese Flächen oder Sample-Points bilden die Auswahleinheiten der ersten Auswahlstufe.

Mit Wahrscheinlichkeiten proportional zur Zahl der Haushalte in den Sample-Points wird nun in jeder Zelle die erforderliche Anzahl Sample-Points gezogen. Ein zusätzlicher Schichtungseffekt ergibt sich dabei durch Sortierung der Auswahleinheiten nach Kreisen, Gemeinden und ggf. nach Stadtbezirken innerhalb jeder Zelle.

⁴

Gemeindetyp (nach BIK Regionen)

0 =	500.000 und mehr Einwohner	(Strukturtyp 1)
1 =	500.000 und mehr Einwohner	(Strukturtyp 2,3,4)
2 =	100.000 bis unter 500.000 Einwohner	(Strukturtyp 1)
3 =	100.000 bis unter 500.000 Einwohner	(Strukturtyp 2,3,4)
4 =	50.000 bis unter 100.000 Einwohner	(Strukturtyp 1)
5 =	50.000 bis unter 100.000 Einwohner	(Strukturtyp 2,3,4)
6 =	20.000 bis unter 50.000 Einwohner	
7 =	5.000 bis unter 20.000 Einwohner	
8 =	2.000 bis unter 5.000 Einwohner	
9 =	unter 2.000 Einwohner	

Strukturtyp 1	Kernbereich der BIK Region
Strukturtyp 2	Verdichtungsbereich der BIK Region
Strukturtyp 3	Übergangsbereich der BIK Region
Strukturtyp 4	Peripherer Bereich

Die Größenzuordnung bezieht sich auf die Einwohnerzahl der betreffenden BIK Region. Bei Gemeinden außerhalb dieser Regionen erfolgt die Zuordnung entsprechend der politischen Gemeindegrößenklasse.

In der zweiten Auswahlstufe werden die für die Stichprobe pro Sample-Point benötigten Haushaltsadressen bestimmt. Dem Interviewer wird eine durch Zufallsauswahl festgelegte Adresse als Startpunkt für einen *Random Walk* übermittelt. Ausgehend von dieser Startadresse folgt der Interviewer nach vorgegebenen Regeln einem Behebungsweg, auf dem jeder dritte Haushalt als Befragungshaushalt definiert ist. Das verwendete Verfahren hierfür wird im folgenden Kap. 2.2 näher beschrieben.

Anzahl und Verteilung der Sample-Points

Die Stichprobe umfasst 250 Sample-Points, die nach dem beschriebenen Verfahren (erste Auswahlstufe) ausgewählt wurden.

Zur Veranschaulichung zeigen wir die Verteilung der Sample-Points nach Bundesländern.

	Sample- points
Schleswig-Holstein	9
Hamburg	5
Niedersachsen	22
Bremen	2
Nordrhein-Westfalen	54
Hessen	20
Rheinland-Pfalz	12
Saarland	4
Baden-Württemberg	29
Bayern	35
Berlin ehemal. Westteil	8
Berlin ehemal. Ostteil	5
Brandenburg	9
Mecklenburg-Vorpommern	5
Sachsen	14
Sachsen-Anhalt	9
Thüringen	8
Insgesamt	250

Innerhalb der Bundesländer sind die Sample-Points entsprechend der Schichtung nach Gemeindegrößen (BIK-Typen) verteilt. Zur Veranschaulichung zeigen wir eine zusammenfassende Verteilung nach West/Ost und drei Gemeindegrößenklassen: den Ballungsräumen mit über 500.000 Einwohnern (BIK 0 und 1), den mittleren verstädterten Regionen mit 100.000 bis unter 500.000 Einwohnern (BIK 2 und 3) und den übrigen, mittel- und kleinstädtisch oder ländlich geprägten Räumen (BIK 4-9).

Verteilung der Sample-Points

	West ⁵	Ost ⁶	Gesamt
BIK 0-1	71	15	86
BIK 2-3	69	13	82
BIK 4-9	60	22	82
Gesamt	200	50	250

Das ehemalige Bundesgebiet West ist demnach mit 200 Sample-Points repräsentiert, das Gebiet der ehemaligen DDR mit 50 Sample-Points. Die drei Regionstypen haben insgesamt einen etwa gleich großen Anteil, wobei im Westen die großstädtischen Ballungsräume und im Osten die kleinstädtisch-ländlichen Gebiete stärker vertreten sind.

2.2 Auswahl der Haushalte über getrennte Adressenerhebung

Das beschriebene Auswahlverfahren folgt dem Modell des „Random-Route“, das in Deutschland für mündlich-persönliche Befragungen von Zufallsstichproben der Bevölkerung das etablierte Verfahren ist. Innerhalb dieses Modells gibt es allerdings eine Bandbreite von Verfahrensweisen im Detail, die – je nach Positionierung am oberen oder unteren Rand – mit erheblichen Qualitäts- und Kostenwirkungen verbunden ist.

Für den Aufbau der neuen Panelstichprobe im Rahmen des SOEP-Stichprobensystems wurde eine Ausgestaltung des Random-Route-Modells gewählt, die deutlich höhere Qualitätsstandards ermöglicht als das Standardmodell. Kernpunkt ist der „getrennte Adressenvorlauf“. Während die Auswahl der Befragungshaushalte normalerweise in einem Arbeitsgang mit der Durchführung der Interviews erfolgt, wurden für die SOEP-Stichprobe H – ebenso wie im Jahr 2000 schon für die Stichprobe F – beide Arbeitsvorgänge getrennt. Dies hat folgende methodische Vorteile:

- Für die eigentliche Befragung werden dem Interviewer feste Adressen vorgegeben. Die Abläufe und Kontrollmöglichkeiten entsprechen praktisch denjenigen einer Adressen-Random-Stichprobe, die beispielsweise mit Adressen aus Einwohnermelderegistern arbeitet.⁷

⁵ Alte Bundesländer mit Westteil Berlins

⁶ Neue Bundesländer mit Ostteil Berlins

⁷ Für ein Haushaltspanel ist eine Einwohnermeldeamtsstichprobe nicht geeignet, weil hier eine Haushaltsstichprobe benötigt wird, während aus Einwohnermelderegistern Personen-Adressen gezogen werden.

- Die Qualität der Adressen – Lage auf dem Begehungsweg, richtige Namen, korrekte Schreibweise – kann teilweise vorab geprüft werden und dann in der Befragungssituation noch einmal gegengecheckt werden. Gerade für eine Panelbefragung – im Hinblick auf die erneute Befragung im Folgejahr, Panelpflege, Adressenrecherche bei Umzügen – ist die Korrektheit der verwendeten Adressen wichtig.
- Die Adressenerhebung kann mit der Erhebung zusätzlicher Informationen verknüpft werden, um verbesserte Möglichkeiten einer Nonresponse-Analyse, also eines Strukturvergleichs von Teilnehmern und Nicht-Teilnehmern an der Befragung zu schaffen. Bei Stichprobe F im Jahr 2000 wurde eine „interviewergestützte Quartiersbeschreibung“ nach Hoffmeyer-Zlotnik vorgenommen. Darauf konnte jetzt verzichtet werden, weil die mittlerweile im SOEP eingeführte Zuspiegelung von Geodaten (MOSAIC-Typen der Fa. Microm) ähnliche Informationen in erweiterter Form bietet. Diese Informationen werden nach Abschluss der Erhebung allen Adressen der Bruttostichprobe zugespielt.
- Die Vorab-Fixierung der eingesetzten Adressen ermöglicht eine klare Beschreibung der Bruttostichprobe und der Bearbeitung jeder einzelnen Adresse im Feldverlauf. TNS Infratest liefert diese Informationen in einem eigenen Datenbestand, dem so genannten „Bruttoband“, ergänzend zu den eigentlichen Befragungsdaten, dem so genannten „Nettoband“.

Die Adressenerhebung für die SOEP-Stichprobe H erfolgte in der Zeit vom 12.-25. Januar 2006. Für alle 250 Sample-Points der Stichprobe wurde ein in der jeweiligen Region lebender Interviewer mit der Adressenerhebung nach fest vorgegebenen Regeln beauftragt. Die eingesetzten Interviewer waren zum Teil SOEP-Interviewer, die bei der späteren Befragung eingesetzt wurden, zum Teil andere, bei der späteren Befragung nicht beteiligte Interviewer.

Die Interviewer hatten 24 Adressen pro Sample-Point aufzulisten. Dies erfolgt auf einem *random walk* mit fest vorgegebener Startadresse. Aufgelistet wird jeder dritte Haushalt auf diesem Begehungsweg. Dabei wird mit den Haushalten kein Kontakt aufgenommen. Notiert werden also „Klingelschild-Adressen“, die nicht eindeutig zu einem Anstaltshaushalt oder einer Firma gehören. Ob sich dahinter wirklich ein Privathaushalt verbirgt (oder ggf. mehrere Privathaushalte, oder eine leer stehende Wohnung), kann endgültig erst in der Befragungsphase geklärt werden.

Die Adressenauflistung erfolgt zunächst mittels eines Erhebungsbogens mit 24 Adress-Zeilen. Jede Adresse ist vom Interviewer dann in eine Bildschirmmaske zu übertragen und wird in elektronischer Form an TNS Infratest übermittelt. Durch dieses Verfahren wird nicht nur Zeit gespart, sondern es werden auch Fehler bei der Datenerfassung (Lesefehler bei Namen und Adressen) vermindert.

Von den 24 übermittelten Adressen sind die ersten 15 als Teil der Bruttostichprobe definiert, die übrigen sind Reserveadressen für den Fall qualitätsneutraler Ausfälle (vgl. dazu näher

Kap. 3.3). Der Gesamtbestand an Adressen, die auf diese Weise erhoben und als Grundlage für die Bildung der Bruttostichprobe übermittelt wurden, betrug 6.000 Adressen.

Der Adressenbestand wurde im Institut zusammengeführt und anhand von Straßen- und Telefonverzeichnissen auf Plausibilität und Vollständigkeit überprüft. Auf dieser Grundlage wurden für jeden Samplepoint Adressenlisten der Haushalte erstellt, die in der Feldphase dann zu bearbeiten waren. Die Bruttostichprobe umfasst 3.750 Haushaltsadressen.

2.3 Die realisierte Nettostichprobe

In den folgenden Kapiteln dieses Berichts werden Durchführung und Erfolg der Erhebung unter verschiedenen methodischen Gesichtspunkten näher beschrieben. Vorab soll kurz das Ergebnis dargestellt werden, also die letztlich realisierte Nettostichprobe in der ersten Befragungswelle der neuen Ergänzungsstichprobe H.

Die realisierte Stichprobe umfasst **1.506 Haushalte**. Damit wurde das angestrebte Ziel erreicht, innerhalb des Stichprobenrahmens eine Zahl von mindestens 1.500 Haushalten für die Teilnahme an der Panelbefragung zu gewinnen.

In den 1.506 Teilnehmerhaushalten leben 2.784 Personen im Befragungsalter ab 16 Jahren. Jede dieser Personen sollte mit einem eigenen *Personenfragebogen* individuell befragt werden. Dies gelang mit einer Erfolgsquote von 94,0% – einem Wert, der fast auf demselben hohen Niveau liegt wie in den schon länger teilnehmenden Haushalten der Panelstichproben A-G. Auf Personenebene umfasst damit der Datenbestand der Erstbefragung **2.616 Interviews**.

3. Durchführung der Erhebung

3.1 Interviewereinsatz und Interviewmethode

Die Befragungen im SOEP erfolgen durch Infratest-Interviewer, die in der Befragungsregion des jeweiligen Panelhaushalts leben. Bei den Längsschnittbefragungen der Altstichproben A – G spielt ergänzend mittlerweile die zentrale Bearbeitung von Panelhaushalten eine erhebliche Rolle, die über telefonische Kontaktinterviews und schriftlich-postalische Befragung erfolgt (vgl. Methodenbericht 2005). Bei einer Erstbefragung neuer Panelhaushalte ist jedoch der persönliche Interviewereinsatz unverzichtbar.

Im Jahr 2005 wurden für die SOEP-Altstichproben 566 Interviewer eingesetzt, die mit wenigen Ausnahmen über mehrjährige SOEP-Erfahrung verfügten, und ähnlich ist es im SOEP 2006. Die Bearbeitung der neuen SOEP-Stichprobe H mit ihren 250 Sample-Points kam als zusätzliche Aufgabe hinzu und erforderte daher zusätzliche Interviewer:

Interviewereinsatz in Stichprobe H

Insgesamt eingesetzte Interviewer	234
darunter:	
– SOEP-erfahrene Interviewer, die auch in den Altstichproben eingesetzt waren	143
– Interviewer ohne SOEP-Erfahrung	91

Die Erstbefragung der neuen Stichprobe H konnte also nicht von den erfahrenen SOEP-Interviewern einfach „mit erledigt“ werden. Erforderlich war vielmehr die Einarbeitung einer erheblichen Zahl neuer Interviewer. Diese waren zwar alle in Face-to-Face-Bevölkerungsumfragen geschult und erfahren, aber nicht mit den spezifischen Anforderungen des SOEP vertraut.

Die projektspezifische Einarbeitung erfolgte anhand des umfangreichen „Interviewerhandbuchs LEBEN IN DEUTSCHLAND“ und einer Hotline zu Projektmitarbeitern bei TNS Infratest in der Startphase der Feldarbeit.

Eine Besonderheit der Erstbefragung in Stichprobe H ist es, dass die CAPI-Technik (CAPI = Computer Assisted Personal Interview) erstmals nicht im Rahmen eines flexiblen *mixed-mode*-Ansatzes, sondern obligatorisch bei allen Interviews eingesetzt werden sollte. Die Erhebungsinstrumente wurden überhaupt nicht als Papierfragebogen ausgeliefert, sondern nur als CAPI-Programm. Dies beinhaltet den Zwang, in einem Mehr-Personen-Haushalt das Interview mit jedem einzelnen Haushaltsmitglied wirklich als persönliches Interview durchzuführen (während im *mixed-mode*-Ansatz für einzelne Haushaltsmitglieder ein Fragebogen hinterlassen werden kann).

Die Beschränkung auf die Methode der CAPI-Interviews hat zudem Auswirkungen auf die Auswahl der Interviewer, die in dem Projekt eingesetzt werden. Sie bedeutet - wie oben gezeigt -, dass viele erfahrene ältere SOEP-Interviewer (die nicht mit Laptop arbeiten), nicht eingesetzt werden konnten. Vielmehr musste ausschließlich auf den Bestand an CAPI-Interviewern im TNS Infratest-Stab zurückgegriffen werden. Die Effekte werden im folgenden Abschnitt zum zeitlichen Verlauf der Feldarbeit erläutert.

3.2 Zeitlicher Verlauf der Feldarbeit

Die Feldarbeit im SOEP 2006 begann für die „alten“ Panelstichproben A-G am 1. Februar 2006.

Die Bearbeitung der neuen Ergänzungsstichprobe H erfolgte parallel dazu, jedoch leicht zeitversetzt ab dem 10. März. Die folgende Übersicht zeigt den gesamten Ablauf:

Terminübersicht für Erstbefragung Stichprobe H

Adressenvorlauf:	12.01.2006 – 25.01.2006
Adressenprüfung und Einsatzvorbereitung:	13.01.2006 – 06.02.2006
Versand der Anschreiben an die Haushalte:	21.02.2006 – 24.02.2006
Versand der Unterlagen an die Interviewer:	07.03.2006
Modemübertragung (CAPI)	09.03.2006
Feldbeginn:	10.03.2006
Telefonische Nachbearbeitung (1):	ab 17.05.2006
Nachbearbeitung Feld:	ab 17.05.2006
Telefonische Nachbearbeitung (2):	04.07.2006 – 25.07.2006
Feldende:	25.07.2006
Datenaufbereitung und -prüfung:	13.04.2006 – 14.08.2006
Geprüfter Datenbestand:	15.08.2006

Die Feldarbeit erstreckte sich damit über 4,5 Monate. Dieser lange Zeitraum erklärt sich teilweise aus den erforderlichen Anstrengungen zur Kontaktierung und Gewinnung der Haushalte. Zusätzlich spielte jedoch die methodische Vorgabe eine Rolle, dass – anders als in den „alten“ Panelstichproben A-G – alle Interviews mit der CAPI-Technik durchzuführen waren. Dies ist gegenüber dem *mixed-mode*-Ansatz eine erhebliche Erschwerung.

Viele von den CAPI-Interviewern haben – wie oben bereits gezeigt - bisher nicht im SOEP mitgearbeitet. Sie gehören aber zum aktiven Kern des Interviewer-Stabes und werden in vielen Projekten eingesetzt. Sie haben die Arbeit an der SOEP-Befragung daher nicht immer sofort aufgenommen, sondern mit zeitlicher Verzögerung, weil Aufgaben in anderen Projekten

teilweise vorrangig bearbeitet wurden. Interviewer, die bisher nicht im SOEP gearbeitet haben, schrecken häufig vor den sehr hohen Anforderungen in diesem Projekt zurück. Sie ziehen es dann vor – schließlich sind sie Freiberufler – ihr Geld mit der Durchführung einfacherer, „normaler“ Befragungen zu verdienen. Manche der ursprünglich eingesetzten Interviewer haben die Aufgabe im SOEP überhaupt nicht angepackt und mussten nach Ablauf der ihnen gesetzten Fristen durch einen anderen Interviewer ersetzt werden.

Dies erklärt, dass die Bearbeitung der Stichproben-Haushalte sich relativ gleichmäßig über den gesamten Zeitraum der Feldarbeit erstreckte.

Monat des letzten Haushaltskontaktes bei Durchführung der Befragung

Anzahl Haushalte nach Monaten	Bruttostichprobe	Realisierte Interviews
März	1.005	450
April	718	263
Mai	826	337
Juni	885	350
Juli	316	106
Gesamt	3.750	1.506

Die von den Interviewern geforderten Anstrengungen lassen sich anhand der dokumentierten Haushaltskontakte verdeutlichen. Um die geforderte Zahl von Haushalts- und Personeninterviews zu erreichen, waren 11.133 Kontakte der Interviewer zu den Stichprobenhaushalten erforderlich.

- 4,583 Kontakte waren erforderlich für die 1.506 Haushalte, die schließlich erfolgreich bearbeitet werden konnten. Das sind im Durchschnitt 3,0 Interviewerkontakte pro Haushalt, in vielen Fällen aber auch erheblich mehr.
- 7097 Kontakte sind dokumentiert bei denjenigen Haushalten, die letztlich *nicht* erfolgreich bearbeitet werden konnten – sei es, weil niemand angetroffen wurde oder die Haushaltsmitglieder sich nicht zur Teilnahme gewinnen ließen oder weil die Adresse sich als Fehladresse erwies (z.B. unbewohnt, Geschäftsadresse usw.). Im Durchschnitt sind das 2,9 Kontakte bzw. Kontaktversuche bis zur (negativen) Klärung des Falles. Es kamen aber auch 9 und mehr Kontakte bzw. Kontaktversuche bis zur Feststellung eines Misserfolgs vor – dies ist für immerhin 110 Haushalte dokumentiert.

3.3 Stichprobenausschöpfung auf der Haushalts- und Personenebene

Die Bruttostichprobe besteht auf der ersten Auswahlstufe, wie oben beschrieben (Kapitel 2.1), aus 250 Sample-Points. Auf der zweiten Auswahlstufe erfolgt über den *random walk* die Auswahl der Befragungshaushalte (vgl. Kapitel 2.2). Den Interviewern wurden pro Sample-Point 15 Adressen als Befragungsadressen übersandt. Von diesen sollten so viele Haushalte wie möglich zur Teilnahme an der Befragung zum LEBEN IN DEUTSCHLAND gewonnen werden. Ersatzadressen für qualitätsneutrale Ausfälle (QNA) wurden den Interviewern auf Bedarf zusätzlich gesandt.

Alle Befragungshaushalte wurden vorab angeschrieben und um ihre Teilnahme gebeten. Dabei wurden sie über das Forschungsvorhaben und die Anforderungen informiert, wobei auch auf die geplanten Wiederholungsbefragungen hingewiesen wurde.

Das Stichprobenbrutto besteht damit aus 3.750 Befragungshaushalten. Qualitätsneutrale Ausfälle werden ersetzt und verändern das effektive Brutto insofern nicht. Sie werden in der Stichprobendokumentation („Bruttoband“) als zusätzliche Adressen aufgenommen.

Qualitätsneutrale Ausfälle (QNA)

Die im Adressenvorlauf ermittelten Adressen sind „Klingelknopf-Adressen“, sind also nach Augenschein – nicht nach vorherigem Kontaktinterview – in die Adressenauflistung aufgenommen worden. Insofern ist zu erwarten, dass die Interviewer in der folgenden Feldarbeit – etwa 2 bis 3 Monate später – in Einzelfällen feststellen, dass es sich um unbewohnte Wohnungen oder Geschäftsadressen handelt. Bewohner zum Zeitpunkt der Adressenermittlung können in der Zwischenzeit auch verstorben oder verzogen sein.⁸

Insgesamt haben die Interviewer 181 Adressen in diesem Sinne als Fehladressen gemeldet, das ist ein QNA-Anteil von 4,8%.

Die im Bruttoband dokumentierte Gesamtstichprobe umfasst damit 3.931 Haushaltsadressen. Von diesen bilden 3.750 das bereinigte Brutto.

Ausfälle und Interviews auf Haushaltsebene

Von den 3.750 Haushalten der Bruttostichprobe konnten 1.506 zur Teilnahme an der Befragung gewonnen werden. Das ist eine Stichprobenausschöpfung von 40,2%.

⁸ Sofern bereits ein Nachfolgehaushalt in die Wohnung eingezogen war, wurde dieser zum Befragungshaushalt.

Im Einzelnen verteilen sich Ausfälle und Interviews wie folgt:

	abs.	%
Bruttofallzahl	3.750	100,0
Haushalt nicht angetroffen	485	12,9
Teilnahme nicht möglich (gesundheitlich oder geistig nicht in der Lage; Sprachprobleme)	172	4,6
Teilnahme verweigert	1.487	39,7
Vom Interviewer nicht korrekt durchgeführt	3	0,0
Ohne Ergebnis zurück/nicht bearbeitet	<u>97</u>	<u>2,6</u>
Summe der Ausfälle	2.244	59,8
Interviews	1.506	40,2

44% der angesprochenen Haushalte haben die Teilnahme an der Befragung verweigert oder waren dazu aus gesundheitlichen, geistigen oder sprachlichen Gründen nicht in der Lage. 12,9% der Haushalte wurden vom Interviewer nie angetroffen – dies trotz teilweise erheblicher Kontaktbemühungen, wie oben dargestellt (Kapitel 3.2).

2.6% der Adressen wurden vom eingesetzten Interviewer nicht bearbeitet oder ohne dokumentierte Bearbeitung zurückgesandt. Dabei handelt es sich überwiegend um Einzeladressen, in einem Fall auch um einen ganzen Sample-Point (Totalausfall).

Eine telefonische Zielpersonenkontrolle in 336 Haushalten, die im Oktober 2006 von der Infratest Feldabteilung durchgeführt wurde, ergab Hinweise darauf, dass ein Interviewer die Befragung mindestens einem Haushalt nicht oder zumindest nicht korrekt durchgeführt hatte. Daraufhin wurden alle drei Haushalte, die von diesem Interviewer bearbeitet wurden und in den neun Personen (davon acht im Befragungsalter) lebten, im Bruttoband und im Nettodatensatz entsprechend gekennzeichnet.

Entwicklung der Stichprobenausschöpfung

Die Stichprobenausschöpfung ist im Vergleich zu früheren Erstbefragungen im Sozio-ökonomischen Panel erneut zurückgegangen. Vergleichbar sind folgende Werte:

	Stichproben- ausschöpfung
SOEP-Stichprobe A im Jahr 1984	60,6%
SOEP-Stichprobe E im Jahr 1998	54,2%
SOEP-Stichprobe F im Jahr 2000	51,0%
SOEP-Stichprobe H im Jahr 2006	40,2%

Dieser Trend ist ähnlich aus anderen Umfrageprojekten, etwa dem ALLBUS, bekannt. Er spiegelt veränderte Einstellungen der Bevölkerung zur Teilnahme an solchen Umfragen wider, aber auch veränderte Arbeitsweisen und Einstellungen der Interviewer in den vorhandenen Stäben für mündlich-persönliche Befragungen. Diese veränderten Bedingungen werden durch große Anstrengungen im Rahmen einzelner Projekte wie dem SOEP beantwortet, können in ihren Auswirkungen letztlich aber nicht aufgehoben werden.

Die rückläufige Stichprobenausschöpfung muss nicht unbedingt eine Verschlechterung der Stichprobenqualität – im Sinne einer abbildungstreuen Wiedergabe der Grundgesamtheit – zur Folge haben. Maßgeblich hierfür ist die Frage, ob die Ausfälle zufällig verteilt sind oder eine Selektivität in Hinblick auf relevante Stichprobenmerkmale aufweisen. Dies wird durch Strukturvergleiche im Zuge der Stichprobengewichtung und, sofern möglich, durch zusätzliche Nonresponse-Analysen überprüft (vgl. dazu Kap. 4).

Stichprobenausschöpfung auf der Personenebene

Zur Methodik der Haushalts-Panels gehört die individuelle Befragung aller Haushaltsmitglieder ab 16 Jahren. Instrument dafür ist der PERSONEN-Fragebogen.

Die Bruttostichprobe auf Personenebene umfasst alle Personen ab 16 Jahren, die in den teilnehmenden Haushalten leben. Die Nettostichprobe umfasst alle Personen, von denen ein auswertbarer PERSONEN-Fragebogen vorliegt.

Die entsprechenden Zahlen in der neuen SOEP-Stichprobe H sind folgende:

	absolut	%
Teilnehmerhaushalte in Welle 1 In diesen Haushalten leben...	1.506	
Personen insgesamt	3.407	
darunter:		
Kinder unter 16 Jahren	623	
Befragungspersonen (= Brutto)	2.784	100
darunter:		
- nie angetroffen	22	0,8
- zur Teilnahme nicht in der Lage, alt/krank	22	0,8
- zur Teilnahme nicht bereit oder Interview nicht durchführbar	<u>124</u>	<u>4,4</u>
Summe Ausfälle	168	6,0
Interviews	2.616	94,0

Die angestrebte individuelle Information ist für 94% der Haushaltsmitglieder in den befragten Haushalten vorhanden. Der erreichte Vollständigkeitsgrad der gewünschten Haushaltsinformation liegt damit auf gleichem Niveau wie in den wiederholt befragten Haushalten der Panelstichproben A-G. Dies ist ein besonders erfreuliches Ergebnis, das nicht unbedingt zu erwarten war – stellte doch die ausschließliche Durchführung der Befragung über mündlich-persönliche CAPI-Interviews eine besondere Anforderung dar. Der Unterschied zwischen deutlich rückläufiger Ausschöpfung auf Haushaltsebene und konstant hoher Ausschöpfung auf Personenebene in Befragungshaushalten kann als ein Hinweis darauf gesehen werden, dass es projektübergreifende Faktoren sind, welche die rückläufige Teilnahmebereitschaft auf Haushaltsebene erklären: hat ein SOEP-Interviewer den Kontakt zu einem Haushalt erfolgreich hergestellt und ein Mitglied im Haushalt zur Teilnahme gewinnen können, sind in der großen Mehrheit der Haushalte auch alle anderen Haushaltsmitglieder zur Teilnahme bereit.

Um eine möglichst hohe Quote von Haushalten mit vollständiger Teilnahme aller Haushaltsmitglieder zu erreichen, wurde eine zentrale Nachbearbeitung von (zunächst) unvollständig realisierten Haushalten durchgeführt. Modell hierfür war die zentrale Bearbeitung in den Panelstichproben A-G, mit der Haushalte bearbeitet werden, die einen Interviewerbesuch nicht wünschen oder vom Interviewer nicht zur Teilnahme gewonnen werden können. Dabei werden zunächst telefonische Kontaktinterviews zur Klärung der Teilnahmebereitschaft geführt. Bei grundsätzlicher Bereitschaft wird der Fragebogen postalisch an die Haushalte übersandt und von diesen selbst ausgefüllt und zurückgesandt.

In der vorliegenden Erstbefragung war dies keine generelle Möglichkeit, weil die Interviews mit der CAPI-Technik durchgeführt werden sollten. Für den Versuch zur Komplettierung unvollständig realisierter Haushalte war sie jedoch die einzige Möglichkeit.⁹

Insgesamt wurden 70 Befragungspersonen aus teilweise realisierten Haushalten in einer telefonischen Nachbearbeitung erneut kontaktiert, davon konnten 38 Personen auf diese Weise zur Mitarbeit gewonnen werden.

4. Stichprobenstrukturen und Non-Response Informationen

Nonresponse-Informationen sind Informationen, die für Teilnehmer und Nichtteilnehmer („Nonrespondenten“) vorliegen oder gegebenenfalls auch nur für Nichtteilnehmer bzw. Ausfälle erhoben werden. Der Strukturvergleich für Teilnehmer und Nichtteilnehmer kann Hinweise auf eine eventuelle Selektivität des Teilnahmeverhaltens und damit auf mögliche Verzerrungen in der Struktur der Nettostichprobe erbringen. Im Sozio-ökonomischen Panel wurden Nonresponse-Informationen bei Erstbefragungen bereits mehrfach genutzt, um Hinweise zur Validität der Stichprobe zu erhalten. Leider gibt es in der Regel – wenn überhaupt – nur wenige Merkmale, die für Teilnehmer und Nichtteilnehmer einer Befragung gleichermaßen vorliegen und im Rahmen der Feldarbeit erhoben werden.

Aus der feldbezogenen Arbeit stehen folgende Merkmale in der Bruttostichprobe zur Verfügung

Regionale Verteilung

1. Bundesland
2. Gemeindegrößenklassen politisch
3. Gemeindegröße nach BIK (Stadtregionen)
4. Regierungsbezirk
5. Siedlungsstruktureller Gemeindetyp nach BFLR

Interviewergestützte Wohnumfeldinformationen

6. Gebäudetyp (Art des Hauses: Ein- vs. Mehrfamilienhäuser etc.)
7. Unterkunftstyp (Wohnhaus vs. verschiedene Wohnheimtypen)
8. Angaben zur Umgebung (Wohn-/Misch-/Gewerbegebiet etc)

⁹ Den Befragungspersonen wurde die Papierfassung des PERSONEN-Fragebogens zugesandt, wie er für die Stichproben A-G vorlag, mit einem Ergänzungsblatt für die zusätzlichen Fragen zu Bildungsabschlüssen. Die an das CAPI-Interview gebundenen kognitiven Tests konnten für diese Personen nicht durchgeführt werden.

Haushaltsmerkmale (unvollständig)

9. Haushaltsgröße
10. Geschlecht des HV
11. Altersgruppe des HV
12. Nationalität des HV, ggf. Sprachkenntnisse
13. Berufstätigkeit/berufliche Stellung des HV

Nur die Regions- und Wohnumfeldinformationen liegen vollständig für die gesamte Bruttostichprobe vor. Die Haushaltsmerkmale stützen sich auf Angaben der Interviewer zu Haushaltsausfällen. Da bei Ausfällen häufig jedoch keinerlei Information zu ermitteln ist, liegen die Angaben unvollständig vor, der Aussagewert für Nonresponse-Analysen ist dementsprechend eingeschränkt. Wir untersuchen im Folgenden Unterschiede in der Struktur der Teilnehmer und Nichtteilnehmer nach regionalen Merkmalen. Für verschiedene Regionen bzw. Regionstypen kann die jeweilige Response-rate ausgewiesen werden.

Ost-West

Wie bei früheren Aufstockungstichproben ist auch bei der Stichprobe H die Ausschöpfung in den östlichen Bundesländern mit 42,1% höher als im Westen, wo sie bei 39,7% liegt. Bei der letzten großen SOEP-Aufstockungstichprobe F lag die Differenz zwischen Ost und West noch bei knapp 6 Prozentpunkten. Der Abstand hat sich mehr als halbiert, weil im Osten der Rückgang der Ausschöpfung stärker ausgeprägt war als in den alten Bundesländern. Bei gleicher Ausschöpfung müsste in der Stichprobe die Zahl der Haushalte im Westen um 15 höher liegen, eine durch die Gewichtung leicht zu korrigierende Verzerrung der Stichprobe.

Politische Gemeindeklassengröße

Bezüglich der Gemeindegröße fällt kein generelles lineares Stadt-Land-Gefälle auf, sondern vor allem eine abfallende Ausschöpfung in Großstädten mit mehr als 500.000 Einwohnern. In diesen konnte nur in 28,8% der Haushalte ein Interview realisiert werden. Läge die *response rate* in den Großstädten auf dem Durchschnittsniveau aller Gemeindegrößenklassen, müsste die Stichprobe dort 70 Panelhaushalte mehr enthalten.

Die Ausschöpfung entlang der BIK-Typen zeigt ebenfalls die Sonderstellung der Großstädte mit mehr als 500.000 Einwohnern, mit Ausschöpfungen von 34,6 und 37,8 in den BIK-Typen 1 und 2.

Stichprobenausschöpfung nach Regionen/Gemeindegröße

	Inter-views	Aus-fälle	QNA	gesamt	Ber. Brutto	Aus-schöp-fung %	Diff. Realisiert-Erwartet	Anteil QNA %
Ost/West								
West	1190	1810	152	3152	3000	39,7	-15	4,8
Ost	316	434	29	779	750	42,1	+15	3,7
Gesamt	1506	2244	181	3931	3750	40,2	0	4,6
Polit Gemeindegrößenklasse (in 1000 Einw.)								
7 (500 u.m.)	177	438	31	646	615	28,8	-70	4,8
6 (100- <500)	266	394	40	700	660	40,3	1	5,7
5 (50- <100)	182	238	25	445	420	43,3	13	5,6
4 (20- <50)	314	391	26	731	705	44,5	31	3,6
3 (5-<20)	375	525	40	940	900	41,7	14	4,3
2 (2-<5)	134	181	14	329	315	42,5	7	4,3
1 (<2)	58	77	5	140	135	43,0	4	3,6
Gesamt	1506	2244	181	3931	3750	40,2	0	4,6
BIK								
0 (>500 Kern)	337	638	50	1025	975	34,6	-55	4,9
1 (>500 Rand)	119	196	9	324	315	37,8	-7	2,8
2 (100-<500 K)	266	394	42	702	660	40,3	1	6,0
3 (100-<500 R)	254	316	19	589	570	44,6	25	3,2
4 (50-<100 K)	46	59	5	110	105	43,8	4	4,5
5 (50-<100 R)	105	150	13	268	255	41,2	3	4,9
6 (20-<50)	176	244	26	446	420	41,9	7	5,8
7 (5-<20)	122	163	11	296	285	42,8	7	3,7
8 (2-<5)	49	41	3	93	90	54,4	13	3,2
9 (<2)	32	43	3	78	75	42,7	2	3,8
	1506	2244	181	3931	3750	40,2	0	4,6

Struktur der Nettostichprobe auf Haushaltsebene: Vor und nach Gewichtung und im Vergleich zur Bevölkerung (lt. Amtlicher Statistik)

	Inter views	% Anteile	% nach Gewich- tung	Bevölkerung	Abweichung in %- Punkten
Haushaltsebene					
Ost/West					
West	1190	79,0	81,5	81,5	0
Ost	316	21,0	18,5	18,5	0
Gesamt	1506	100	100	100	0
Polit GGK (in 1000 Einw.)					
7 (500 u.m.)	177	11,8	15,9	16,5	-0,6
6 (100- <500)	266	17,7	17,6	17,5	0,1
5 (50- <100)	182	12,1	12,2	9,1	3,1
4 (20- <50)	314	20,8	18,5	18,0	0,5
3 (5-<20)	375	24,9	23,7	24,1	-0,4
2 (2-<5)	134	8,9	9,0	8,6	0,4
1 (<2)	58	3,9	3,1	6,1	-3,0
BIK					
0	337	22,4	26,8	26,7	0,1
1	119	7,9	8,0	8,1	-0,1
2	266	17,7	17,2	17,1	0,1
3	254	16,9	14,6	14,7	-0,1
4	46	3,1	2,7	2,5	0,2
5	105	7,0	7,6	7,7	-0,1
6	176	11,7	10,8	11,0	-0,2
7	122	8,1	7,6	7,6	0
8	49	3,3	2,6	2,7	-0,1
9	32	2,1	1,9	2,0	-0,1

Durch die Stichprobenanlage und die Ausfallstrukturen ergeben sich Abweichungen der Stichprobe von bekannten Soll-Strukturen der Grundgesamtheit. Wie für die SOEP-Stichproben A–F wurde auch für die Aufstockungstichprobe H eine Designgewichtung vorgenommen, wobei hierbei keine Hochrechnungsfaktoren an die Datensätze angespielt werden, sondern Gewichtungsfaktoren. Die primäre Erhebungsebene im SOEP ist die Haushaltsebene. Für die Personengewichtungsfaktoren bedeutet dies, dass die Gewichtungsfaktoren der Haushaltsebene angespielt werden und anschließend eine Gewichtung auf Personenebene nach den unten genannten Merkmalen vorgenommen wird.

Haushaltsgewichtung

Die Haushaltsgewichtung erfolgt mit den Rändern:

- Bundesland nach BIK (neue Systematik)
- West/Ost nach Haushaltsgröße
- West/Ost nach Nationalität des Haushaltsvorstandes (deutsch/nichtdeutsch)

Daran anschließend wird der Haushaltsfaktor an den Personendatensatz angespielt

Der Personensatzes wird gewichtet mit den Rändern

West: Altersgruppen (4) x Nielsen x Geschlecht – deutsch (Alter 4 Gruppen)
 Altersgruppen (4) x Geschlecht – (Bula zusammengefasst) – nichtdeutsch

Ost: Altersgruppen (4) x Geschlecht x Bundesland (deutsch/nichtdeutsch)

West: Altersgruppen x Geschlecht x Nielsen – deutsch (Alter in Dekaden)
 Altersgruppen x Geschlecht – nichtdeutsch (Alter in Dekaden)

Ost: Altersgruppen x Geschlecht x Bundesland - (Alter in Dekaden)

West: Familienstand – verheiratet/ledig/verwitwet/geschieden

Ost: Familienstand – Geschlecht – verheiratet/ nicht verheiratet

West: deutsch/nichtdeutsch x Altersgruppen (5er Gruppen)

Ost : Geschlecht x Altersgruppen (5er-Gruppen)

West: deutsch/nichtdeutsch x Bundesland x Geschlecht

Ost : Bundesland x Geschlecht

Um Ausreißer in den Faktoren zu vermeiden, werden die Faktoren bei der Personengewichtung "gekapt". Für den iterativen Gewichtungslauf festgelegt, dass kein Faktor kleiner als 0.150 und keiner größer als 6.000 sein darf.

Für statistische Analysen, in deren Rahmen Rückschlüsse auf die Grundgesamt durchgeführt werden sollen, wird dringend empfohlen, die Gewichtungsfaktoren zu berücksichtigen.

Erweiterte Non-Response Analysen

Für die nunmehr vorliegende Aufstockungsstichprobe H wurden von TNS Infratest mikrogeographische Daten der Firma MICROM an die Bruttostichprobe angespielt. Dadurch liegt eine Vielzahl von (Proxy-) Variablen für die Modellierung von Ausfallgründen in wissenschaftlichen Analysen vor. Für die Aufstockungsstichprobe H liegen bislang vorläufige Ergebnisse von Nonresponse-Analysen von PD Jörg-Peter Schräpler, Professor Jürgen Schupp und Professor Gert G. Wagner vor.¹⁰ Die Modelle zur Schätzung der spezifizierten Einflussfaktoren unterscheiden zwischen drei Haupteinflussdimensionen: direkt erhobene Zielgruppencharakteristika, Interviewermerkmale und verschiedene mikrogeographische Proxy-Variablen, die das Haushaltungsfeld beschreiben. Dabei wurde zunächst der Einfluss der spezifizierten Variablen auf die Wahrscheinlichkeit ein Interview zu realisieren, untersucht, anschließend drei Ausfallgründe: erstens eine explizite „harte“ Verweigerung, zweitens die in der Feldzeit nicht mögliche Kontaktherstellung und drittens die Angabe, „nicht in der Lage“ zu sein, an der Befragung teilzunehmen. Letztere Angabe kann objektiv nachvollziehbare Gründe (z.B. gesundheitliche) haben, aber auch einen möglichen Hinweis auf eine Art „weiche“ Verweigerung geben, weil Zielpersonen in einer face-to-face Situation mit einem Interviewer nicht explizit verweigern möchten.

Die Ergebnisse der Non-Response Analysen sind komplex, noch als vorläufig einzustufen und werfen noch einige wichtige Fragen auf, insbesondere was die Robustheit der Schätzer für einzelne Variablengruppen angeht. Trotzdem werden sie hier präsentiert, weil sie einige interessante Tendenzen aufzeigen. Wir beschränken uns auf einige Kommentare zu wichtigen Ergebnissen, die für die Durchführung der Befragung einen unmittelbaren oder mittelbaren Bezug haben. Übersicht 1 gibt einen ausführlicheren Überblick.

Bei der Teilnahmebereitschaft zeigt sich, dass insbesondere Haushalte mit einem höheren sozialen Status und entsprechenden Lebensumfeldcharakteristika eher zur Teilnahme an der Befragung gewonnen werden können. Bestätigt wird auch der aus der einfachen Stichprobenstruktur sichtbare Befund, demzufolgende bei Zielhaushalten in Großstädten

¹⁰ Jörg-Peter Schräpler/Jürgen Schupp/Gert G. Wagner: Who Are the Nonrespondents? An Analysis Based on a New Subsample of the German Socio-Economic Panel (SOEP) including Microgeographic Characteristics and Survey-Based Interviewer Characteristics, unveröffentlichtes Manuskript Bochum/Berlin 2007.

besonders schwer Interviews realisiert werden können. Weitere Analysen bedarf es aus unserer Sicht zu dem über eine MICROM Proxy-Variable getesteten Alterseffekt, demzufolge die Teilnahmebereitschaft bei den unter 35-Jährigen überdurchschnittlich ist: Dies widerspricht Erfahrungen aus der Feldarbeit und liegt auch quer zur Stichprobenstruktur auf Personenebene.

Interviewerebene

Alter und Geschlecht des Interviewers sind für sich genommen keine signifikanten Prädiktoren für die Ausschöpfung, ein Interaktionsterm erweist sich indes als statistisch signifikant: Junge männliche Interviewer erreichen eher unterdurchschnittliche Ausschöpfungen, bei den Interviewerinnen ist kein Alterseffekt ersichtlich.

Ein – wenn man so will „positiver“ – Interviewereffekt zeigt sich im Hinblick auf das Ausmaß an Involvierung und damit Erfahrung der SOEP-Interviewer: je höher die Zahl der Haushalte, die der Interviewer bearbeitet, desto tendenziell höher ist die Ausschöpfung. Indirekt wird damit auch der Faktor „Erfahrung“ im Projekt erfasst: es entspricht der SOEP-Praxis, die Zahl zu betreuender Haushalte pro Interviewer an Erfahrungswerte mit dem Interviewer zu knüpfen. Direkt bestätigt wird damit eine wichtige strategische Weichenstellung für die Feldarbeit, nämlich zugunsten einer Einsatzpraxis, bei der keine niedrige Obergrenzen für die pro Interviewer maximale Anzahl an Interviews definiert wird, um Interviewereffekte zu begrenzen. Im Rahmen der Vorbereitung der SOEP-Haupterhebung wird bei der Planung des Interviewereinsatzes darauf geachtet, einen der Panelstabilität zuträglichen „positiven“ Interviewereffekt nutzbar zu machen: Interviewern, die sich im Rahmen des schwierigen, weil komplexen Projekts vielfach bewährt haben, werden in der Regel eine größere Anzahl von Haushalten zugeteilt als jüngeren und weniger erfahrenen Interviewern. Dies wird bei neuen Aufstockungsstichproben und notwendig werdenden Umbesetzungen von Haushalten ebenfalls berücksichtigt. Dadurch sinkt mittelfristig auch der Bedarf an Interviewerumbesetzungen, weil die erfahrenen und bewährten Interviewer mit einer vergleichsweise großen Zahl zu bearbeitender Haushalte in der Regel auch die längsten „project tenures“ aufweisen.

Ein dritter Befund, den es noch genauer zu prüfen gilt, zeigt, dass Interviewer, die in der Interviewerbefragung in besonderem Maße dazu geneigt haben, ihre Antworten an etablierten Normen sozialer Erwünschtheit auszurichten, überzufällig häufig Haushalte berichten, die nicht in der Lage seien, teilzunehmen. Dies könnte bei einem Teil der Interviewer ein erfolgskonformes Verhalten andeuten, also eine Neigung die Zahl der Verweigerungen zu reduzieren und statt dessen einen Haushalt im Zweifelsfall eher „als nicht in der Lage“ teilzunehmen, einzustufen. Infratest wird weitere Schritte unternehmen und prüfen, ob im Zweifelsfall tatsächlich einige Interviewer dazu tendieren, Haushalte als nicht teilnahmefähig einzustufen, obwohl es sich um Verweigerer handelt. Allerdings ist zu bedenken, dass „echte

Verweigerer“ auch dazu neigen anzugeben, dass sie nicht in der Lage sind an der Befragung teilzunehmen, um dem Interviewer eine vergleichsweise „freundliche Absage“ zu erteilen.

Haushaltsbezogene Proxy-Informationen

Die mikrogeographischen Angaben der Firma Microm liegen auf Straßenabschnittsebene vor. Die Angaben umfassen ein Vielzahl von Merkmalen wie Wohnumfeld, soziale Schichtungsmerkmale (SINUS-Milieus), Schätzungen der familien- und altersstrukturellen Zusammensetzung der Bewohner eines Straßenabschnitts und eine viele anderer Angaben. Wichtig ist dabei festzuhalten, dass es sich um Proxy-Informationen handelt, also nicht direkt auf Haushaltsebene erhobene Informationen. Die „Treffgenauigkeit“ von Proxy-Informationen kann variieren, zum Beispiel in Abhängigkeit des Ausmaßes an Homogenität bzw. Heterogenität in einer mikrogeografischen Einheit (etwa einem Straßenabschnitt). So sind für städtische Wohngebiete mit einer starken sozialen Durchmischung der Bewohner im Hinblick auf Alter, Familienstruktur, Bildungshintergrund oder Einkommen größere Ungenauigkeiten bei den Proxy-Informationen zu erwarten als in sozialstrukturell eher homogen zusammengesetzten Wohngebieten (etwa Straßenabschnitten mit überwiegend Einfamilien- oder Reihenhäusern in städtischen Vororten). Insofern sind Ergebnisse von komplexen, multivariaten Mehrebenenmodellen, mit Vorsicht zu interpretieren, wenn es um den Einfluss mikrogeographischen Proxy-Informationen handelt. Wir verzichten daher auf eine ausführliche Zusammenfassung der Nonresponse-Analysen.

Insgesamt zeigen die bislang vorliegenden Nonresponse Analysen statistisch signifikante Einflussgrößen auf verschiedenen Ebenen, der Interviewerebene, der Haushaltsebene und der geographischen Kontextebene. Die signifikanten partiellen Regressionskoeffizienten in multivariaten Mehrebenenmodellen sind im Kontext eher niedriger erklärter Varianz in der abhängigen Variable zu sehen. Die systematischen Effekte, die sich in den Non-Response-Analysen für einzelne Variablen zeigen, bedeuten also nicht, dass die Qualität der Stichprobe durch starke Verzerrungen leidet.

Übersicht über die Ergebnisse der Non-Response Analyse¹

	Teilnahme	Verweigerung	Kontaktherstellung	Angabe nicht in der Lage teilzunehmen
Einflussfaktoren Zielgruppe	+ wahrscheinlicher - weniger wahrscheinlich	+ wahrscheinlicher - weniger wahrscheinlich	+ schwieriger - leichter	+ wahrscheinlicher - weniger wahrscheinl.
Sinus Milieu	+ Etablierte	+ Bürgerliche Mitte + Experimentalisten + Moderne Performer	+: Experimentalisten +: Moderne Performer	+ Obere Schicht Konservative + Traditionsverwurzelte + Bürgerliche Mitte + Moderne Performer
Alter	+ Alter <= 35 Jahre	+ Alter > 45 - 50 years	+: Alter > 45-50 Jahre +: Alter > 55 - 60 Jahre	
Familienstruktur		+ Familien mit Kindern	+ Single-Haushalte	
Interviewer				
Geschlecht				+ männlicher Interviewer
Alter		+ Alter < 40 und männlich.		
Bildungsstatus		+ höherer Bildungsabschluss	+: höherer Bildungsabschluss	
Anzahl Haushalte	+ hohe Anzahl	+ niedrige Anzahl		+ niedrige Anzahl Haushalte
SOEP Erfahrung	+ große Erfahrung	+ niedrige Erfahrung		
Persönlichkeitsmerkmale	+ rücksichtsvoll	+ nicht rücksichtsvoll	+: nicht kreativ	+ ausgeprägte Tendenz zu sozialer Erwünschtheit
	+ hohe Lebenszufriedenheit + zurückhaltend	+ niedrige Lebenszufriedenheit + nicht zurückhaltend	+: nicht zurückhaltend +: lebhaftes Phantasie hat	+ wenig kommunikativ + geringe Lebenszufriedenheit
	+ nicht leicht nervös wird	+ hohes geduldig		+ eher faul + wenig wissbegierig + leicht nervös wird + nicht zurückhaltend + hohe Geduldbereitschaft
Gebiet	+ kein Großstadt	+ (Groß-)Städte	+: (Groß-)Städte	- keine (Groß-)Stadt
MOSAIC Typ	+ Familien mit hohem Einkommen + neue Privathäuser	+ Einfache städtische Zeilenbausiedlung - für soziale Aufsteiger, in Hochprestigeberufen	- Jüngere Dorfbewohner - Selbständige in guten neuen Einzelhäusern	- sozial schwache Kleinstädter weniger wahrscheinlich: gut verdienende Familien in neueren Eigenheimen
	+ Ältere Familien am Stadtrand + Selbständige in neuen Häusern + neue, qualitative hochwertige Doppelhaushälften + soziale Aufsteiger,	- Ältere Familien am Stadtrand - Peripherie: Dörfer in Randlagen - Selbständige in guten neuen Einzelhäusern	- alte Ortskerne - einfache Häuser im ländlichen Raum - einfache Berufe auf dem Land	- einfache Berufe auf dem Land

gehobene Berufe		
Umzugs- häufigkeit	+ hoch	+ größere Umzugswahrscheinlichkeit

1 Zusammenstellung auf Basis Schräpler/Schupp/Wagner 2007, S.25, Tabelle 7.

5. Datenprüfung und übergebener Datenbestand

Datenerfassung, Datenprüfung und Datenbereinigung erfolgen in der neuen SOEP-Stichprobe H nach denselben Verfahren wie in den SOEP-Altstichproben. Eine erneute Darstellung erübrigt sich hier; verwiesen wird auf die vorliegenden Methodenberichte zum SOEP.

Der an das DIW übergebene Datenbestand zur Stichprobe H umfasst folgende Bestandteile:

	Fallzahl
Nettodaten HAUSHALTS-Fragebogen	1.506
Nettodaten PERSONEN-Fragebogen	2.616
Bruttodaten Haushalte	3.931
Bruttodaten Personen ¹¹	3.407

¹¹ Nur Personen in Teilnehmer-Haushalten; einschließlich Kindern.