

Statistisches Bundesamt



Handbuch zum Mikrozensus-Panel 2001–2004

April 2009

Vorwort

Mit dem Mikrozensus-Panel 2001–2004 steht das zweite Mikrozensus-Panel der Wissenschaft als faktisch anonymisierter Datensatz zur Verfügung. Das vorliegende Handbuch soll den Nutzern/Nutzerinnen des faktisch anonymisierten Mikrozensus-Panels 2001–2004 den Umgang mit dem Datenmaterial erleichtern. Aufbau, Struktur und Inhalt orientieren sich weitgehend am Handbuch zum Mikrozensus-Panel 1996–1999. Die Abweichungen betreffen in erster Linie die Bildung des Mikrozensus-Panels sowie die intertemporalen Veränderungen in Bezug auf die Erhebungsmerkmale, die zwar im vorliegenden Handbuch nicht dokumentiert sind, aber dem Schlüsselverzeichnis entnommen werden können. Das Handbuch zum Mikrozensus-Panel 2001–2004 bietet also eine sukzessive Informationsverdichtung von einer allgemeinen Erläuterung des Mikrozensus über die generelle Nutzung des Mikrozensus als Datenquelle für Längsschnittauswertungen bis hin zur Beschreibung der Erstellung des faktisch anonymisierten Mikrozensus-Panels 2001–2004 und die dazugehörige Bereitstellung von Hochrechnungsfaktoren für Längsschnittauswertungen.

Im ersten Kapitel werden zunächst einige allgemeine Erläuterungen und methodische Informationen zum Mikrozensus selbst gegeben, die das Erhebungsdesign sowie das, für die Erhebungsjahre 1996 bis 2004 geltende, Frageprogramm betreffen. Des Weiteren wird kurz auf die hierarchische Datenstruktur des Mikrozensus und der sich hieraus ergebenden Erhebungs- und Analyseeinheiten eingegangen. Danach wird im zweiten Kapitel erläutert

- inwieweit der Mikrozensus als Datenquelle für Längsschnittauswertungen genutzt werden kann,
- wie das Mikrozensus-Panel 2001–2004 gebildet wurde,
- in welchem Umfang Datensätze längsschnittanalytisch analysiert werden können – angesprochen ist hier die Stichprobenentwicklung des Mikrozensus-Panels 2001–2004 –,
- welche Auswertungsstrategien in einem Mikrozensus-Panel herangezogen werden können,
- welche Längsschnittinformationen das Mikrozensus-Panel 2001–2004 enthält,
- wie der Datensatzaufbau des Mikrozensus-Panels 2001–2004 gestaltet ist und

- welche Erhebungs- und Analyseeinheiten in einem Mikrozensus-Panel vorliegen und in welchem Umfang sie genutzt werden können.

Welche allgemeinen und spezifischen Schutzvorkehrungen zur Erstellung des faktisch anonymisierten Mikrozensus-Panels 2001–2004 getroffen wurden (einschließlich des Auswahlplans für die Substichprobenziehung), welche Stichprobenentwicklung im faktisch anonymisierten Mikrozensus-Panel 2001–2004 vorliegt und wie die Hochrechnungsfaktoren für die Analyse von Querschnittsanalysen im faktisch anonymisierten Mikrozensus-Panel 2001–2004 gebildet wurden, sind Gegenstand des dritten Kapitels. Anders als das Mikrozensus-Panel 1996–1999 beinhaltet das Mikrozensus-Panel 2001–2004 auch Personen aus Gemeinschaftsunterkünften. Die im Querschnitt hochgerechneten Randverteilungen des faktisch anonymisierten Mikrozensus-Panels 2001–2004 können damit zwar mit den veröffentlichten Tabellen der Fachserien verglichen werden, dennoch wird an der Praxis festgehalten für ausgewählte Merkmale der Erhebungsjahre 2001 bis 2004 (unter Beachtung der jeweiligen Hochrechnung) die Randverteilungen der Mikrozensus-Originalfiles mit den entsprechenden Randverteilungen der Querschnitte aus dem faktisch anonymisierten Mikrozensus-Panel 2001–2004 gegenüberzustellen. Aufgrund des umfangreichen Datenmaterials sind die Tabellen in einer gesonderten Datei hinterlegt. Allerdings enthält das Handbuch zur Illustration einige wenige Tabellen. Mit kurzen Erläuterungen zu weiteren Metainformationen zum Mikrozensus-Panel, dem Schlüsselverzeichnis mit Beschreibungen zum Datensatzaufbau, Variablenamen und Variablenetiketten und den Klassifikationen zu den Wirtschaftszweigen, den Berufen und der Hauptfachrichtung des Abschlusses an einer (Fach-)Hochschule schließt das Kapitel ab.

Gewichtungsverfahren zur Behandlung des Selektivitätsbias werden im vierten Kapitel beschrieben. Im Rahmen des vom Bundesministerium für Bildung und Forschung¹ und der Deutschen Forschungsgemeinschaft geförderten Verbundprojektes "Aufbereitung und Bereitstellung des Mikrozensus als Panelstichprobe" wurden drei Ansätze verfolgt, die auch im Mikrozensus-Panel 2001–2004 herangezogen wurden. Der erste Ansatz schätzt Bleibewahrscheinlichkeiten durch ein Logit-Modell unter Berücksichtigung sozio-demografischer Merkmale. Eine Kalibrierung unter Verwendung bekannter Totalwerte der Grundgesamtheit verfolgt der zweite Ansatz von Gewichtungsverfahren. In einem dritten Ansatz werden die beiden vorgenannten Ansätze kombiniert, indem die geschätzten Bleibewahrscheinlichkeiten im Eingangsgewicht für die Kalibrierung verwendet werden.

Im abschließenden fünften Kapitel werden Methoden zur Aufdeckung und Reduktion von Verzerrungen infolge räumlicher Mobilität beispielhaft an Erwerbsübergängen erläutert. Bezüglich der Biasreduktion werden drei Verfahren vorgestellt:

1. der Einbezug von Kontrollvariablen,
2. die Gewichtung mit dem Kehrwert der Immobilitätswahrscheinlichkeit und

¹Förderkennzeichen des Projekts: 07SWF06A.

3. die Schätzung eines Modells für Kontingenztabellen mit fehlenden Werten.

Für die Umsetzung des ersten Verfahrens steht ein SAS-Makro zur Verfügung. Ein gesondertes Makro für den Gewichtungsansatz wird nichtbereitgestellt, da in den meisten Statistikanalyseprogrammen die Zuweisung von Gewichten durch ein Statement erfolgt. Die Schätzung von Modellen für Kontingenztabellen mit fehlenden Werten wurden mit dem Statistikprogramm *ℓEM* durchgeführt².

Bitte beachten Sie, dass dieses Kapitel unverändert aus dem Handbuch zum Mikrozensus-Panel 1996-1999 übernommen wurde. Die in diesem Kapitel enthaltenen Beispiele wurden demzufolge alle mit dem einem vorläufigen faktisch anonymisierten Mikrozensus-Panel 1996–1999 erstellt, das im Rahmen des zuvor erwähnten Verbundprojektes allen Verbundteilnehmern zur Verfügung stand.

²Das Statistikprogramm *ℓEM* wurde von Jeroen Vermunt entwickelt und steht kostenlos im Internet als Download zur Verfügung: URL: <http://www.uvt.nl/faculteiten/fsw/organisatie/departementen/mto/software2.html> (letzter Zugriff: 23. März 2008). Auf der selben Internetseite befindet sich auch ein detailliertes Benutzerhandbuch sowie zahlreiche Beispiele zum Download.

Inhaltsverzeichnis

1	Allgemeine und methodische Informationen zum Mikrozensus	17
1.1	Allgemeine Erläuterungen zum Mikrozensus	17
1.2	Das Erhebungsdesign des Mikrozensus ab 1990	21
1.3	Frageprogramm und Auswahlätze des Mikrozensus für die Erhebungsjahre 1996 bis 2004	26
1.4	Erhebungs- und Analyseeinheiten im Mikrozensus	29
1.5	Weiterführende Literatur	30
2	Der Mikrozensus als Datenquelle für Panelauswertungen	33
2.1	Vorbemerkung	33
2.2	Die Bildung des Mikrozensus-Panels 2001–2004	35
2.3	Die Stichprobenentwicklung des Mikrozensus-Panels 2001–2004	43
2.3.1	Umfang der Zusammenführungen im Mikrozensus-Panel 2001–2004 nach den Zusammenführungsstufen	43
2.3.2	Die Stichprobenentwicklung im Mikrozensus-Panel 2001–2004 auf der Personenebene	46
2.3.3	Die Stichprobenentwicklung im Mikrozensus-Panel 2001–2004 auf der Haushaltsebene	49
2.4	Auswertungskonzepte in einem Mikrozensus-Paneldatensatz	51
2.5	Längsschnittinformationen im Mikrozensus-Paneldatensatz 2001–2004	56
2.6	Der Datensatzaufbau des Mikrozensus-Panels 2001–2004	60
2.7	Erhebungs- und Analyseeinheiten im Mikrozensus-Panel	61
2.8	Weiterführende Literatur	62
3	Das faktisch anonymisierte Mikrozensus-Panel 2001–2004	63
3.1	Maßnahmen zur Erstellung des faktisch anonymisierten Mikrozensus-Panels 2001–2004	64
3.1.1	Allgemeine Schutzvorkehrungen	64
3.1.2	Spezifische Anonymisierungsmaßnahmen	65
3.2	Die Stichprobenentwicklung des faktisch anonymisierten Mikrozensus-Panels 2001–2004	68

3.3	Das Variablenspektrum des faktisch anonymisierten Mikrozensus-Panels 2001–2004	75
3.4	Die Bildung angepasster Hochrechnungsfaktoren für Querschnittsanalysen im Mikrozensus-Panel 2001–2004	76
3.5	Randverteilungen ausgewählter Merkmale in den Mikrozensus-Originalfiles und den Querschnitten des faktisch anonymisierten Mikrozensus-Panels 2001–2004	80
3.6	Metainformationen zum Mikrozensus-Panel 2001–2004	87
3.6.1	Das Schlüsselverzeichnis des faktisch anonymisierten Mikrozensus-Panels 2001–2004	88
3.6.2	Das systematische Verzeichnis der Wirtschaftszweige des Statistischen Bundesamtes in der Fassung für den Mikrozensus – Ausgabe 1993	92
3.6.3	Das systematische Verzeichnis der Wirtschaftszweige des Statistischen Bundesamtes in der Fassung für den Mikrozensus – Ausgabe 2003	93
3.6.4	Klassifikationen beruflicher Tätigkeiten	96
3.6.5	Gliederung der Hauptfachrichtung des Abschlusses an einer (Fach-)Hochschule bzw. eines beruflichen Abschlusses – Ausgabe 2003	99
3.7	Weiterführende Literatur	100
4	Das Hochrechnungsverfahren für Längsschnittauswertungen im Mikrozensus-Panel 2001–2004	101
4.1	Die Grundidee der Hochrechnung	101
4.2	Verallgemeinerte Schätzgleichungen	103
4.3	Grundgesamtheit und Stichprobe des Mikrozensus-Panels 2001–2004	105
4.4	Die Kalibrierung des Mikrozensus-Panels 2001–2004	106
4.5	Die im Mikrozensus-Panel 2001–2004 bereitgestellten Hochrechnungs- und Mobilitätsfaktoren	110
4.6	Weiterführende Literatur	111
5	Möglichkeiten zur Korrektur von Verzerrungen aufgrund räumlicher Immobilität	113
5.1	Das Auftreten von räumlicher Mobilität	113
5.2	Methoden zur Biasaufdeckung	114
5.2.1	Beispiel	116
5.2.2	Die Größenordnung des Mobilitätsbias	117
5.2.3	Makro hausman.sas	118
5.3	Methoden zur Biasreduktion	118
5.3.1	Verwendung von Kontrollvariablen	119
5.3.1.1	Makro logit.sas	122

5.3.2	Gewichtung mit dem Kehrwert der Immobilitätswahrscheinlichkeit	123
5.3.2.1	Beispiel	124
5.3.2.2	Makro	127
5.3.3	Schätzung eines Modells für Kontingenztabelle mit fehlenden Werten	127
5.3.3.1	Beispiel	129
5.3.3.2	IEM Programme	131
5.4	Weiterführende Literatur	137

Tabellenverzeichnis

Tabelle 1.1: Erhebungstermine und Auswahlsätze der Merkmalsbereiche der Mikrozensus-erhebungen 1996 bis 2004	26
Tabelle 1.2: Anzahl der zu befragenden Personen und Anteil der in das Unterstichprobenprogramm einbezogenen Personen im dritten Rotationsviertel der fünften Stichprobe in den einzelnen Erhebungsjahren	29
Tabelle 2.1: Kombinationstypen bezüglich des Vorliegens bzw. Nicht-Vorliegens von Informationen in einem Mikrozensus-Panel über vier Erhebungszeitpunkte	37
Tabelle 2.2: Konstellation von Kombinationstypen und Kennung von Verlusten und Gewinnen, die zu einem plausibel zusammengeführten Personendatensatz führen	38
Tabelle 2.3: Zugelassene Abweichungen von ± 1 Jahr im Geburtsjahr in „Abhängigkeit erklärter“ Haushaltstypen	41
Tabelle 2.4: (Un-)Plausible Datensätze im Mikrozensus-Panel 2001–2004, differenziert nach Erhebungsjahr, Haushaltstyp und Zusammenführungsebene	45
Tabelle 2.5: Umfang (un-)plausibler Datensätze auf der Personenebene im Mikrozensus-Panel 2001–2004, differenziert nach Haushaltstypen in den einzelnen Erhebungsjahren	48
Tabelle 2.6: Umfang (un-)plausibler Datensätze auf der Haushaltsebene im Mikrozensus-Panel 2001–2004, differenziert nach Haushaltstypen in den einzelnen Erhebungsjahren	50
Tabelle 2.7: Zusammenhang von Auswertungskonzept und Werteselektion erklärter Personensätze	56
Tabelle 2.8: Längsschnittinformationen im Mikrozensus-Panel 2001–2004	57
Tabelle 3.1: Auswahlplan der Substichprobenziehung zur Generierung eines faktisch anonymisierten Mikrozensus-Panels 2001–2004	66
Tabelle 3.2: Die Stichprobenentwicklung des faktisch anonymisierten Mikrozensus-Panels 2001–2004 auf der Personenebene	69

Tabelle 3.3: Die Stichprobenentwicklung des faktisch anonymisierten Mikrozensus-Panels 2001–2004 auf der Haushaltsebene	72
Tabelle 3.4: Auswahlwahrscheinlichkeit für eine Person aus der Unterstichprobe des Mikrozensus-Originalfiles in die Unterstichprobe des faktisch anonymisierten Mikrozensus-Panels 2001–2004 zu gelangen	77
Tabelle 3.5: Variablennamen und -labels der angepassten Hochrechnungsfaktoren in den einzelnen Erhebungsjahren des Mikrozensus-Panels 2001–2004	79
Tabelle 3.6: Bevölkerung am Ort der Hauptwohnung nach ausgewählten Merkmalen im MZ-Originalfile und dem Querschnitt des MZ-Panels im Erhebungsjahr 2002	82
Tabelle 3.7: Metainformationen zum Mikrozensus-Panel 2001–2004	87
Tabelle 3.8: Ausfallart sowie zugehöriger Variablenwert und Wertelabel im Mikrozensus-Paneldatensatz 2001–2004	91
Tabelle 3.9: Die Gliederungsebenen der Klassifikation der Wirtschaftszweige des Statistischen Bundesamtes in der Ausgabe von 1993	92
Tabelle 3.10: Differenzen zwischen der Kodierung und Aufgliederung von NACE Rev. 1.1 und WZ 2003	94
Tabelle 3.11: Die Gliederungsebenen der Klassifikation der Wirtschaftszweige von ISIC Rev. 3.1, NACE Rev. 1.1 und WZ 2003	95
Tabelle 3.12: Änderungen und zugehörige Kennzeichnungen zwischen den Wirtschaftszweige klassifikationen WZ 2003 und WZ 93	96
Tabelle 3.13: Die Gliederungsebenen der Klassifikation der Berufe des Statistischen Bundesamtes in der Ausgabe von 1992	96
Tabelle 3.14: Die Gliederungsebenen der Internationalen Standardklassifikation der Berufe zur Verwendung innerhalb der Europäischen Gemeinschaft in der Ausgabe von 1988	97
Table 4.1: Kovariablen in den verallgemeinerten Schätzgleichungen zur Ermittlung der Bleibewahrscheinlichkeiten im Mikrozensus-Panel 2001–2004	104
Table 4.2: Variablennamen der einzelnen Längsschnitt- und Mobilitätsgewichte im Mikrozensus-Panel 2001–2004	110
Tabelle 5.1: Übergangswahrscheinlichkeiten zwischen Arbeitslosigkeit bzw. Nichterwerbsperson (1996) und Erwerbstätigkeit (t=1997, 1998, 1999). Datenbasis: SOEP	117
Tabelle 5.2: Übergangswahrscheinlichkeiten zwischen Arbeitslosigkeit bzw. Nichterwerbsperson (1996) und Erwerbstätigkeit (t=1997, 1998, 1999) bei Kontrolle nach Altersgruppe. Datenbasis: SOEP	120
Tabelle 5.3: Logistische Regression: Übergang von Arbeitslosigkeit (1996) nach Erwerbstätigkeit (1998). Datenbasis: SOEP	121

Tabelle 5.4: Logistische Regression: Übergang von Nichterwerbsperson (1996) nach Erwerbstätigkeit (1998). Datenbasis: SOEP	122
Tabelle 5.5: (Im-)mobilitätsmodell: Zeitintervall 1996-1998	125
Tabelle 5.6: Verhältnis von Biaskorrektur zu Bias für die Benutzung der GewichtungsvARIABLEN über die Immobilitätswahrscheinlichkeit. Datenbasis: SOEP und MZ-Panel	126
Tabelle 5.7: Verhältnis von Biaskorrektur zu Bias bei Benutzung eines loglinearen Modells. Datenbasis: SOEP und MZ-Panel	130
Tabelle 5.8: Verhältnis von Biaskorrektur zu Bias bei Benutzung eines loglinearen Modells. Datenbasis: SOEP und MZ-Panel	131

Abbildungsverzeichnis

Abbildung 2.1: Zusammenführbare Rotationsviertel der Mikrozensus-erhebungen 1996 bis 2004	35
Abbildung 2.2: Umfang der Analyseeinheiten in einer Paneluntersuchung mit vier Befragungszeitpunkten ohne temporäre Ausfälle nach dem Analysekonzept mit Berücksichtigung aller verfügbaren Fälle – schematische Darstellung	52
Abbildung 2.3: Umfang der Analyseeinheiten in einer Paneluntersuchung mit vier Befragungszeitpunkten ohne temporäre Ausfälle nach dem Weiterverfolgungskonzept – schematische Darstellung	53
Abbildung 2.4: Umfang der Analyseeinheiten in einer Paneluntersuchung mit vier Befragungszeitpunkten ohne temporäre Ausfälle nach dem Rückverfolgungskonzept – schematische Darstellung	54
Abbildung 2.5: Umfang der Analyseeinheiten in einer Paneluntersuchung mit vier Befragungszeitpunkten ohne temporäre Ausfälle nach dem Analysekonzept mit Berücksichtigung der vollständigen Fälle– schematische Darstellung	55
Abbildung 4.1: Definition der beiden betrachteten Mengen für die Randanpassung des Längsschnitts 2001–2004	107

Kapitel 1

Allgemeine und methodische Informationen zum Mikrozensus

Robert Herter-Eschweiler (Statistisches Bundesamt, Gruppe VIII C)

Hier im ersten Kapitel des Handbuchs erfolgt zunächst eine allgemeine Beschreibung des Mikrozensus, eine Erläuterung zu seinem, seit 1990 bestehenden Erhebungsdesign und eine Darstellung seines Frageprogramms, das für die Geltungsdauer des Mikrozensusgesetzes vom Januar 1996 angewendet wurde.

1.1 Allgemeine Erläuterungen zum Mikrozensus

Was ist der Mikrozensus?	Der Mikrozensus ist eine amtliche Stichprobenerhebung über die Bevölkerung und den Arbeitsmarkt, an der etwa 1 Prozent aller in Deutschland wohnenden Personen beteiligt sind.
Seit wann gibt es den Mikrozensus?	Den Mikrozensus gibt es im früheren Bundesgebiet bereits seit 1957 und in den neuen Bundesländern und Berlin-Ost seit 1991.
Ist die Durchführung des Mikrozensus gesetzlich geregelt?	Ja. Rechtsgrundlage ist das Gesetz zur Durchführung einer Repräsentativstatistik über die Bevölkerung und den Arbeitsmarkt sowie die Wohnsituation der Haushalte (Mikrozensusgesetz 1996 – MZG 1996) vom 17. Januar 1996 (BGBl. I S. 34) – (URL: http://bundesrecht.juris.de/bundesrecht/mikrozensusg_1996/gesamt.pdf (letzter Zugriff: 23. März 2008)) – in Verbindung mit dem Bundesstatistikgesetz (BStatG) vom 22. Januar 1987 (BGBl. I S. 462, 565), zuletzt geändert durch Artikel 16 des Gesetzes vom 21. August 2002 (BGBl. I S. 3222) – (URL: http://www.gesetze-im-internet.de/bundesrecht/bstatg_1987/gesamt.pdf (letzter Zugriff: 23. März 2008)) – .

Ist das Mikrozensusgesetz zeitlich befristet?	Ja. Damit der Gesetzgeber die Rechtsgrundlage des Mikrozensus an den gesellschaftlichen Wandel anpassen kann, ist das Mikrozensusgesetz 1996 – wie auch die vorangegangenen Mikrozensusgesetze – zeitlich befristet und gilt für die Jahre 1996 bis 2004.
Welche Aufgaben hat der Mikrozensus?	<p>Der Mikrozensus dient der Bereitstellung statistischer Informationen über die wirtschaftliche und soziale Lage der Bevölkerung sowie über die Erwerbstätigkeit, den Arbeitsmarkt und die Ausbildung (Mehrzweckstichprobe). Er schreibt die Ergebnisse der Volkszählung in den Jahren fort, in denen es keine gibt.</p> <p>Darüber hinaus dient er der Rationalisierung anderer amtlicher Statistiken, wie z.B. der Einkommens- und Verbrauchsstichprobe sowie als Hochrechnungsrahmen nicht amtlicher Stichprobenerhebungen.</p>
Wofür werden die Mikrozensusergebnisse verwendet?	Die Mikrozensusergebnisse sind eine unverzichtbare Informationsquelle für Parlament, Regierung, Verwaltung, Wissenschaft und Öffentlichkeit in Bund und Ländern. Sie gehen in Regierungsberichte, in das Jahresgutachten des Sachverständigenrates zur Begutachtung der gesamtwirtschaftlichen Entwicklung ein und bilden die Grundlage für die laufende Arbeitsmarkt- und Berufsforschung, den jährlichen Rentenversicherungsbericht der Bundesregierung und vieles andere mehr.
Wer führt die Erhebung durch?	Die organisatorische und technische Vorbereitung des Mikrozensus erfolgt im Statistischen Bundesamt. Die Durchführung der Befragung und die Aufbereitung obliegt den Statistischen Landesämtern (dezentrale Statistik).
In welchen Abständen wird der Mikrozensus durchgeführt?	Der Mikrozensus wird jährlich mit einer festen Berichtswoche erhoben. In der Regel handelt es sich hierbei um die letzte feiertagsfreie Woche im April eines Jahres.
Wie lange bleiben die Personen in der Stichprobe?	Im Mikrozensus wird jährlich ein Viertel aller in der Stichprobe enthaltenen Auswahlbezirke ausgetauscht (Verfahren der partiellen Rotation). Folglich verbleibt jede Person maximal drei Jahre in der Stichprobe (entspricht vier Erhebungszeitpunkte)
Wer wird befragt?	Befragt werden alle Personen eines Auswahlbezirks, die nach den §§ 12 bis 16 des Melderechtsrahmengesetzes meldepflichtig sind.
Wie wird die Befragung durchgeführt?	Der Mikrozensus ist überwiegend eine persönliche Befragung durch Interviewer/innen (face-to-face). Die Interviewbefragung ist die schnellste und für die Befragten mit dem geringsten Aufwand verbundene Befragungsmethode. Allerdings steht den Befragten auch offen, die Auskünfte schriftlich zu erteilen.

Zu welchen Themen werden die einzelnen Angaben erhoben?

Das jährliche Grundprogramm des Mikrozensus umfasst u.a. Merkmale der Person (Alter, Geschlecht, Staatsangehörigkeit usw.), den Familien- und Haushaltszusammenhang sowie darüber hinaus die Merkmale Haupt- und Nebenwohnung, Erwerbstätigkeit, Arbeitsuche, Arbeitslosigkeit, Nichterwerbstätigkeit, Kind im Vorschulalter, Schüler/in, Student/in, allgemeiner und beruflicher Ausbildungsabschluss, Quellen des Lebensunterhalts sowie Angaben zur gesetzlichen Rentenversicherung, zur Pflegeversicherung und zur Höhe des Individual- und Haushaltseinkommens.

Im jährlichen Ergänzungsprogramm werden u.a. zusätzliche Fragen zur Erwerbstätigkeit gestellt und Angaben zu einer früheren Erwerbstätigkeit sowie zur beruflichen und allgemeinen Aus- und Fortbildung erfragt.

Im Rahmen der vierjährigen Zusatzprogramme werden u.a. Angaben zum Berufs- und Ausbildungspendeln, zur Wohnsituation, zur Krankenversicherung und zur Gesundheit und Behinderteneigenschaft erhoben.

Wie werden die Mikrozensusergebnisse dargestellt?

Mikrozensusergebnisse werden als Summenwerte für die Darstellungseinheiten Personen, Lebensgemeinschaften, Familien oder Haushalte zur Verfügung gestellt. Neben Bundesergebnissen sind auch vielfältige Ergebnisse für die einzelnen Bundesländer verfügbar. Einer tieferen regionalen Gliederung sind jedoch Grenzen gesetzt, so dass nur Ergebnisse für große Nachweisungsgruppen zur Verfügung gestellt werden können.

Wo werden die Ergebnisse des Mikrozensus veröffentlicht?

Die Bundesergebnisse des Mikrozensus werden als Online-Produkte im Statistik-Shop und in den verschiedenen Fachserien (Fachserie 1, Reihe 4.1.1: Stand und Entwicklung der Erwerbstätigkeit; Reihe 4.1.2: Beruf, Ausbildung und Arbeitsbedingungen der Erwerbstätigen; Fachserie 1, Reihe 3: Haushalte und Familien; Fachserie 13, Reihe 1: Versicherte in der Kranken- und Rentenversicherung sowie Fachserie 12, Reihe S.3: Fragen zur Gesundheit (unregelmäßig)) und Querschnittsveröffentlichungen des Statistischen Bundesamtes veröffentlicht. Zu aktuellen Themen werden Ergebnisse in einer Pressekonferenz und Pressemitteilungen bekannt gegeben. Die Pressemitteilungen sind kostenlos im Internetangebot des Statistischen Bundesamtes abrufbar. Neben Bundesergebnissen sind auch vielfältige Ergebnisse für die Länder verfügbar, die von den jeweiligen Statistischen Landesämtern veröffentlicht werden.

**Wie genau ist der
Mikrozensus?**

Auf Grund des Stichprobencharakters des Mikrozensus ist ein Stichprobenfehler zu berücksichtigen. Für hochgerechnete Besetzungszahlen unter 5 000, d.h. für weniger als 50 Fälle in der Stichprobe, geht der einfache relative Standardfehler über 15 Prozent hinaus. Solche Ergebnisse haben nur noch einen geringen Aussagewert. Hochgerechnete Besetzungszahlen unter 5 000 werden demzufolge in den Veröffentlichungen des Mikrozensus nicht ausgewiesen.

**Wie vollständig sind
die Auskünfte der
Befragten?**

Die Antwortquote im Mikrozensus beträgt im Bundesdurchschnitt etwa 97 Prozent; 3 Prozent fallen aus, überwiegend weil sie nicht erreichbar sind. Diese so genannten bekannten Ausfälle werden im Mikrozensus durch ein besonderes Verfahren (Kompensationsverfahren) ausgeglichen. Bei Fragen mit freiwilliger Auskunftserteilung sind die "Ohne-Angabe"-Quoten je nach Merkmal sehr unterschiedlich.

**Integration der
EU-Arbeitskräfte-
erhebung**

Seit 1968 ist die Europäische Arbeitskräfteerhebung in den Mikrozensus integriert. Es bot sich an, weil der größte Teil des Mikrozensus-Frageprogramms mit dem der EU-Arbeitskräfteerhebung übereinstimmt. Damit reduziert sich die Belastung der Befragten, der organisatorische Aufwand sowie die Kosten.

Die Ergebnisse der EU-Arbeitskräfteerhebung werden unter anderem herangezogen für die Verteilung der Mittel aus dem EU-Regional- und Sozialfonds.

Geregelt wird die EU-Arbeitskräfteerhebung in der Verordnung (EWG) Nr. 3711/1991 des Rates vom 16. Dezember 1991 zur Durchführung einer jährlichen Stichprobenerhebung über Arbeitskräfte in der Gemeinschaft (Abl. EG Nr. L 351, S.1). Diese wurde ersetzt durch die Verordnung (EG) Nr. 577/98 des Rates vom 9. März 1998 zur Durchführung einer Stichprobenerhebung über Arbeitskräfte in der Gemeinschaft (URL: http://eur-lex.europa.eu/LexUriServ/site/de/oj/1998/l_077/l_07719980314de00030007.pdf) (letzter Zugriff: 23. März 2008).

1.2 Das Erhebungsdesign des Mikrozensus ab 1990

Auswahlgrundlage	Alte Bundesländer: Volkszählung 1987 Neue Bundesländer: Zentrales Einwohnerregister 1990
Aktualisierung	Jährlich mit Hilfe der Bautätigkeitsstatistik
Auswahlverfahren	Einstufig geschichtete Klumpenstichprobe (Klumpen = Auswahlinheiten)
Auswahlseinheiten	Auswahlseinheiten sind so genannte Auswahlbezirke. Sie werden aus mehreren bewohnten Gebäuden, einem einzelnen bewohnten Gebäude oder Teilen eines bewohnten Gebäudes mit folgenden Zuordnungsregeln zu den einzelnen Gebäudegrößenklassen gebildet.

Nr.	Grundauswahl	Aktualisierung
1	Gebäude mit 1 bis 4 Wohnungen, die zu Auswahlbezirken mit einer Richtgröße von 12 Wohnungen zusammengefasst werden. Die Zusammenfassung erfolgte soweit möglich in der Reihenfolge der Hausnummern in einer Straße.	Gebäude mit 1 bis 4 Wohnungen, die zu Auswahlbezirken mit einer Richtgröße von 6 Wohnungen zusammengefasst wurden.
2	Gebäude mit 5 bis 10 Wohnungen, die jeweils eigene Auswahlbezirke bilden.	Gebäude mit 5 bis 8 Wohnungen, die jeweils eigene Auswahlbezirke bilden.
3	Gebäude mit 11 oder mehr Wohnungen werden in Auswahlbezirke mit einer Richtgröße von 6 Wohnungen (unter weitgehender Verwendung vollständiger Geschosse) zerlegt.	Gebäude mit 9 oder mehr Wohnungen werden in Auswahlbezirke mit einer Richtgröße von 6 Wohnungen (unter weitgehender Verwendung vollständiger Geschosse) zerlegt.

Fortsetzung der Übersicht auf nachfolgender Seite.

Fortsetzung der Übersicht.

Nr.	Grundausswahl	Aktualisierung
4	<p>Gebäude mit vermuteten Gemeinschaftsunterkünften bilden eine eigene Schicht <i>"und zwar genau dann, wenn die Zahl der Personen eines Gebäudes mindestens $4*(k+4)$ betrug. ($k \geq 1$: Zahl der Wohnungen des Gebäudes) oder wenn ein bewohntes Gebäude keine Wohnung aufwies"</i> (Meyer 1994: 107).</p> <p>Die Richtgröße beträgt 15 Personen. Umfasst eine Gemeinschaftsunterkunft deutlich mehr als 15 Personen, wird eine Teilung der Gemeinschaftsunterkunft über die Anfangsbuchstaben des Familiennamens der in der Gemeinschaftsunterkunft lebenden Personen vorgenommen. Jeder dieser Teile bildet dann einen eigenen Auswahlbezirk.</p>	<p>Es gelten die gleichen Regeln wie bei der Grundausswahl.</p> <p>Aus den Meldungen der Bautätigkeitsstatistik ist jedoch bekannt, ob eine Gemeinschaftsunterkunft vorliegt oder nicht.</p>

Schichtung

Bundesland, Regierungsbezirk, regionale Anpassungsschicht, regionale Schicht und Gebäudegrößenklasse.

Regionale Anpassungsschichten sind regionale Einheiten mit durchschnittlich 500 000 Einwohnern, die durch Zusammenlegungen von Kreisen und kreisfreien Städten bzw. einzelnen kreisfreien Städten gebildet wurden.

Regionale Schichten sind regionale Einheiten, die durch Zusammenlegungen von Kreisen und kreisfreien Städten gebildet wurden. Ihre durchschnittliche Einwohnerzahl beziffert sich auf durchschnittlich 250 000 Einwohner, wobei Großstädte mit 200 000 oder mehr Einwohnern eine eigene regionale Schicht bilden und sehr große Städte wie beispielsweise Berlin und Hamburg in mehrere regionale Schichten unterteilt wurden.

Fortsetzung der Übersicht auf nachfolgender Seite.

Fortsetzung der Übersicht.

Auswahltechnik**Grundauswahl:**

1. Sortierung der Auswahlbezirke nach regionalen Untergruppen (= regionale Einheiten von etwa 100 000 Einwohnern oder mehr, die nicht zwingend kreis-scharf sind), Kreis, Gemeindegrößenklasse, Gemeinde und Auswahlbezirksnummer.
2. Je 100 in der Reihenfolge der Sortierung aufeinander folgende Auswahlbezirke bilden eine so genannte "Zone".
3. Zufällige Zuordnung der Auswahlbezirke einer Zone zu den Zahlen 0–99 (= Stichprobennummer).

Zusammenfassung aller Auswahlbezirke mit gleicher Stichprobennummer, womit eine Zerlegung der Gesamtheit in 100 1%–Stichproben gegeben ist. Vier in der Reihenfolge der Sortierung aufeinander folgende Zonen bilden einen Block.
4. Zufällige Zuordnung der Zahlen 1 bis 4 an jede Zone eines Blocks.

Auswahlbezirke mit gleicher Zonennummer gehören dem gleichen Rotationsviertel an, womit eine Zerlegung der 1%–Stichproben in vier Rotationsviertel zu je 0,25% gegeben ist.
5. Ermittlung von 20 1% - Stichproben durch zufällige Ziehung der Ordnungsnummern der Stichproben.

Aktualisierung:

1. Einordnung der Neubaugebiete in Gebäudegrößenklassen, die annähernd gleich groß sind und pro regionaler Schicht in nur einer Schicht – der Neubauschicht – zusammengefasst werden.
2. Sortierung der Neubaugebiete nach Aktualisierung und regionaler Kennung.
3. Systematische Auswahl mit Zufallsstart und entsprechender Zuweisung auf die Rotationsviertel und weiteren Teilstichproben.

Fortsetzung der Übersicht auf nachfolgender Seite.

Fortsetzung der Übersicht.

Hochrechnung

Zweistufiges Verfahren:

1. Kompensation der bekannten Ausfälle auf der Haushaltsebene innerhalb der regionalen Untergruppen für jeweils 19 Kompensationsklassen, die sich aus der Kombination folgender Merkmale ergeben:
 - (a) Haushaltsgröße (differenziert nach 1, 2, 3 oder mehr Personen),
 - (b) Staatsangehörigkeit der Haushaltsbezugsperson (Deutsche/r, Ausländer/in),
 - (c) für Deutsche: Wohnrechtsverhältnis der Haushaltsbezugsperson (Haupt-/Nebenwohnsitz),
 - (d) für 1-Personenhaushalte:
 - Geschlecht,
 - Altersgruppenzugehörigkeit (unter 60 Jahre, 60 Jahre oder älter).
2. Anpassung der Stichprobenergebnisse an Eckzahlen aus der laufenden Bevölkerungsfortschreibung auf der Ebene der regionalen Anpassungsschichten für Deutsche und Ausländer/innen in der Gliederung nach dem Geschlecht.

Für Soldaten/Soldatinnen und Grundwehr- sowie Zivildienstleistende erfolgt eine gesonderte Anpassung auf der Ebene der Regierungsbezirke bzw. Bundesländer auf der Basis von Bestandsmeldungen des Verteidigungs- bzw. Innenministeriums.

Aus der Multiplikation des haushaltsbezogenen Kompensationsfaktors und des personenbezogenen Anpassungsfaktors ergibt sich der endgültige Personenfaktor.

Da der Personenfaktor zwischen den Personen eines Haushalts variiert, sind die Ergebnisse haushaltsbezogener Auswertungen abhängig von der Person, die als Haushaltsstellvertreter/in eingesetzt wird. Zur Aufhebung dieser Inkohärenz wurde ein Haushaltsfaktor gebildet, der dem ungewogenen arithmetischen Mittel der Personenfaktoren der zum Haushalt gehörenden Personen entspricht.

Erhebungseinheiten

Alle Personen, die nach den §§ 12 bis 16 des Melderechtsrahmengesetzes meldepflichtig in einem, in die Stichprobe einbezogenen, Auswahlbezirk wohnen.

Fortsetzung der Übersicht auf nachfolgender Seite.

Fortsetzung der Übersicht.

Teilnahme	Verpflichtend nach § 7, Abs. 1 des Mikrozensusgesetzes vom 17. Januar 1996 (MZG 1996) (BGBl. I S. 34).
Auskunftserteilung	Verpflichtend nach § 7 MZG 1996; bei einigen Fragen hat der Gesetzgeber jedoch eine freiwillige Beantwortung zugelassen (§ 7, Abs. 4 MZG 1996).
Proxy-Angaben	Zwar besteht generell für jedes volljähriges Haushaltsmitglied eine Auskunftspflicht, Auskünfte über andere Haushaltsmitglieder können aber nach § 7, Abs. 2, Nr. 1 MZG 1996 von einer "Vertrauensperson" im Haushalt gegeben werden (= so genannte Proxy-Angaben).
Erhebungsmethoden	Nach § 8 MZG 1996 ist ein Methodenmix zugelassen: <ul style="list-style-type: none">• mündlich gegenüber einem/einer Interviewer/in oder• schriftlich als Selbstausfüller/in.
Zeitbezug der Angaben	Feste Berichtswoche; in der Regel die letzte feiertagsfreie Woche im April eines Jahres.
Durchführung	Die technisch-organisatorische und konzeptionelle Vorbereitung erfolgt im Statistischen Bundesamt, wohingegen die Durchführung und die Aufbereitung den Statistischen Landesämtern obliegt (dezentrale Statistik).

1.3 Frageprogramm und Auswahlsätze des Mikrozensus für die Erhebungsjahre 1996 bis 2004

Alle Haushalte und Personen, die in den ausgewählten Auswahlbezirken wohnberechtigt sind, sind als Erhebungseinheiten im Mikrozensus zu gesetzlich festgelegten Tatbeständen aus verschiedenen Themenbereichen (Mehrzweckstichprobe) zu befragen (siehe § 4 MZG 1996). Die festgelegten Tatbestände können in ein Grund-, ein Ergänzungs- und ein Zusatzprogramm untergliedert werden. Die Fragen zum Grundprogramm werden mit vollem Auswahlatz von 1 Prozent jährlich erhoben. Ebenfalls jährlich erhoben werden die Fragen zum Ergänzungsprogramm, im Gegensatz zum Grundprogramm jedoch mit einem verringerten Auswahlatz, der im Bundesdurchschnitt maximal 0,5 Prozent betragen kann. Die Fragen des Zusatzprogramms werden in einem Abstand von vier Jahren erhoben, teils mit dem Standardauswahlatz von 1 Prozent, teils mit einem verringerten Auswahlatz von bundesdurchschnittlich maximal 0,5 Prozent. Die nachfolgende Tabelle weist die einzelnen Merkmalsbereiche nach Erhebungsjahr und Auswahlatz aus. Rechtsgrundlage ist das Mikrozensusgesetz von 1996.

Tabelle 1.1: Erhebungstermine und Auswahlätze der Merkmalsbereiche der Mikrozensususerhebungen 1996 bis 2004

Merkmalsbereich (— = Auskunftserteilung freiwillig)	§ 4 MZG 1996	Erhebungsjahr und Auswahlatz in %									
		1996	1997	1998	1999	2000	2001	2002	2003	2004	
1 Grundprogramm											
1.1 Merkmale der Person, Familien-, Haushaltszusammenhang, Staatsangehörigkeit, Haupt- und Nebenwohnung ¹⁾	Abs. 1, Nr. 1a,k	1	1	1	1	1	1	1	1	1	
1.2 Angaben zur gesetzlichen Rentenversicherung	Abs. 1, Nr. 1b	1	1	1	1	1	1	1	1	1	
1.3 Angaben zur Pflegeversicherung	Abs. 1, Nr. 1b	1	1	1	1	—	—	—	1	—	
1.4 Quelle des Lebensunterhalts, Höhe des Einkommens	Abs. 1, Nr. 1c	1	1	1	1	1	1	1	1	1	
1.5 Allgemeine und berufliche Ausbildung, Besuch von Kindergarten, Schule, Hochschule ²⁾	Abs. 1, Nr. 1d,e	1	1	1	1	1	1	1	1	1	
1.6 Erwerbstätigkeit, Arbeitslosigkeit und Arbeitsuche, Nichterwerbspersonen	Abs. 1, Nr. 1f-j	1	1	1	1	1	1	1	1	1	

Fortsetzung der Tabelle auf nachfolgender Seite

Fortsetzung der Tabelle 1.1

Merkmalsbereich (— = Auskunftserteilung freiwillig)	§ 4 MZG 1996	Erhebungsjahr und Auswahlsatz in %									
		1996	1997	1998	1999	2000	2001	2002	2003	2004	
2	Ergänzungsprogramm³⁾										
2.1	Berufliche und allgemeine Aus- und Fortbildung	Abs. 1, Nr. 2a	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0,5
2.2	Ergänzende Angaben zur Erwerbstätigkeit	Abs. 1, Nr. 2b	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0,5
2.3	Frühere Erwerbstätigkeit	Abs. 1, Nr. 2c	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0,5
2.4	Situation ein Jahr vor der Erhebung	Abs. 1, Nr. 2d	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0,5
2.5	Pflegebedürftigkeit, Leistungen einer Pflegeversicherung	Abs. 1, Nr. 2e	0,5	0,5	0,5	0,5	—	—	—	0,5	—
3	Zusatzprogramm										
3.1	Zusatzangaben zur beruflichen Ausbildung	Abs. 2, Nr. 1a	1	—	—	—	1	—	—	—	1
3.2	Pendlereigenschaft, -merkmale	Abs. 2, Nr. 1b	1	—	—	—	1	—	—	—	1
3.3	Zusatzangaben für Ausländer	Abs. 2, Nr. 2	1	—	—	—	1	—	—	—	1
3.4	Zusatzangaben zur Erwerbstätigkeit	Abs. 2, Nr. 3,4	1	—	—	—	1	—	—	—	1
3.5	Zusatzangaben zur Erwerbstätigkeit ³⁾ ...	Abs. 3, Nr. 2a	—	0,5	—	—	—	0,5	—	—	—
3.6	Private und betriebliche Altersvorsorge ³⁾⁴⁾	Abs. 3, Nr. 1,2b	—	0,5	—	—	—	0,5	—	—	—
3.7	Fragen zur Wohnsituation	Abs. 4	—	—	1	—	—	—	1	—	—
3.8	Angaben zur Krankenversicherung	Abs. 5, Nr. 1	—	—	—	1	—	—	—	1	—
3.9	Angaben zur Pflegeversicherung	Abs. 5, Nr. 1	—	—	—	1	—	—	—	1	—
3.10	Angaben zur Gesundheit ³⁾	Abs. 5, Nr. 2	—	—	—	0,5	—	—	—	0,5	—
3.11	Pflegebedürftigkeit, Leistungen einer Pflegeversicherung ³⁾	Abs. 5, Nr. 2	—	—	—	0,5	—	—	—	0,5	—

1) Die Angaben zum Eheschließungsjahr, zur Wohn- und Lebensgemeinschaft und zur Aufenthaltsdauer für Ausländer sind freiwillig.

2) Die Angaben zum allgemeinen und beruflichen Ausbildungsabschluß sind für Personen ab dem 51. Lebensjahr freiwillig.

3) In Anlehnung an die Genauigkeitsanforderungen für die Arbeitskräfteerhebung ist der Auswahlsatz des Ergänzungsprogramms sowie der Zusatzprogramme 3.5, 3.6, 3.10 und 3.11 auf Regierungsbezirksebene unterschiedlich (0,4%, 0,6%, 0,8% oder 1%). Im Bundesdurchschnitt kann er 0,5% betragen.

4) Die Angabe zur privaten Altersvorsorge ist freiwillig.

Quelle: Emmerling und Riede 1997: 167 (Übersicht 3).

Da im Rahmen der Unterstichprobe auch die Datenanforderungen der EU-Arbeitskräfteerhebung erfüllt werden, orientiert sich der Auswahlsatz für die Unterstichprobe an den Genauigkeitsanforderungen für die EU-Arbeitskräfteerhebung¹. Aufgrund dieser Genauigkeitsanforderungen kann der Auswahlsatz der Unterstichprobe auf der Ebene der Regierungsbezirke 0,4, 0,6, 0,8 oder 1 Prozent betragen. Im Bundesdurchschnitt ergibt sich für die Erhebungsjahre 1996 bis 2004 eine disproportionale Unterstichprobe mit einem Auswahlsatz von etwa 0,45 Prozent.

Da die Unterstichprobe so angelegt ist, dass einerseits die Belastung der Befragten gleichmäßig auf die Auswahleinheiten verteilt ist und andererseits eine Entlastung der zu befragenden Personen in der ersten und vierten Befragung erreicht werden soll, sind von Erhebungsjahr zu Erhebungsjahr wechselnde Rotationsteile für das Unterstichprobenprogramm vorgesehen. Damit ergeben sich unmittelbare Auswirkungen auf die Analysemöglichkeiten innerhalb eines Rotationsviertels, da der Auswahlsatz der an der Unterstichprobe beteiligten Personen innerhalb eines Rotationsviertels von Erhebungsjahr zu Erhebungsjahr variiert.

Wie nachfolgender Tabelle 1.2 – in der exemplarisch die Anteilwerte der an der Unterstichprobe beteiligten Personen in den einzelnen Bundesländern an dem dritten Rotationsviertel der fünften Stichprobe (RV 05/3) ausgewiesen sind – entnommen werden kann, sind im ersten Erhebungsjahr eines neu in die Auswahl gelangten Rotationsviertels rund 7,5 Prozent der zu befragenden Personen auch in das Unterstichprobenprogramm involviert. In den beiden darauf folgenden Erhebungsjahren erhöht sich dieser Anteil auf etwa 84 Prozent. Im vierten und letzten Erhebungsjahr werden wiederum die selben Regierungsbezirke wie im ersten Erhebungsjahr des dritten Rotationsviertels der fünften Stichprobe für die Unterstichprobe herangezogen, sodass sich wiederum ein Auswahlsatz von etwa 7,5 Prozent der befragten Personen ergibt, die in das Unterstichprobenprogramm eingebunden sind. Aus diesem Rotationsschema des Unterstichprobenprogramms innerhalb eines Rotationsviertels und der unterschiedlichen Größe der Regierungsbezirke in den einzelnen Bundesländern folgt des Weiteren, dass die Auswahlsätze des Unterstichprobenprogramms zwischen den Bundesländern unterschiedlich ausfallen.

¹Gemäß der Verordnung (Nr. 577/98 vom 9. März 1998) zur Durchführung einer Stichprobenerhebung über Arbeitskräfte in der Europäischen Gemeinschaft darf nach Art. 3, Abs. 1 der einfache relative Stichprobenfehler für eine Gruppe von Arbeitslosen, die 5 Prozent der Bevölkerung im erwerbsfähigen Alter ausmacht auf der Regionaleinheit NUTS II (entspricht in Deutschland der Ebene der Regierungsbezirke) höchstens 8 Prozent der betreffenden Bevölkerungsgruppe betragen. Ausgenommen hiervon sind Regionen mit weniger als 300 000 Einwohnern.

Im Falle einer jährlichen Erhebung im Frühjahr, wie dies auf den Mikrozensus zutrifft, wird mindestens ein Viertel der Erhebungseinheiten der Stichprobe der vorhergehenden Erhebung entnommen und mindestens ein Viertel in die Stichprobe der nächsten Erhebung einbezogen (Art. 3, Abs. 3 der Verordnung Nr. 577/98).

Tabelle 1.2: Anzahl der zu befragenden Personen und Anteil der in das Unterstichprobenprogramm einbezogenen Personen im dritten Rotationsviertel der fünften Stichprobe in den einzelnen Erhebungsjahren

Bundesland	Befragte Personen im zweiten Rotationsviertel der vierten Stichprobe im Erhebungsjahr . . .							
	2001		2002		2003		2004	
	Insgesamt	darunter %-Anteil in UStp.	Insgesamt	darunter %-Anteil in UStp.	Insgesamt	darunter %-Anteil in UStp.	Insgesamt	darunter %-Anteil in UStp.
SH	6 366	—	6 309	79,5	6 441	79,4	6 311	—
HH	3 718	—	3 700	81,0	3 706	79,3	3 786	—
NI	16 847	4,0	17 031	84,8	17 085	83,7	16 596	4,3
HB	1 440	100,0	1 471	100,0	1 456	100,0	1 428	100,0
NW	37 921	—	37 263	79,7	37 432	79,8	36 857	—
HE	13 024	14,3	13 370	87,3	13 463	87,4	13 394	14,5
RP	9 182	21,2	9 226	90,5	9 069	89,3	8 820	20,5
BW	22 674	—	22 489	80,6	22 613	79,7	22 749	—
BY	28 758	11,2	28 836	87,9	28 889	88,2	28 856	11,4
SL	2 249	18,6	2 201	100,0	2 219	100,0	2 144	20,0
BE	7 137	7,4	7 157	86,8	7 254	89,1	7 245	5,7
BB	5 776	—	5 761	80,6	5 630	80,7	5 681	—
MV	3 359	—	3 392	78,6	3 443	79,8	3 490	—
SN	9 712	4,7	9 661	84,2	9 723	86,3	9 598	4,9
ST	5 665	50,2	5 704	100,0	5 810	100,0	5 659	51,7
TH	5 270	—	5 470	81,3	5 339	80,0	5 275	—
D	179 095	7,5	179 041	84,4	179 572	84,3	177 889	7,5

UStp: Unterstichprobe

SH=Schleswig-Holstein, HH=Hamburg, NI=Niedersachsen, HB=Bremen, NW=Nordrhein-Westfalen, HE=Hessen, RP=Rheinland-Pfalz, BW=Baden-Württemberg, BY=Bayern, SL=Saarland, BE=Berlin, BB=Brandenburg, MV=Mecklenburg-Vorpommern, SN=Sachsen, ST=Sachsen-Anhalt, TH=Thüringen, D=Deutschland

1.4 Erhebungs- und Analyseeinheiten im Mikrozensus

Infolge der Totalerhebung aller wohnberechtigten Personen innerhalb eines, in den Mikrozensus einbezogenen, Auswahlbezirks und der Vergabe von Ordnungsnummern zur technisch-organisatorischen Durchführung, weist der Mikrozensus eine hierarchische Datenstruktur auf, die von den Erhebungseinheiten der Auswahlbezirke über die potenziellen Analyseeinheiten der Wohnungen und Haushalte bis hin zur Person reicht. Des Weiteren können Analysen auf der Ebene der Familien und Lebensgemeinschaften vorgenommen werden. Die beiden letzten Analyseeinheiten

sind jedoch nicht Teil der Erhebung oder Durchführung des Mikrozensus, sondern werden erst im Rahmen der Datenaufbereitung abgegrenzt.

Alle einer Person übergeordneten Analyseeinheiten können mit Identifikatoren eindeutig bestimmt werden. Für die Analyseeinheiten der Haushalte, Familien und Lebensgemeinschaften werden bereits einige Typisierungen bereitgestellt, die sich auf die Bezugsperson bzw. auf den/die Partner/in der Bezugsperson des Haushalts, der Familie oder Lebensgemeinschaft beziehen. Darüber hinaus werden eine Reihe unterschiedlicher Haushalts- und Familientypisierungsmerkmale (Generationen, verwandtschaftliche Beziehungen etc.) gebildet. Zu beachten ist bei diesen übergeordneten Kontextmerkmalen, dass das Koresidenzprinzip gilt, d.h. es werden nur die Beziehungen zu den Personen erhoben und abgebildet, die in einem gemeinsamen Haushalt leben.

1.5 Weiterführende Literatur

- Emmerling, D.; Riede, T. (1997): *40 Jahre Mikrozensus*, in: *Wirtschaft und Statistik*; 3/1997: 160-174.
- Heidenreich, H.-J. (1994): *Hochrechnung des Mikrozensus ab 1990*; S. 112-123, in: Gabler, S.; Hoffmeyer-Zlotnik, J.; Krebs, D. (Hrsg.): *Gewichtung in der Umfragepraxis*. Opladen: Westdeutscher Verlag.
- Heidenreich, H.-J.; Nöthen, M. (2002): *Der Wandel der Lebensformen im Spiegel des Mikrozensus*, in: *Wirtschaft und Statistik*; 1/2002: 26-38.
- Krug, W.; Nourney, M.; Schmidt, J. (1999): *Wirtschafts- und Sozialstatistik. Gewinnung von Daten*. München, Wien: R. Oldenbourg, 5., völlig neubearbeitete Auflage.
- Lengerer, A.; Bohr, J.; Janßen, A. (2005): *Haushalte, Familien und Lebensformen im Mikrozensus – Konzepte und Typisierungen*, ZUMA-Arbeitsbericht Nr. 2005/05. Mannheim: Zentrum für Umfragen, Methoden und Analysen. URL: http://www.gesis.org/Publikationen/Berichte/ZUMA_Arbeitsberichte/05/AB_05_05.pdf (letzter Zugriff: 23. März 2008)
- Meyer, K. (1994): *Zum Auswahlplan des Mikrozensus ab 1990*; S. 106-111, in: Gabler, S.; Hoffmeyer-Zlotnik, J.; Krebs, D. (Hrsg.): *Gewichtung in der Umfragepraxis*. Opladen: Westdeutscher Verlag.
- Rönsch, H. (1990): *Regionalisierung des Mikrozensus. Basis Volkszählung 1987*, in: *Statistische Rundschau für das Land Nordrhein-Westfalen*; **42**: 24-27.
- Schimpl-Neimanns, B. (2006): *Filekonzept zum Mikrozensuspanel*. Arbeitspapier #12, Methodenverbund "Aufbereitung und Bereitstellung des Mikrozensus als Panelstichprobe". URL: <http://www.destatis.de/jetspeed/portal/cms/Sites/>

destatis/Internet/DE/Content/Wissenschaftsforum/MethodenVerfahren/Mikrozensus/Arbeitspapiere/Arbeitspapier12,property=file.pdf (letzter Zugriff: 23. März 2008)

Statistisches Bundesamt in Zusammenarbeit mit den Statistischen Landesämtern (Hrsg.) (ohne Jahr): *Mikrozensus 1996. Interviewerhandbuch, Teil 3: Begehungsanleitung.*

Kapitel 2

Der Mikrozensus als Datenquelle für Panelauswertungen

Robert Herter-Eschweiler (Statistisches Bundesamt, Gruppe VIII C)

2.1 Vorbemerkung

Analog zum vorherigen Kapitel, in dem eine allgemeine Beschreibung des Mikrozensus gegeben wurde, wird hier erläutert, wie sich der Mikrozensus als Datenquelle für Längsschnittauswertungen nutzen lässt. Die in diesem Kapitel angegebenen Fallzahlen beziehen sich daher stets auf das Gesamtfile des Mikrozensus-Panels 2001–2004 und nicht auf das faktisch anonymisierte Mikrozensus-Panelfile 2001–2004.

Wesentlich für die Nutzung des Mikrozensus für Panelauswertungen ist das Konzept der partiellen Rotation der Auswahlbezirke (jährlich wird ein Viertel der Auswahlbezirke ausgetauscht), sowie der Tatsache, dass alle meldepflichtigen Personen eines ausgewählten Auswahlbezirks zu befragen sind.

Die partielle Rotation der Auswahlbezirke wurde bereits im Oktober 1962 zur Kostenreduktion und Verbesserung der Ergebnisgenauigkeit eingeführt. Damit war zwar prinzipiell die Voraussetzung zur Nutzung des Mikrozensus als Panel frühzeitig gegeben, sie wurde jedoch wenig genutzt¹. Ab 1982 waren Panelauswertungen infolge der Vorgaben für das Trennen und Löschen der Ordnungsnummern zum Zusammenführen der Stichprobeneinheiten des Mikrozensus über die unterschiedlichen Erhebungsjahre nicht mehr realisierbar. Dies ist erst wieder seit dem Mikrozensusgesetz vom 17. Januar 1996 (BGBl. I S. 34.) möglich.

Infolge der partiellen Rotation der Auswahlbezirke können zum einen Zweijahrespanels auf der Basis von jeweils drei Rotationsvierteln – also einer Auswahl von

¹Zu den vorliegenden Paneluntersuchungen aus dieser frühen Zeit siehe Linke (1969) sowie Mayer (1979, 1983).

75 Prozent der Auswahlbezirke – gebildet werden, wobei sich die aufeinander folgenden Zweijahreslängsschnitte in je zwei von drei Rotationsvierteln überlappen. Weiterhin können Dreijahreslängsschnitte auf der Basis von jeweils zwei Rotationsvierteln – also einer Auswahl von 50 Prozent der Auswahlbezirke – gebildet werden. In den aufeinander folgenden Dreijahreslängsschnitten überlappt sich dabei jeweils eines von zwei Rotationsvierteln. Schließlich können Vierjahreslängsschnitte auf der Basis eines Rotationsviertels (Auswahlsatz: 25 Prozent der Auswahlbezirke) gebildet werden, bei denen sich keine Rotationsviertel mehr überlappen. In der nachfolgenden Abbildung 2.1 sei dies beispielhaft an je zwei aufeinander folgenden Paneldatensätzen mit unterschiedlicher Befragungsdauer dargestellt.

Das Zweijahreslängsschnitt 1996/1997 setzt sich zusammen aus dem vierten Rotationsviertel der dritten Stichprobe sowie dem ersten und zweiten Rotationsviertel der vierten Stichprobe und das Zwei-Wellen-Panel 1997/1998 aus dem ersten, zweiten und dritten Rotationsviertel der vierten Stichprobe.

Damit, bezogen auf das Erhebungsjahr 1997, überlappen sich die beiden Zweijahreslängsschnitte mit dem ersten und zweiten Rotationsviertel der vierten Stichprobe (hellgraue Kästchen). Analog hierzu sind die Panels mit einer Befragungsdauer von drei bzw. vier Erhebungszeitpunkten zu interpretieren, wobei deutlich zu erkennen ist, dass es bei einem Vierjahreslängsschnitt keine Überlappungen mehr gibt.

Längsschnittdatensätze auf der Basis des Mikrozensus decken damit zwar einen relativ kurzen Beobachtungszeitraum ab, ermöglichen aber aufgrund ihres großen Stichprobenumfangs inhaltlich differenzierte Analysen für ein breites Merkmalspektrum.

Der maximale Anteil identischer Auswahlbezirke von 75 Prozent, 50 Prozent bzw. 25 Prozent wird jedoch nicht auf der Haushalts- und Personenebene erreicht, weil infolge der Anlage des Mikrozensus als Wohnungsstichprobe aus dem Auswahlbezirk wegziehende Haushalte und Personen nicht weiter befragt werden, sondern durch die nachziehenden Bewohner/innen ersetzt werden. Dies führt einerseits zu methodischen Komplikationen in Verbindung mit räumlicher Mobilität: für fortziehende Einheiten fehlen die Angaben nach dem Fortzug und für zuziehende Einheiten liegen – bis auf wenige Retrospektivangaben – keine Informationen vor dem Zuzug vor. Andererseits bietet dieses Verfahren der Ersetzung fortziehender durch zuziehende Einheiten den Vorteil der automatischen Erfassung von Populationsveränderungen.

Abbildung 2.1: Zusammenführbare Rotationsviertel der Mikrozensus-erhebungen 1996 bis 2004

Stp.Nr./ RV	Erhebungsjahr									
	1996	1997	1998	1999	2000	2001	2002	2003	2004	
	a	b	c	d	e	f	g	h	i	
Zweijahreslängsschnitte										
03/3	4. Befr.									
03/4	3. Befr.	4. Befr.								
04/1	2. Befr.	3. Befr.	4. Befr.							
04/2	1. Befr.	2. Befr.	3. Befr.	4. Befr.						
04/3		1. Befr.	2. Befr.	3. Befr.	4. Befr.					
04/4			1. Befr.	2. Befr.	3. Befr.	4. Befr.				
05/1				1. Befr.	2. Befr.	3. Befr.	4. Befr.			
05/2					1. Befr.	2. Befr.	3. Befr.	4. Befr.		
05/3						1. Befr.	2. Befr.	3. Befr.	4. Befr.	
05/4							1. Befr.	2. Befr.	3. Befr.	
06/1								1. Befr.	2. Befr.	
06/2									1. Befr.	
Dreijahreslängsschnitte										
03/3	4. Befr.									
03/4	3. Befr.	4. Befr.								
04/1	2. Befr.	3. Befr.	4. Befr.							
04/2	1. Befr.	2. Befr.	3. Befr.	4. Befr.						
04/3		1. Befr.	2. Befr.	3. Befr.	4. Befr.					
04/4			1. Befr.	2. Befr.	3. Befr.	4. Befr.				
05/1				1. Befr.	2. Befr.	3. Befr.	4. Befr.			
05/2					1. Befr.	2. Befr.	3. Befr.	4. Befr.		
05/3						1. Befr.	2. Befr.	3. Befr.	4. Befr.	
05/4							1. Befr.	2. Befr.	3. Befr.	
06/1								1. Befr.	2. Befr.	
06/2									1. Befr.	
Vierjahreslängsschnitte										
03/3	4. Befr.									
03/4	3. Befr.	4. Befr.								
04/1	2. Befr.	3. Befr.	4. Befr.							
04/2	1. Befr.	2. Befr.	3. Befr.	4. Befr.						
04/3		1. Befr.	2. Befr.	3. Befr.	4. Befr.					
04/4			1. Befr.	2. Befr.	3. Befr.	4. Befr.				
05/1				1. Befr.	2. Befr.	3. Befr.	4. Befr.			
05/2					1. Befr.	2. Befr.	3. Befr.	4. Befr.		
05/3						1. Befr.	2. Befr.	3. Befr.	4. Befr.	
05/4							1. Befr.	2. Befr.	3. Befr.	
06/1								1. Befr.	2. Befr.	
06/2									1. Befr.	

Stp.Nr.: Stichprobennummer; RV: Rotationsviertel; Befr.: Befragung

2.2 Die Bildung des Mikrozensus-Panels 2001–2004

Zur organisatorischen Vorbereitung und Durchführung der Erhebung verwendet der Mikrozensus ein Ordnungsnummernsystem, in dem die Zugehörigkeit zu einem Auswahlbezirk, zu einer Wohnung und zu einem Haushalt festgehalten wird. Mit den Vorgaben für das Trennen und Löschen der Ordnungsnummern des ab 1996 geltenden Mikrozensusgesetzes ist es möglich, an die befragten Personen und Haushalte

über die vier Erhebungszeitpunkte hinweg identische Ordnungsnummern zu vergeben und über diese dann die Querschnittsdatensätze mit den identischen Einheiten zusammenzuführen.

Mit der Orientierung des Mikrozensus auf Querschnittsergebnisse kann zurzeit ohne zusätzlichen Aufwand keine vollständige Längsschnittkonsistenz der Ordnungsnummern gewährleistet werden. Untersuchungen zeigen insbesondere die Schwachstellen auf der Personenebene auf. Von mehreren untersuchten Alternativen für eine rein maschinelle Zusammenführung – zentral im Statistischen Bundesamt – mit möglichst hoher Zuverlässigkeit erscheint der Identifikator am besten geeignet, der sich aus der Ordnungsnummer für den Auswahlbezirk und den Haushalt einerseits und dem Geburtsjahr sowie dem Geschlecht der Person innerhalb eines Haushalts andererseits zusammensetzt.

Ausgehend von den Erfahrungen der ersten Erstellung eines Mikrozensus-Panels konnte als ein weiterer Aspekt, der zu einer einschränkenden Verknüpfung der Querschnittsdatensätze führt, die Datenqualität identifiziert werden. Die Inkonsistenzen hinsichtlich der Angaben zum Geburtsjahr, dem Geschlecht oder den Angaben zu einer haushalts- bzw. personenbezogenen Veränderung im Zeitablauf (gemeint sind hiermit Fort- und Zuzüge von Haushalten oder Personen sowie der Tod oder die Geburt einer Person) sollten ebenfalls in einem standardisierten Verfahren berücksichtigt werden. Für die Erstellung des Mikrozensus-Panels 2001–2004 wurde daher ein sequentielles Verfahren entwickelt, in dem sukzessive die Bedingungen, nach denen Personensätze als plausibel zusammengeführt gelten, gelockert werden. So werden Abweichungen von ± 1 Jahr beim Geburtsjahr zugelassen als auch die Auslassung der Angabe zum Geschlecht als Identifikator-Merkmal. Zudem müssen unvollständige Verläufe nicht immer durch Kennungen über Verluste und Gewinne auf der Haushaltsebene "erklärt" werden können und es wird als zusätzlicher Identifikator zur Zusammenführung von Personensätzen das Merkmal "Eheschließungsjahr" aufgenommen. Das sequentielle Verfahren, nach dem die Erstellung des Mikrozensus-Panels 2001–2004 erfolgte, wird nachfolgend kurz beschrieben.

Aufgrund räumlicher Mobilität und den demographischen Ereignissen Geburt und Tod liegen bei der Bildung eines Mikrozensus-Panels über vier Erhebungszeitpunkte nicht zu allen Erhebungszeitpunkte Informationen vor. Wird pro Erhebungszeitpunkt notiert, ob für einen Personensatz eine Information vorliegt oder nicht, ergeben sich insgesamt 16 Kombinationsmöglichkeiten, die in Abhängigkeit von Informationen über haushalts- und/oder personenbezogenen Veränderungsangaben im Zeitablauf als plausibel bzw. unplausibel eingestuft werden können. Eingeteilt werden die Veränderungsangaben in Verluste (Fortzüge von Haushalten oder Personen sowie Sterbefälle) und Gewinne (Zuzüge von Haushalten oder Personen sowie Geburten). Zu der Einstufung eines Zusammenführungstyps in Abhängigkeit von Verlusten und Gewinnen siehe Tabelle 2.2.

Tabelle 2.1: Kombinationstypen bezüglich des Vorliegens bzw. Nicht-Vorliegens von Informationen in einem Mikrozensus-Panel über vier Erhebungszeitpunkte

Kombinationsnummer	Es liegen Angaben vor für das Erhebungsjahr			
	t_1	t_2	t_3	t_4
1	Ja	Nein	Nein	Nein
2	Ja	Ja	Nein	Nein
3	Ja	Ja	Ja	Nein
4	Ja	Ja	Ja	Ja
5	Nein	Ja	Nein	Nein
6	Nein	Ja	Ja	Nein
7	Nein	Ja	Ja	Ja
8	Nein	Nein	Ja	Nein
9	Nein	Nein	Ja	Ja
10	Nein	Nein	Nein	Ja
11	Ja	Nein	Ja	Nein
12	Ja	Nein	Ja	Ja
13	Ja	Nein	Nein	Ja
14	Ja	Ja	Nein	Ja
15	Nein	Ja	Nein	Ja
0	Nein	Nein	Nein	Nein

Tabelle 2.2: Konstellation von Kombinationstypen und Kennung von Verlusten und Gewinnen, die zu einem plausibel zusammengeführten Personendatensatz führen

Kombinationstyp	Bedingung(en) bezüglich Verlusten/Gewinnen die gelten müssen, damit ein Datensatz mit nebenstehendem Kombinationstyp als plausibel eingestuft werden kann
1: x - - -	Verlust in 2002 oder Ausfall in 2002, 2003 und 2004 oder Ausfall in 2002 und Verlust in 2003 oder Ausfall in 2002 und 2003 und Verlust in 2004
2: x x - -	Verlust in 2003 oder Ausfall in 2003 und 2004 oder Ausfall in 2003 und Verlust in 2004
3: x x x -	Verlust in 2004 oder Ausfall in 2004
4: x x x x	keine Bedingung(en)
5: - x - -	Gewinn in 2002 und Verlust in 2003 oder Gewinn in 2002 und Ausfall in 2003 und 2004 oder Gewinn in 2002 und Ausfall in 2003 und Verlust in 2004 oder Ausfall in 2001 und Verlust in 2003 oder Ausfall in 2001, 2003 und 2004 oder Ausfall in 2001 und 2003 und Verlust in 2004
6: - x x -	Gewinn in 2002 und Verlust in 2004 oder Gewinn in 2002 und Ausfall in 2004 oder Ausfall in 2001 und Verlust in 2004 oder Ausfall in 2001 und 2004

Fortsetzung der Tabelle auf nächster Seite

Kombinationstyp	Bedingung(en) bezüglich Verlusten/Gewinnen die gelten müssen, damit ein Datensatz mit nebenstehendem Kombinationstyp als plausibel eingestuft werden kann
7: - x x x	Gewinn in 2002 oder Ausfall in 2001
8: - - x -	Gewinn in 2003 und Verlust in 2004 oder Gewinn in 2003 und Ausfall in 2004 oder Ausfall in 2001 und 2002 und Verlust in 2004 oder Ausfall in 2001, 2002 und 2004
9: - - x x	Gewinn in 2003 oder Ausfall in 2001 und 2002
10: - - - x	Gewinn in 2004 oder Ausfall in 2001, 2002 und 2003
11: x - x -	Ausfall in 2002 und Verlust in 2004 oder Ausfall in 2002 und 2004
12: x - x x	Ausfall in 2002
13: x - - x	Ausfall in 2002 und 2003
14: x x - x	Ausfall in 2003
15: - x - x	Gewinn in 2002 und Ausfall in 2003 oder Ausfall in 2001 und 2003

Die nummern entsprechen dem Kombinationstyp in Tabelle 2.1

x: Es liegen Angaben für ein Erhebungsjahr vor

-: Es liegen keine Angaben für ein Erhebungsjahr vor

Nach diesen Kriterien wird auf jeder Zusammenführungsstufe ein Datensatz als plausibel oder unplausibel eingestuft. Nach welchem Verfahren die Personensätze auf den verschiedenen Stufen zusammengeführt werden, ist nachfolgend kurz beschrieben.

Zusammenführungsstufe 1: Ein Datensatz gilt als plausibel zusammengeführt wenn:

- a) der Zusammenführungstyp auf der Personen- und Haushaltsebene übereinstimmt,

- b) unvollständige Verläufe oder abweichende Zusammenführungstypen zwischen Personen- und Haushaltsebene über die Kennung von Verlusten, Gewinnen oder Ausfällen "erklärt" werden können und
- c) die Anzahl plausibler Personensätze im Zeitablauf übereinstimmt mit der Anzahl der Personensätze in einem Haushalt, für die Angaben vorliegen.

Die Zusammenführung der Personensätze erfolgt über die Identifikatoren: Bundesland, Regierungsbezirk, Auswahlbezirk, Haushaltsnummer, Geburtsjahr und Geschlecht.

Zusammenführungsstufe 2: Alle nach Zusammenführungsstufe 1 nicht zusammengeführten Datensätze gelten als plausibel zusammengeführt, wenn:

- a) der Zusammenführungstyp auf der Personen- und Haushaltsebene übereinstimmt,
- b) unvollständige Verläufe oder abweichende Zusammenführungstypen zwischen Personen- und Haushaltsebene über die Kennung von Verlusten, Gewinnen oder Ausfällen "erklärt" werden können und
- c) die Anzahl plausibler Personensätze im Zeitablauf übereinstimmt mit der Anzahl der Personensätze in einem Haushalt, für die Angaben vorliegen.

Die Zusammenführung der Personensätze erfolgt nur in Privathaushalten mit zwei oder mehr Personen und wird in zwei Schritten vollzogen:

Schritt 1: Führe die Personensätze zusammen über die Identifikatoren: Bundesland, Regierungsbezirk, Auswahlbezirk, Haushaltsnummer, Geburtsjahr und Geschlecht.

Schritt 2: Führe die nach Schritt 1 noch nicht zusammengeführten Personensätze zusammen über die Identifikatoren: Bundesland, Regierungsbezirk, Auswahlbezirk, Haushaltsnummer, Eheschließungsjahr und Geschlecht.

Zusammenführungsstufe 3: Alle nach Zusammenführungsstufe 2 nicht zusammengeführten Datensätze gelten als plausibel zusammengeführt, wenn:

- a) der Zusammenführungstyp auf der Personen- und Haushaltsebene
 - ohne Lücken (Kombinationstypen 1 bis 10) übereinstimmt bzw.
 - mit Lücken (Kombinationstypen 11 bis 15) durch Ausfallkennungen "erklärt" werden kann,
- b) die Anzahl plausibler Personensätze im Zeitablauf übereinstimmt mit der Anzahl der Personensätze in einem Haushalt, für die Angaben vorliegen.

Die Zusammenführung der Personensätze erfolgt über die Identifikatoren: Bundesland, Regierungsbezirk, Auswahlbezirk, Haushaltsnummer, Geburtsjahr und Geschlecht.

Die Zusammenführungsstufe 3 unterscheidet sich von der Zusammenführungsstufe 1 dadurch, dass bei unvollständigen Verläufen auf der Haushaltsebene, diese nicht durch Verluste und Gewinne "erklärt" werden müssen. Es genügt, dass die Anzahl der als plausibel eingestuften Personensätze über alle Erhebungszeitpunkte mit der Anzahl der Personensätze in einem Haushalt übereinstimmt, für die Angaben vorliegen.

Zusammenführungsstufe 4: Alle nach Zusammenführungsstufe 3 nicht zusammengeführten Datensätze gelten als plausibel zusammengeführt, wenn:

- a) der Zusammenführungstyp auf der Personen- und Haushaltsebene übereinstimmt,
- b) unvollständige Verläufe oder abweichende Zusammenführungstypen zwischen Personen- und Haushaltsebene über die Kennung von Verlusten, Gewinnen oder Ausfällen "erklärt" werden können und
- c) die Anzahl plausibler Personensätze im Zeitablauf übereinstimmt mit der Anzahl der Personensätze in einem Haushalt, für die Angaben vorliegen.

Die Zusammenführung der Personensätze erfolgt nur in Privathaushalten mit zwei oder mehr Personen in Abhängigkeit bestimmter Haushalts- und Personentypisierungen (siehe Tabelle 2.3) und wird in zwei Schritten vollzogen:

Schritt 1: Führe die Personensätze zusammen über die Identifikatoren: Bundesland, Regierungsbezirk, Auswahlbezirk, Haushaltsnummer, Geburtsjahr und Geschlecht.

Schritt 2: Führe die nach Schritt 1 noch nicht zusammengeführten Personensätze zusammen über die Identifikatoren: Bundesland, Regierungsbezirk, Auswahlbezirk, Haushaltsnummer, Geburtsjahr ± 1 und Geschlecht.

Tabelle 2.3: Zugelassene Abweichungen von ± 1 Jahr im Geburtsjahr in Abhängigkeit "erklärter" Haushaltstypen

Erklärter Haushaltstyp	darin alternierender Personentyp	darin Personentyp, bei dem eine Abweichung von ± 1 Jahr im Geburtsjahr zugelassen wird
x x x -	x - x -	- x - -
x x x x	x - x x	- x - -
	x x - x	- - x -
	x - - x	- x x -
	x - x -	- x - x
- x x x	- x - x	- - x -

Zusammenführungsstufe 5: Alle nach Zusammenführungsstufe 4 nicht zusammengeführten Datensätze gelten als plausibel zusammengeführt, wenn:

- a) der Zusammenführungstyp auf der Personen- und Haushaltsebene übereinstimmt,
- b) unvollständige Verläufe oder abweichende Zusammenführungstypen zwischen Personen- und Haushaltsebene über die Kennung von Verlusten, Gewinnen oder Ausfällen "erklärt" werden können und
- c) die Anzahl plausibler Personensätze im Zeitablauf übereinstimmt mit der Anzahl der Personensätze in einem Haushalt, für die Angaben vorliegen.

Die Zusammenführung der Personensätze erfolgt nur in Privathaushalten mit zwei oder mehr Personen und wird in zwei Schritten vollzogen:

Schritt 1: Führe die Personensätze zusammen über die Identifikatoren: Bundesland, Regierungsbezirk, Auswahlbezirk, Haushaltsnummer, Geburtsjahr und Geschlecht.

Schritt 2: Führe die nach Schritt 1 noch nicht zusammengeführten Personensätze zusammen über die Identifikatoren: Bundesland, Regierungsbezirk, Auswahlbezirk, Haushaltsnummer und Geburtsjahr.

Zusammenführungsstufe 6: Alle nach Zusammenführungsstufe 5 nicht zusammengeführten Datensätze gelten als plausibel zusammengeführt, wenn:

- a) der Zusammenführungstyp auf der Personen- und Haushaltsebene übereinstimmt und
- b) unvollständige Verläufe oder abweichende Zusammenführungstypen zwischen Personen- und Haushaltsebene über die Kennung von Verlusten, Gewinnen oder Ausfällen "erklärt" werden können.

Die Zusammenführung der Personensätze erfolgt über die Identifikatoren: Bundesland, Regierungsbezirk, Auswahlbezirk, Haushaltsnummer, Geburtsjahr und Geschlecht.

Im Gegensatz zur Zusammenführungsstufe 1 wird hier nicht verlangt, dass die Anzahl der plausibel zusammengeführten Personensätze mit der Anzahl der Personensätze in einem Haushalt übereinstimmt, für die Angaben vorliegen. Die Plausibilität bezieht sich lediglich auf die Personenebene.

Zusammenführungsstufe 7: Alle nach Zusammenführungsstufe 6 nicht zusammengeführten Datensätze gelten als plausibel zusammengeführt, wenn:

- a) der Zusammenführungstyp auf der Personen- und Haushaltsebene
– ohne Lücken (Kombinationstypen 1 bis 10) übereinstimmt bzw.

- mit Lücken (Kombinationstypen 11 bis 15) durch Ausfallkennungen "erklärt" werden kann und
- b) abweichende Zusammenführungstypen zwischen Personen- und Haushaltsebene über die Kennung von Verlusten, Gewinnen oder Ausfällen "erklärt" werden können.

Die Zusammenführung der Personensätze erfolgt über die Identifikatoren: Bundesland, Regierungsbezirk, Auswahlbezirk, Haushaltsnummer, Geburtsjahr und Geschlecht.

Zusammenführungsstufe 8: Alle nach Zusammenführungsstufe 7 nicht zusammengeführten Datensätze gelten als plausibel zusammengeführt, wenn der Zusammenführungstyp auf der Personenebene über die Kennung von Verlusten und Gewinnen "erklärt" werden kann.

Die Zusammenführung der Personensätze erfolgt über die Identifikatoren: Bundesland, Regierungsbezirk, Auswahlbezirk, Haushaltsnummer, Geburtsjahr und Geschlecht.

2.3 Die Stichprobenentwicklung des Mikrozensus-Panels 2001–2004

Im vorhergehenden Unterkapitel wurde die Bildung des Mikrozensus-Panels 2001–2004 beschrieben. In welchem Umfang sich in den jeweiligen Phasen des Zusammenführungsprozesses Personensätze für den Mikrozensus-Paneldatensatz 2001–2004 zusammenführen lassen (Abschnitt 2.3.1) und welche Entwicklung sich im Zeitverlauf des Längsschnitts für die Personenebene (Abschnitt 2.3.2) bzw. für die Haushaltsebene (Abschnitt 2.3.2) ergibt, ist Gegenstand nachfolgender Ausführungen.

2.3.1 Umfang der Zusammenführungen im Mikrozensus-Panel 2001–2004 nach den Zusammenführungsstufen

Zunächst richtet sich der Blick auf die Auszählung im Zusammenhang mit den verschiedenen Zusammenführungsstufen (siehe Tabelle 2.4). Hier zeigt sich, dass nach Abschluss aller Zusammenführungsstufen fast 97 % der Datensätze unter Berücksichtigung von räumlicher Mobilität und den demographischen Ereignissen Geburt und Tod als längsschnittplausibel eingestuft werden können. Im gleichen Ausmaß können die Datensätze aus Privathaushalten als längsschnittkonsistent angesehen

werden. In Gemeinschaftsunterkünften variiert hingegen dieser Anteil zwischen 88 % und rund 90 %.

Privathaushalte und Gemeinschaftsunterkünfte unterscheiden sich aber nicht nur im Hinblick auf den Umfang plausibler Datensätze, sondern auch im Hinblick darauf, auf welcher Zusammenführungsebene die meisten Datensätze als plausibel eingestuft werden. In Privathaushalten ist dies auf der Zusammenführungsstufe 1, in der zwischen 75 % und 77 % aller plausiblen Datensätze zusammengeführt werden. An zweiter Stelle rangiert die Zusammenführungsstufe 3, in der immerhin noch zwischen 12 % und 14 % der Datensätze als plausibel zusammengeführt werden. Danach folgen die Zusammenführungsstufe 6 mit einem Anteil von rund 2.7 % an allen plausiblen zusammengeführten Datensätzen und die Zusammenführungsstufe 2 mit einem entsprechenden Anteil von etwa 2.4 %. Den restlichen Zusammenführungsebenen 4, 5, 7 und 8 kommt nur eine untergeordnete Rolle zu, da ihr Anteil an allen plausibel zusammengeführten Datensätzen zusammengenommen unter 3 % beträgt.

Anders gestaltet sich die Anteilsstruktur bei den Gemeinschaftsunterkünften. Mit einem Anteil zwischen 25 % und 38 % sind die meisten plausibel zusammengeführten Datensätze auf der achten Zusammenführungsstufe zu verzeichnen. Mit Anteilswerten zwischen 24 % und 26 % nimmt die Zusammenführungsstufe 6 den zweiten Rang ein. An dritter Position befindet sich die Zusammenführungsstufe 3 mit Anteilswerten zwischen 17 % und 26 %. Zusammenführungsanteile von 8 % bis 10 % ergeben sich auf der ersten Zusammenführungsstufe. Fast vernachlässigbar ist die Zusammenführungsstufe 7 mit einem Anteil von 0.4 % an allen plausiblen Datensätzen. Die Zusammenführungsstufen 2, 4 und 5 wurden in den Gemeinschaftsunterkünften nicht umgesetzt, sodass hier per Festlegung auch keine Zusammenführungen vorliegen können.

Tabelle 2.4: (Un-)Plausible Datensätze im Mikrozensus-Panel 2001–2004, differenziert nach Erhebungsjahr, Haushaltstyp und Zusammenführungsebene

Personensätze	Angaben beziehen sich auf das Erhebungsjahr											
	2001			2002			2003			2004		
	Insgesamt	davon in PHH	GU	Insgesamt	davon in PHH	GU	Insgesamt	davon in PHH	GU	Insgesamt	davon in PHH	GU
Insgesamt	179 095	177 314	1 781	179 041	177 322	1 719	179 570	177 891	1 679	177 889	176 124	1 765
davon plausibel zusammengeführt	173 414	171 852	1 562	173 410	171 871	1 539	173 959	172 436	1 523	172 447	170 879	1 568
Insgesamt	136 895	136 707	188	133 062	132 889	173	133 641	133 491	150	134 752	134 610	142
davon auf der Zusammenführungsebene	4 180	4 180	—	4 374	4 374	—	4 395	4 395	—	4 246	4 246	—
1	21 693	21 216	477	25 028	24 664	364	24 879	24 591	288	22 565	22 252	313
2	2 076	2 076	—	2 176	2 176	—	2 176	2 176	—	2 110	2 110	—
3	2 082	2 082	—	2 227	2 227	—	2 226	2 226	—	2 182	2 182	—
4	5 258	4 825	433	5 207	4 770	437	5 228	4 791	437	5 128	4 691	437
5	109	103	6	34	30	4	135	126	9	148	121	27
6	1 121	663	458	1 302	741	561	1 279	640	639	1 316	667	649
7	5 681	5 462	219	5 631	5 451	180	5 611	5 455	156	5 442	5 245	197
Restposition												
Insgesamt	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0
davon plausibel zusammengeführt	96,8	96,9	87,7	96,9	96,9	89,5	96,8	96,9	90,7	96,9	97,0	90,6
Insgesamt	76,4	77,1	10,6	74,3	74,9	10,1	74,4	75,0	8,9	75,8	76,4	8,1
davon auf der Zusammenführungsebene	2,3	2,4	—	2,4	2,5	—	2,5	2,5	—	2,4	2,4	—
1	12,1	12,0	26,8	14,0	13,9	21,2	13,9	13,8	17,2	12,7	12,6	17,7
2	1,2	1,2	—	1,2	1,2	—	1,2	1,2	—	1,2	1,2	—
3	1,2	1,2	—	1,2	1,3	—	1,2	1,3	—	1,2	1,2	—
4	2,9	2,7	24,3	2,9	2,7	25,4	2,9	2,7	26,0	2,9	2,7	24,8
5	0,1	0,1	0,3	0,0	0,0	0,2	0,1	0,1	0,5	0,1	0,1	1,5
6	0,6	0,4	25,7	0,7	0,4	32,6	0,7	0,4	38,1	0,7	0,4	36,8
7	3,2	3,1	12,3	3,1	3,1	10,5	3,1	3,1	9,3	3,1	3,0	11,2
8												
Restposition												

PHH = Privathaushalt; GU = Gemeinschaftsunterkunft; — = Wert per Definition ausgeschlossen

2.3.2 Die Stichprobenentwicklung im Mikrozensus-Panel 2001–2004 auf der Personenebene

Bezüglich der Stichprobenentwicklung im Mikrozensus-Panel 2001–2004 ist zu beachten, dass bei den als plausibel ausgewiesenen Angaben sowohl personelle Veränderungen (Zuzüge, Geburten, Fortzüge und Sterbefälle) als auch Veränderungen berücksichtigt wurden, die den gesamten Haushalt betreffen (Zu-/Fortzug bzw. Tod des Haushalts).

Weiterhin ist zu beachten, dass bei vier Erhebungszeitpunkten mit der Kennung, ob für den jeweiligen Erhebungszeitpunkt personelle Angaben vorliegen oder nicht, sich insgesamt 16 Kombinationsmöglichkeiten ergeben. Die Kombinationsmöglichkeit, nach der zu keinem Erhebungszeitpunkt personelle Angaben vorliegen, scheidet aus und Kombinationsmöglichkeiten, bei denen Angaben zu zwei oder drei, von temporären Ausfällen unterbrochenen Erhebungszeitpunkten vorliegen, werden zu einer Kategorie zusammengefasst. Bei diesen temporären Ausfällen ist zu beachten, dass es sich grundsätzlich um Ausfälle aller Haushaltsmitglieder handelt, weil der zu befragende Haushalt im betreffenden Erhebungsjahr nicht erreicht wurde. Auf der Personenebene selbst kann es keine temporären Ausfälle geben, da für alle Mitglieder eines Haushalts Auskunftspflicht besteht. Für abwesende Haushaltsmitglieder (wegen Urlaub, Krankenhausaufenthalt, Montage u.ä.), die im zu befragenden Haushalt als wohnberechtigt gemeldet sind, werden die Angaben von einem volljährigen, auskunftspflichtigen Haushaltsmitglied gemacht.

Unter diesen Setzungen können von den jeweils rund 179 000 Datensätzen pro Erhebungsjahr knapp 97 % als plausibel eingestuft werden (siehe Tabelle 2.5). Vollständig zusammengeführt, d.h. es liegen Angaben über alle vier Erhebungszeitpunkte vor, konnten rund 67 % der Datensätze. Für ca. 28,5 % liegen aufgrund räumlicher Mobilität oder den demographischen Ereignissen Geburt und Tod nur unvollständige Verlaufsangaben vor. Der Anteil von Datensätzen mit temporären Ausfällen beziffert sich auf durchschnittlich 0,75 %.

Ausgehend von rund 177 250 Datensätzen pro Erhebungsjahr von Personen aus Privathaushalten, zeigt sich eine nahezu gleiche Anteilstruktur hinsichtlich der als plausibel definierten Datensätzen. Knapp 97 % der Datensätze können als plausibel zusammengeführt betrachtet werden. Diese 97 % unterteilen sich in etwa 67 % immobile Datensätze, d.h. für diese liegen Angaben zu allen vier Erhebungszeitpunkten vor, 28,5 % der Datensätze weisen eine räumliche Mobilität bzw. das demographische Ereignis Geburt oder Tod auf und bei ca. 0,75 % der Datensätze ist ein temporärer Ausfall zu verzeichnen.

Anders gestaltet sich die Anteilstruktur der plausibel zusammengeführten Datensätzen bei Personen aus Gemeinschaftsunterkünften. Von den durchschnittlich 1 750 Datensätzen von Personen aus Gemeinschaftsunterkünften können durchschnittlich etwa 89 % plausibel zusammengeführt werden, wovon aber lediglich für durchschnittlich 38 % Angaben zu allen vier Erhebungszeitpunkten vorliegen. Unvollständige Verlaufsangaben liegen bei ca. 51 % der Datensätze vor. Bei lediglich

0,3 % ist ein temporärer Ausfall zu verzeichnen.

Bei den Datensätzen von Personen aus Gemeinschaftsunterkünften ist nicht nur eine andere Anteilstruktur in der Gestalt einer niedrigeren Immobilität und dementsprechend höheren Gewinn- und Verlustanteilen festzustellen, es zeigen sich auch größere Variationen in den einzelnen Anteilwerten zwischen den einzelnen Erhebungsjahren. So beziffert sich die größte Differenz bei den Immobilien von Personen aus Privathaushalten auf 0,7 Prozentpunkte wohingegen der entsprechende Differenzwert von Personen aus Gemeinschaftsunterkünften 2,2 Prozentpunkte beträgt.

Tabelle 2.5: Umfang (un-)plausibler Datensätze auf der Personenebene im Mikrozensus-Panel 2001–2004, differenziert nach Haushaltstypen in den einzelnen Erhebungsjahren

Anzahl der Datensätze auf der Personenebene	Angaben im Erhebungsjahr ...							
	2001		2002		2003		2004	
	Anzahl	%	Anzahl	%	Anzahl	%	Anzahl	%
	Insgesamt							
Insgesamt	179 095	100,0	179 041	100,0	179 572	100,0	177 889	100,0
davon mit:								
- unplausiblen Angaben	5 681	3,2	5 631	3,1	5 611	3,1	5 442	3,1
- plausiblen Angaben	173 414	96,8	173 410	96,9	173 961	96,9	172 447	96,9
- Angaben im Erhebungsjahr...								
2001 2002 2003 2004								
X - - -	24 199	13,5	—	—	—	—	—	—
X X - -	14 405	8,0	14 405	8,0	—	—	—	—
X X X -	12 406	6,9	12 406	6,9	12 406	6,9	—	—
X X X X	120 550	67,3	120 550	67,3	120 550	67,1	120 550	67,8
- X - -	—	—	6 431	3,6	—	—	—	—
- X X -	—	—	3 385	1,9	3 385	1,9	—	—
- X X X	—	—	15 499	8,7	15 499	8,6	15 499	8,7
- - X -	—	—	—	—	6 090	3,4	—	—
- - X X	—	—	—	—	14 832	8,3	14 832	8,3
- - - X	—	—	—	—	—	—	19 775	11,1
Summe	171 560	95,8	172 676	96,4	172 762	96,2	170 656	95,9
- einem Ausfall in mindestens einem der anderen Erhe- bungsjahre ^{a)}	1 854	1,0	734	0,4	1 199	0,7	1 791	1,0
	davon in Privathaushalten							
Insgesamt	177 314	100,0	177 322	100,0	177 893	100,0	176 124	100,0
davon mit:								
- unplausiblen Angaben	5 462	3,1	5 451	3,1	5 455	3,1	5 245	3,0
- plausiblen Angaben	171 852	96,9	171 871	96,9	172 438	96,9	170 879	97,0
- Angaben im Erhebungsjahr...								
2001 2002 2003 2004								
X - - -	23 721	13,4	—	—	—	—	—	—
X X - -	14 125	8,0	14 125	8,0	—	—	—	—
X X X -	12 267	6,9	12 267	6,9	12 267	6,9	—	—
X X X X	119 891	67,6	119 891	67,6	119 891	67,4	119 891	68,1
- X - -	—	—	6 234	3,5	—	—	—	—
- X X -	—	—	3 342	1,9	3 342	1,9	—	—
- X X X	—	—	15 282	8,6	15 282	8,6	15 282	8,7
- - X -	—	—	—	—	5 903	3,3	—	—
- - X X	—	—	—	—	14 556	8,2	14 556	8,3
- - - X	—	—	—	—	—	—	19 365	11,1
Summe	170 004	95,9	171 141	96,5	171 241	96,3	169 094	96,0
- einem Ausfall in mindestens einem der anderen Erhe- bungsjahre ^{a)}	1 848	1,0	730	0,4	1 197	0,7	1 785	1,0
	Gemeinschaftsunterkünften							
Insgesamt	1 781	100,0	1 719	100,0	1 679	100,0	1 765	100,0
davon mit:								
- unplausiblen Angaben	219	12,3	180	10,5	156	9,3	197	11,2
- plausiblen Angaben	1 562	87,7	1 539	89,5	1 523	90,7	1 568	88,8
- Angaben im Erhebungsjahr...								
2001 2002 2003 2004								
X - - -	478	26,8	—	—	—	—	—	—
X X - -	280	15,7	280	16,3	—	—	—	—
X X X -	139	7,8	139	8,1	139	8,3	—	—
X X X X	659	37,0	659	38,3	659	39,2	659	37,3
- X - -	—	—	197	11,5	—	—	—	—
- X X -	—	—	43	2,5	43	2,6	—	—
- X X X	—	—	217	12,6	217	12,9	217	12,3
- - X -	—	—	—	—	187	11,1	—	—
- - X X	—	—	—	—	276	16,4	276	15,6
- - - X	—	—	—	—	—	—	410	23,2
Summe	1 556	87,4	1 535	89,3	1 521	90,6	1 562	88,5
- einem Ausfall in mindestens einem der anderen Erhe- bungsjahre ^{a)}	6	0,3	4	0,2	2	0,1	6	0,3

a) Es ist zu beachten, dass es auf der Personenebene keine temporären Ausfälle geben kann, da für alle Mitglieder eines Haushalts eine Auskunftspflicht besteht. Temporäre Ausfälle betreffen also stets alle Mitglieder eines Haushalts, weil der Haushalt im entsprechenden Erhebungsjahr nicht erreicht werden konnte.

2.3.3 Die Stichprobenentwicklung im Mikrozensus-Panel 2001–2004 auf der Haushaltsebene

Bei den Haushalten insgesamt (= Privathaushalte und Gemeinschaftsunterkünfte) liegt der Anteil plausible Zusammenführungsanteil geringfügig niedriger als auf der Personenebene. Von den etwa 83 500 Haushalten konnten rund 93 % plausibel zusammengeführt werden (siehe Tabelle 2.6). Diese 93 % differenzieren sich in etwa 68 % immobile und 23 % mobile Haushalte. Bei den restlichen 2 % ist in mindestens einem der Erhebungsjahre ein temporärer Ausfall zu verzeichnen.

Eine fast gleich Struktur ist bei den Privathaushalten festzustellen. Von den rund 83 000 Privathaushalten konnten etwa 93 % plausibel zusammengeführt werden. Davon entfielen 68 % auf immobile, 23 % auf mobile Privathaushalte und 2 % auf Privathaushalte mit mindestens einem temporären Ausfall in einem der Erhebungsjahre

Die bei den Privathaushalten erreichten Anteilwerte plausibel zusammengeführter Haushalte sind bei den Gemeinschaftsunterkünften nicht zu erzielen. Im Erhebungsjahr 2001 konnten von den 243 Gemeinschaftsunterkünften lediglich rund 41 % als plausibel zusammengeführt betrachtet werden. Dieser Anteilwert reduziert sich sukzessive von Erhebungsjahr zu Erhebungsjahr und erreicht im Jahr 2004 einen Wert von ca. 31 %. Dementsprechend niedrig ist auch der Anteil immobilier Gemeinschaftsunterkünfte, der zwischen 24 % und 31 % variiert. Mobilien Gemeinschaftsunterkünfte kommt – mit Ausnahme im Erhebungsjahr 2001 – ebenso wie Gemeinschaftsunterkünften mit mindestens einem temporären Ausfall in einem der Erhebungsjahre eine untergeordnete Bedeutung zu.

Diese Befunde zeigen auf, dass für Gemeinschaftsunterkünfte zwar Analysen auf der Personenebene durchgeführt werden können, nicht aber auf der Ebene der Gemeinschaftsunterkünfte selbst. Die Ursache hierfür ist sicherlich in nicht erheblichem Ausmaß der nur unzureichenden Erfassung der Zu-/und Fortzüge sowie Sterbefällen zuzuschreiben.

Tabelle 2.6: Umfang (un-)plausibler Datensätze auf der Haushaltsebene im Mikrozensus-Panel 2001–2004, differenziert nach Haushaltstypen in den einzelnen Erhebungsjahren

Anzahl der Datensätze auf der Haushaltsebene	Angaben im Erhebungsjahr ...							
	2001		2002		2003		2004	
	Anzahl	%	Anzahl	%	Anzahl	%	Anzahl	%
	Insgesamt							
Insgesamt	83 160	100,0	83 454	100,0	84 066	100,0	83 632	100,0
davon mit:								
- unplausiblen Angaben	4 802	5,8	6 076	7,3	6 246	7,4	6 841	8,2
- plausiblen Angaben	78 358	94,2	77 378	92,7	77 820	92,6	78 791	91,8
- Angaben im Erhebungsjahr...								
2001 2002 2003 2004								
X - - -	8 426	10,1	—	—	—	—	—	—
X X - -	5 863	7,1	5 863	7,0	—	—	—	—
X X X -	4 327	5,2	4 327	5,2	4 327	5,1	—	—
X X X X	56 797	68,3	56 797	68,1	56 797	67,6	56 797	67,9
- X - -	—	—	1 936	2,3	—	—	—	—
- X X -	—	—	1 204	1,4	1 204	1,4	—	—
- X X X	—	—	5 723	6,9	5 723	6,8	5 723	6,8
- - X -	—	—	—	—	1 859	2,2	—	—
- - X X	—	—	—	—	6 345	7,5	6 345	7,6
- - - X	—	—	—	—	—	—	7 058	8,4
Summe	75 413	90,7	75 850	90,9	76 255	90,7	75 924	90,8
- einem Ausfall in mindestens einem der anderen Erhe- bungsjahre	2 945	3,5	1 528	1,8	1 565	1,9	867	1,0
	davon							
	Privathaushalte							
Insgesamt	82 917	100,0	83 269	100,0	83 875	100,0	83 424	100,0
davon mit:								
- unplausiblen Angaben	4 658	5,6	5 961	7,2	6 121	7,3	6 698	8,0
- plausiblen Angaben	78 259	94,4	77 308	92,8	77 754	92,7	76 726	92,0
- Angaben im Erhebungsjahr...								
2001 2002 2003 2004								
X - - -	8 397	10,1	—	—	—	—	—	—
X X - -	5 857	7,1	5 857	7,0	—	—	—	—
X X X -	4 326	5,2	4 326	5,2	4 326	5,2	—	—
X X X X	56 739	68,4	56 739	68,1	56 739	67,6	56 739	68,0
- X - -	—	—	1 936	2,3	—	—	—	—
- X X -	—	—	1 204	1,4	1 204	1,4	—	—
- X X X	—	—	5 719	6,9	5 719	6,8	5 719	6,9
- - X -	—	—	—	—	1 858	2,2	—	—
- - X X	—	—	—	—	6 345	7,6	6 345	7,6
- - - X	—	—	—	—	—	—	7 056	8,5
Summe	75 319	90,8	75 781	91,0	76 191	90,8	75 859	90,9
- einem Ausfall in mindestens einem der anderen Erhe- bungsjahre	2 940	3,5	1 527	1,8	1 563	1,9	867	1,0
	Gemeinschaftsunterkünfte							
Insgesamt	243	100,0	185	100,0	191	100,0	208	100,0
davon mit:								
- unplausiblen Angaben	144	59,3	115	62,2	125	65,4	143	68,8
- plausiblen Angaben	99	40,7	70	37,8	66	34,6	65	31,3
- Angaben im Erhebungsjahr...								
2001 2002 2003 2004								
X - - -	29	11,9	—	—	—	—	—	—
X X - -	6	2,5	6	3,2	—	—	—	—
X X X -	1	0,4	1	0,5	1	0,5	—	—
X X X X	58	23,9	58	31,4	58	30,4	58	27,9
- X - -	—	—	0	—	—	—	—	—
- X X -	—	—	0	—	0	—	—	—
- X X X	—	—	4	2,2	4	2,1	4	1,9
- - X -	—	—	—	—	1	0,5	—	—
- - X X	—	—	—	—	1	0,5	1	0,5
- - - X	—	—	—	—	—	—	2	1,0
Summe	94	38,7	69	37,3	65	34,0	65	31,3
- einem Ausfall in mindestens einem der anderen Erhe- bungsjahre	5	2,1	1	0,5	1	0,5	0	—

2.4 Auswertungskonzepte in einem Mikrozensus-Paneldatensatz

Wie aus der Stichprobenentwicklung ersichtlich, liegen aufgrund räumlicher Mobilität und den demographischen Ereignissen Geburt und Tod sowie temporärer Ausfälle nicht für alle Personen Informationen zu allen Erhebungszeitpunkten vor. Bleiben die temporären Ausfälle unberücksichtigt, ergeben sich für die Analysen mit einem Mikrozensus-Panel in Abhängigkeit des einbezogenen Informationspotenzials verschiedenen Auswertungsstrategien, unter anderem:

- im Sinne eines Analysekonzepts mit Berücksichtigung aller verfügbaren Fälle²,
- im Sinne eines Weiterverfolgungskonzepts,
- im Sinne eines Rückverfolgungskonzepts oder
- im Sinne eines Analysekonzepts mit Berücksichtigung der vollständigen Fälle³.

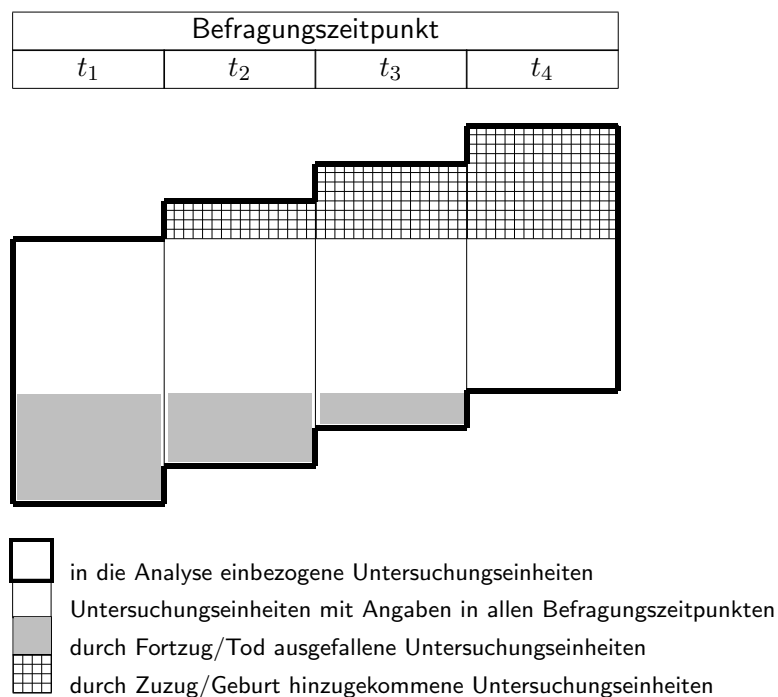
Die vorstehenden Konzepte werden nachfolgend kurz erläutert. Im Rahmen eines **Analysekonzepts mit Berücksichtigung aller verfügbaren Fälle** werden alle Untersuchungseinheiten in die Analyse einbezogen, und zwar unabhängig davon, über welchen Zeitraum für eine Untersuchungseinheit Informationen vorliegen (siehe nachfolgende Abbildung).

Mit diesem Konzept werden alle Informationen eines Mikrozensus-Paneldatensatzes ausgenutzt. Auch gestattet dieses Konzept einen Vergleich der fort- und zuziehenden Personen in den Übergängen zweier Erhebungszeitpunkte, so dass wanderungsbedingte Ausfälle und Ausfälle aufgrund der demographischen Ereignisse Geburt und Tod in thematischen Analysen unmittelbar berücksichtigt werden können.

²In der Literatur wird dieses Auswertungskonzept als "complete information analysis" oder "available case analysis" bezeichnet. Hier wird jedoch eine deutsche Terminologie gewählt, die von Toutenburg et al. (2004) verwendet wird.

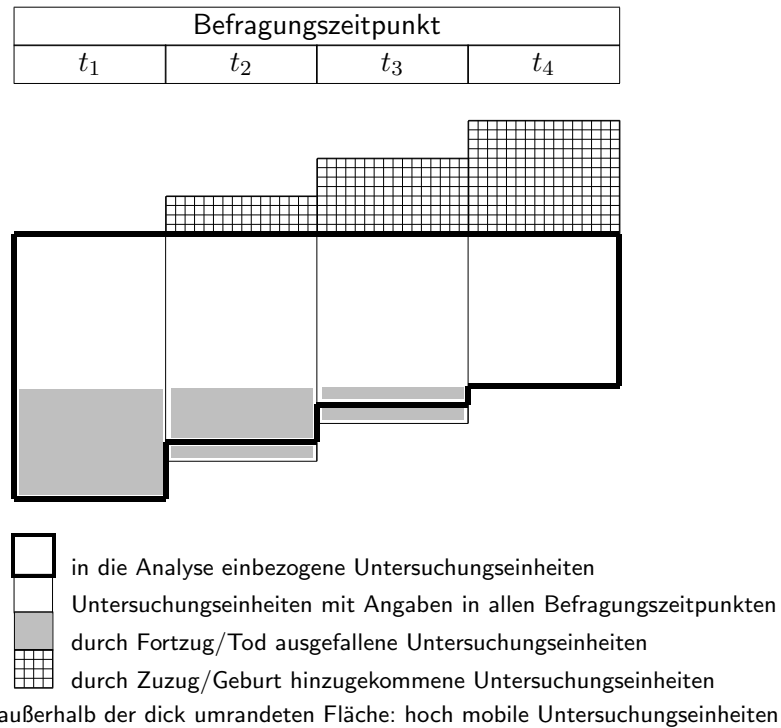
³In der Literatur wird dieses Analysekonzept oft als "complete cases analysis" bezeichnet. Aber auch hier wird eine deutsche Terminologie gewählt, die wie zuvor von Toutenburg et al. (2004) verwendet wird.

Abbildung 2.2: Umfang der Analyseeinheiten in einer Paneluntersuchung mit vier Befragungszeitpunkten ohne temporäre Ausfälle nach dem Analysekonzept mit Berücksichtigung aller verfügbaren Fälle – schematische Darstellung



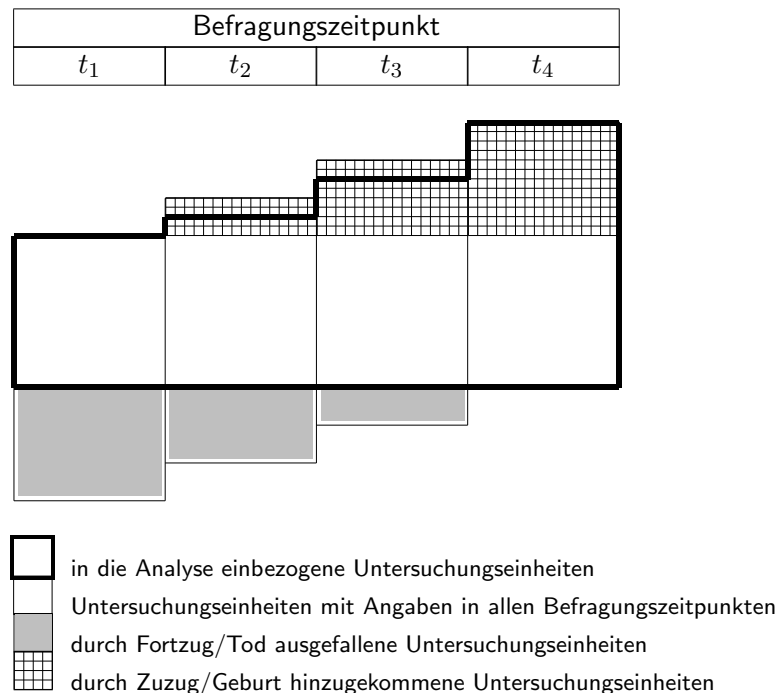
Beim **Weiterverfolgungskonzept** werden nur die Untersuchungseinheiten in die Analyse einbezogen, für die Informationen zum ersten Befragungszeitpunkt vorliegen (siehe Abbildung 2.3). Zu nachfolgenden Befragungszeitpunkten neu hinzugekommene Untersuchungseinheiten (in der Abbildung 2.3 als gerasterte Fläche wiedergegeben) und hoch mobile Untersuchungseinheiten, also Untersuchungseinheiten, die nach dem Befragungszeitpunkt t_1 in einen Auswahlbezirk zugezogen sind bzw. geboren wurden und den Auswahlbezirk vor dem Befragungszeitpunkt t_4 wieder verlassen haben bzw. gestorben sind (in der Abbildung 2.3 dargestellt als graue Fläche außerhalb der dick umrandeten Fläche), werden nicht berücksichtigt. Dieses Konzept konzentriert sich damit auf die "Verluste" in einem Mikrozensus-Panel.

Abbildung 2.3: Umfang der Analyseeinheiten in einer Paneluntersuchung mit vier Befragungszeitpunkten ohne temporäre Ausfälle nach dem Weiterverfolgungskonzept – schematische Darstellung



Analog zum Weiterverfolgungskonzept werden beim **Rückverfolgungskonzept** nur die Untersuchungseinheiten in die Analyse einbezogen, für die Informationen zum letzten Erhebungszeitpunkt vorlagen. Analysen, die auf diesem Konzept basieren, berücksichtigen also nur die "Gewinne" in einem Mikrozensus-Panel Datensatz.

Abbildung 2.4: Umfang der Analyseeinheiten in einer Paneluntersuchung mit vier Befragungszeitpunkten ohne temporäre Ausfälle nach dem Rückverfolgungskonzept – schematische Darstellung



Gerasterte Fläche außerhalb der dick umrandeten Fläche: hoch mobile Untersuchungseinheiten

Mit einem Vergleich der Ergebnisse aus Weiter- und Rückverfolgungskonzept lassen sich zwar Rückschlüsse auf wanderungsbedingte Ausfälle und Ausfälle aufgrund der demographischen Ereignisse Geburt und Tod ziehen, zu beachten ist hierbei jedoch, dass

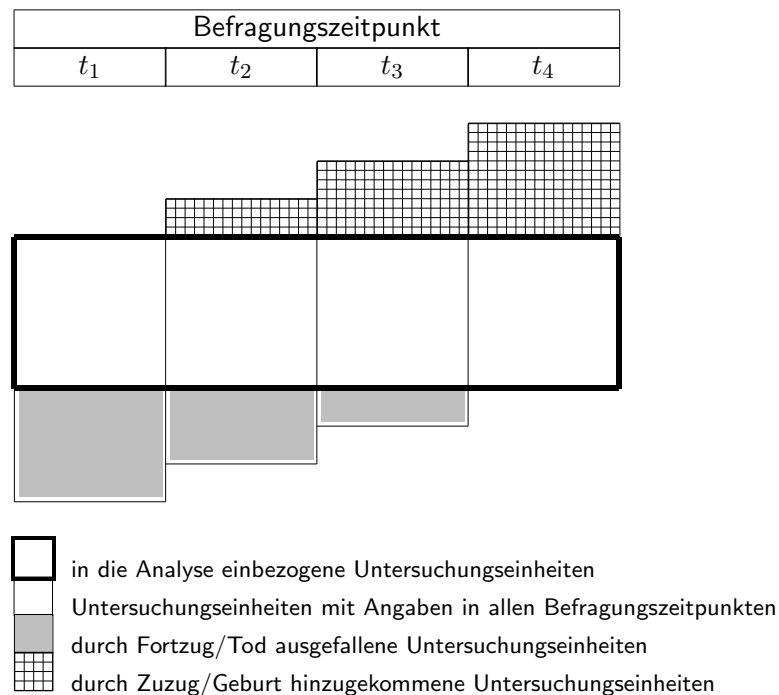
- (a) im Weiterverfolgungskonzept die Personen, die nach dem Erhebungszeitpunkt t_1 zugezogen oder geboren wurden, aus der Analyse ausgeschlossen werden und
- (b) im Rückverfolgungskonzept die Personen aus der Analyse ausgeschlossen bleiben, die vor dem Befragungszeitpunkt t_4 fortgezogen oder verstorben sind.

Unberücksichtigt bleiben also insbesondere die hoch mobilen Personen, die in einen bestehenden Mikrozensus-Auswahlbezirk nach dem Befragungszeitpunkt t_1 zugezogen und diesen bereits vor Durchführung der letzten (vierten) Befragung verlassen haben. Diese hoch mobilen Personen finden sich in den Abbildungen 2.3 und 2.4 wieder, und zwar als graue Fläche (Abbildung 2.3) bzw. gerasterte Fläche (Abbildung 2.4) außerhalb der dick umrandeten Fläche.

Es ist allerdings zu beachten, dass in geringem Ausmaß auch die Ereignisse Geburt und Tod in Verbindung mit räumlicher Mobilität dazu führen können, dass Personen weder zum ersten noch zum vierten Befragungszeitpunkt in den Mikrozensus einbezogen sind. Dies ist dann der Fall, wenn die beiden Ereignisse räumliche Mobilität und demographisches Ereignis bei einer Person/einem Haushalt nach dem ersten und vor dem letzten Befragungszeitpunkt stattfinden.

Als ein weiteres Auswertungskonzept bietet sich das **Analysekonzept mit Berücksichtigung nur vollständiger Fälle** an. Hierbei werden nur die Untersuchungseinheiten in die Analyse einbezogen, für die Informationen für alle Erhebungszeitpunkte vorliegen. Untersuchungseinheiten, die während des Erhebungszeitraums den Auswahlbezirk verlassen bzw. in den Auswahlbezirk neu hinzukommen, bleiben ausgeschlossen.

Abbildung 2.5: Umfang der Analyseeinheiten in einer Paneluntersuchung mit vier Befragungszeitpunkten ohne temporäre Ausfälle nach dem Analysekonzept mit Berücksichtigung der vollständigen Fälle – schematische Darstellung



Für Paneluntersuchungen muss in Abhängigkeit des jeweils herangezogenen Auswertungskonzeptes stets eine Vorselektion der Datensätze gemäß der Kennzeichnungsvariablen `perk1` (wenn die Auswertungen auf der Personenebene stattfinden sollen) bzw. `hherk1` (wenn die Auswertungen nur für Personen durchgeführt

werden sollen, für die alle Personensätze eines Haushalts als erklärt gelten) vorgenommen werden. Welche Werteselektion bei welchem Auswertungskonzept zur Anwendung kommen muss, ist nachfolgender Tabelle zu entnehmen; temporäre Ausfälle sind hierbei ausgeschlossen.

Tabelle 2.7: Zusammenhang von Auswertungskonzept und Werteselektion erklärter Personensätze

Auswertungskonzept	Werteselektion bei ...	
	Personenauswertungen	Personenauswertungen in Haushalten, in denen alle Personensätze plausibel sind
Analysekonzept mit Berücksichtigung aller verfügbaren Fälle	$1 \leq \text{perk1} \leq 10$	$1 \leq \text{hherk1} \leq 10$
Weiterverfolgungskonzept	$1 \leq \text{perk1} \leq 4$	$1 \leq \text{hherk1} \leq 4$
Rückverfolgungskonzept	$\text{perk1} = 4, 7, 9, 10$	$\text{hherk1} = 4, 7, 9, 10$
Analysekonzept mit Berücksichtigung nur vollständiger Fälle	$\text{perk1} = 4$	$\text{hherk1} = 4$

2.5 Längsschnittinformationen im Mikrozensus-Paneldatensatz 2001-2004

Infolge räumlicher Mobilität (Fort- und Zuzüge), den demographischen Ereignissen Geburt und Tod, Ausfällen sowie nicht zusammenführbaren Personensätzen, liegen nicht für alle Personen in allen Erhebungsjahren Angaben vor. Das Mikrozensus-Panel 2001-2004 ist somit ein unbalanciertes Panel. Zur Kennzeichnung der Verlaufsangaben bezüglich räumlicher Mobilität, demographischen Ereignissen, Ausfall und Zusammenführungserfolg wurden einige Hilfsmerkmale erzeugt, die nachfolgend kurz erläutert werden (vgl. Tabelle 2.8). Die ausführlichen Kennzeichnungen dieser Verlaufangaben sind dem Schlüsselverzeichnis zu entnehmen.

Zusätzlich zu diesen Hilfsmerkmalen, die im Datensatz den Variablen aus den einzelnen Erhebungsjahren vorangestellt sind, enthält der Mikrozensus-Paneldatensatz des Weiteren die Merkmale Bundesland, Auswahlbezirksnummer, Haushaltsnummer, Geburtsjahr und Geschlecht als so genannte "Längsschnittmerkmale". Diese Merkmale weisen für alle Erhebungsjahre, in denen Angaben für eine Person bzw. für einen Haushalt vorliegen, (eventuell korrigierte) Werte auf, und zwar unabhängig vom Zusammenführungserfolg. Vorgenommene Korrekturen bei den Angaben zum Geburtsjahr und zum Geschlecht wurden in gesonderten Variablen (beim Geburtsjahr in der Variablen `flag_gj` und beim Geschlecht in der Variablen `flag_sex`) dokumentiert. Liegen für das Geburtsjahr bzw. für das Geschlecht über

alle Erhebungsjahre hinweg, in denen eine Person befragt wurde, identische Werte vor, weist die jeweilige Kennzeichnungsvariable (`flag_gj` bzw. `flag_sex`) den Wert 0 auf, in allen anderen Fällen den Wert 1.

Gemäß den Datenschutzvorschriften wurden die Nummern für die Auswahlbezirke und Haushalte systemfrei vergeben. Die Auswahlbezirke werden, bezogen auf den Gesamtdatensatz, von 1 bis n durchnummeriert und die Privathaushalte innerhalb eines Auswahlbezirks von 1 bis n . Gemeinschaftsunterkünften wird die Haushaltsnummer 0 zugewiesen.

Tabelle 2.8: Längsschnittinformationen im Mikrozensus-Panel 2001-2004

Längsschnittinformationen im Mikrozensus-Panel 2001-2004	
Variablenname	Inhalt
<code>erfolg</code>	<p>Kennzeichnung der Zusammenführungsebene</p> <p>Die Variable <code>erfolg</code> gibt Auskunft darüber, auf welcher Ebene eine Personensatz zusammengeführt wurde.</p>
<code>perkl</code>	<p>Kennung plausibler Personensätze</p> <p>Die Variable <code>perkl</code> kennzeichnet den Zusammenführungserfolg auf der Personenebene. Aus ihr lässt sich erkennen für welche Erhebungszeitpunkte Personensätze plausibel – d.h. unter Berücksichtigung räumlicher Mobilität, demographischen Ereignissen und temporären Ausfällen – zusammengeführt werden konnten. Die genaue Kennzeichnung des Zusammenführungsstatus kann dabei dem Variablenlabel entnommen werden, das sich aus insgesamt acht Zeichen zusammensetzt. Jeweils zwei aufeinander folgende Zeichen kennzeichnen den Personenstatus in einem bestimmten Erhebungsjahr.</p> <p>Dabei informieren die beiden ersten Zeichen über den Personenstatus im Erhebungsjahr 2001, die Zeichen an dritter und vierter Stelle über den Personenstatus im Erhebungsjahr 2002, die Zeichen an fünfter und sechster Position über den Personenstatus im Erhebungsjahr 2003 und schließlich die Zeichen an siebter und achter Position über den Personenstatus im Erhebungsjahr 2004.</p> <p>Zur Kennzeichnung eines Personenstatus werden folgende Zeichenfolgen herangezogen:</p> <ul style="list-style-type: none"> (a) zwei aufeinander folgende Ziffern geben das Erhebungsjahr an und weisen darauf hin, dass für das entsprechende Erhebungsjahr Personenangaben vorliegen; (b) die Zeichenfolge "xx" weist darauf hin, dass in dem entsprechenden Erhebungsjahr zwar keine Personenangaben vorliegen, dies aber auf Grund vorliegender Veränderungen (Fort-, Zuzug, Geburt oder Tod) als erklärt gelten kann;

Fortsetzung der Tabelle auf nächster Seite

Längsschnittinformationen im Mikrozensus-Panel 2001-2004	
Variablenname	Inhalt
	<p>Nachfolgend sind einige Beispiele zur Interpretation der Variablenlabels gegeben:</p> <p>01xxxxxx Es liegen nur für das Erhebungsjahr 2001 Angaben vor und die Person ist im Übergang 2001 → 2002 fortgezogen oder verstorben.</p> <p>xx02xxxx Es liegen nur für das Erhebungsjahr 2002 Angaben vor und die Person ist im Übergang 2001 → 2002 zugezogen oder geboren und im Übergang 2002 → 2003 fortgezogen oder verstorben.</p>
hherkl	<p>Kennung plausibler Privathaushalte</p> <p>Die Variable <code>hherkl</code> kennzeichnet analog zu <code>perkl</code> den Zusammenführungserfolg auf der Ebene der Privathaushalte.</p> <p>Für die Kennzeichnung eines Datensatzes als plausibel auf der Haushaltsebene ist es zwar erforderlich, dass alle Personensätze eines Haushalts plausibel zusammengeführt wurden, aufgrund individueller Veränderungen ist es aber nicht erforderlich, dass <code>perkl=hherkl</code> gelten muss. So kann ein Haushalt plausible Angaben zu vier Erhebungszeitpunkten ausweisen, während für eine im Haushalt lebende Person nur für drei Erhebungszeitpunkte plausible Angaben vorliegen, weil sie beispielsweise nach der dritten Erhebung verstarb.</p> <p>Die Kennzeichnung eines plausiblen Privathaushalts erfolgt analog der Kennzeichnung eines plausiblen Personendatensatzes.</p>
hhk	<p>Kennung eines Datensatzes für Auswertungen auf der Haushaltsebene</p> <p>Die Variable <code>hhk</code> kennzeichnet innerhalb eines, als plausibel gekennzeichneten Haushalts einen Personendatensatz für den gilt:</p> $\text{perkl} = \text{hherkl}$ <p>Vor der Zuweisung wurde eine Vorsortierung nach dem Geburtsjahr der Personen innerhalb eines Haushalts vorgenommen, so dass die älteste Person in einem Haushalt, für die vorstehende Bedingung erfüllt ist, als "Stellvertreter/in" eingesetzt wird.</p> <p>Bitte beachten Sie, dass diese Stellvertreterposition unabhängig von der, im Querschnitt vorliegenden Kennung der Haushaltsbezugs-person vorgenommen wurde, und das für diese Stellvertreterposition auch keine weiteren Typisierungen gebildet wurden.</p>
gewinnh	<p>Kennzeichnung des Zuzugs eines Haushalts</p> <p>Mit der Variablen <code>gewinnh</code> wird der Zuzug eines Haushalts gekennzeichnet.</p>

Fortsetzung der Tabelle auf nächster Seite

Längsschnittinformationen im Mikrozensus-Panel 2001-2004	
Variablenname	Inhalt
verlusth	<p>Kennzeichnung der Art eines Verlustes auf der Haushaltsebene Mit der Variablen <code>verlusth</code> wird die Art eines Verlustes auf der Haushaltsebene gekennzeichnet, wobei unterschieden wird zwischen:</p> <ul style="list-style-type: none"> (a) dem Fortzug eines Haushalts, (b) dem Tod eines Haushalts sowie (b) dem Tod/Fortzug eines Haushalts.
gewinnp	<p>Kennzeichnung der Art eines Zugewinns auf der Personenebene Mit der Variablen <code>gewinnp</code> wird die Art eines Zugewinns auf der Personenebene gekennzeichnet, wobei unterschieden wird zwischen der Geburt und dem Zuzug einer Person.</p>
verlustp	<p>Kennzeichnung der Art eines Verlustes auf der Personenebene Mit der Variablen <code>verlustp</code> wird die Art eines Verlustes auf der Personenebene gekennzeichnet, wobei analog zur Haushaltsebene unterschieden wird zwischen:</p> <ul style="list-style-type: none"> (a) dem Fortzug einer Person, (b) dem Tod einer Person sowie (b) dem Tod/Fortzug einer Person.

2.6 Der Datensatzaufbau des Mikrozensus-Panels 2001–2004

Im Mikrozensus-Paneldatensatz 2001–2004 wird der übliche Datensatzaufbau des Mikrozensus beibehalten, d.h. die Variablen der einzelnen Erhebungsjahre werden pro Person horizontal in Blöcken (im so genannten Wide-Format) abgelegt.

Mit einer solchen Anordnung der Daten ist die Beibehaltung der originalen Variablennamen nicht mehr möglich⁴. Um die Umbenennungen der Variablennamen gegenüber dem jeweiligen Mikrozensus-Originalfile so gering wie möglich zu halten und unter der Vorgabe, dass die Maximallänge eines Variablennamens acht Zeichen nicht überschreiten sollte, wurde von den ursprünglichen Variablennamen in jedem Erhebungsjahr das erste Zeichen wie folgt ersetzt:

Erhebungsjahr	alter Variablenname ^{a)}	neuer Variablenname ^{a)}
2001	ef...	ff...
2002	ef...	gf...
2003	ef...	hf...
2004	ef...	if...

a) Die Auslassungspunkte symbolisieren die jeweilige Feldnummer

Mit der Vergabe der neuen Variablennamen wird also an die Vergabe der Variablennamen im Mikrozensus-Panel 1996–1999 angeknüpft, wonach ein Variablenname in einem Mikrozensus-Panel für das erste Erhebungsjahr eines neuen Mikrozensusgesetzes mit dem Buchstaben a beginnt, der Variablenname des zweiten Erhebungsjahres mit dem Buchstaben b usw.

Daraus ergibt sich folgender Datensatzaufbau im Mikrozensus-Panel 2001–2004 für die Variablen aus den Mikrozensus-Originalfiles:

Daten- satz von ...	Variablenname im MZ-Panel 2001–2004 für das Erhebungsjahr ...											
	2001			2002			2003			2004		
	ff1	...	ffp	gf1	...	gfq	hf1	...	hfr	if1	...	ifs
Person 1												
Person 2												
⋮												
Person n												

p = letzte Feldnummer der Merkmale im Erhebungsjahr 2001

q = letzte Feldnummer der Merkmale im Erhebungsjahr 2002

r = letzte Feldnummer der Merkmale im Erhebungsjahr 2003

s = letzte Feldnummer der Merkmale im Erhebungsjahr 2004

⁴Die erhobenen Informationen des Mikrozensus werden in einem festen Datensatzformat abgespeichert, so dass die Werte einer Variablen immer an der gleichen Spaltenposition eines Datensatzes stehen. Die Variablennamen setzen sich in der Regel daher aus zwei Bestandteilen zusammen. Aus dem Präfix *ef*, mit dem alle Variablennamen beginnen und für "Eingabefeld" steht, und der daran anschließenden Spaltenposition im ursprünglichen Datensatz. In einigen Ausnahmefällen ist der Spaltenposition noch ein Zusatz angefügt, der bei einer hierarchischen Merkmalsausprägung (wie beispielsweise der Wirtschafts- oder Berufsklassifikation) die entsprechende Hierarchieebene anzeigt.

Für Analysen mit wiederholten Messungen wird teilweise ein anderer Datenaufbau erwartet, nämlich, dass die Werte der wiederholten Messungen einer Untersuchungseinheit nicht nebeneinander, sondern untereinander angeordnet sind. Diese vertikale Datenordnung wird auch als Long-Format bezeichnet. Da eine entsprechende Umorganisation der Daten, vom Wide- in ein Long-Format, mit den gängigen Statistikpaketen (wie beispielsweise SPSS⁵, SAS⁶ oder Stata⁷) für die Nutzer/innen mit einem geringen Aufwand verbunden ist, wurde auf die Bereitstellung eines gesonderten Datensatzes des Mikrozensus-Panels 2001–2004 im Long-Format verzichtet.

2.7 Erhebungs- und Analyseeinheiten im Mikrozensus-Panel

Das Mikrozensus-Panel weist wie die Mikrozensus-Originalfiles die gleiche hierarchische Datenstruktur bezüglich der Erhebungs- und potenziellen Analyseeinheiten auf. Diese – nicht nur für Forscher/innen — informativen Inhalte können derzeit in vollem Umfang nur im Querschnitt, nicht jedoch im Längsschnitt ausgewertet werden. Die Gründe hierfür sind zum einen, dass die Identifikatoren für Gebäude, Wohnungen, Familien und Lebensgemeinschaften nicht längsschnittkonsistent sind und zum anderen können sich die Beziehungen innerhalb der Einheiten Wohnung, Haushalt, Familie oder Lebensgemeinschaft infolge eines Wechsels der jeweiligen Bezugsperson ändern. Ein solcher Wechsel der Bezugsperson kann darin begründet sein, dass die Bezugsperson räumlich mobil ist, sie im Beobachtungszeitraum verstarb oder dass in den einzelnen Querschnittsdaten verschiedene Haushaltsmitglieder als Bezugsperson eingesetzt wurden.

⁵Bei SPSS steht hierfür der Befehl VARSTOCASES zur Verfügung.

⁶Mit SAS kann dies mit der Prozedur PROC TRANSPOSE bewerkstelligt werden.

⁷In Stata kann dies mit dem Kommando RESHAPE durchgeführt werden.

2.8 Weiterführende Literatur

- Heidenreich, H.-J. (2002): *Längsschnittdaten aus dem Mikrozensus: Basis für neue Analysemöglichkeiten*, in: Allgemeines Statistisches Archiv; **86**: 213-231.
- Heidenreich, H.-J.; Nöthen, M. (2002): *Der Wandel der Lebensformen im Spiegel des Mikrozensus*, in: Wirtschaft und Statistik; 1/2002: 26-38.
- Heidenreich, H.-J.; Herter-Eschweiler, R. (2002): *Längsschnittdaten aus dem Mikrozensus. Basis für neue Analysemöglichkeiten*, in: Wirtschaft und Statistik; 8/2002: 669-680.
- Herter-Eschweiler, R. (2003): *Längsschnittdaten aus dem Mikrozensus. Basis für neue Analysemöglichkeiten. Dokumentationsband* Bonn: Statistisches Bundesamt (unveröffentlichtes Manuskript).
- Lengerer, A.; Bohr, J.; Janßen, A. (2005): *Haushalte, Familien und Lebensformen im Mikrozensus – Konzepte und Typisierungen*, ZUMA-Arbeitsbericht Nr. 2005/05. Mannheim: Zentrum für Umfragen, Methoden und Analysen. URL: http://www.gesis.org/Publikationen/Berichte/ZUMA_Arbeitsberichte/05/AB_05_05.pdf (letzter Zugriff: 23. März 2008)
- Schimpl-Neimanns, B. (2006): *Filekonzept zum Mikrozensuspanel*. Arbeitspapier #12, Methodenverbund "Aufbereitung und Bereitstellung des Mikrozensus als Panelstichprobe". URL: <http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Content/Wissenschaftsforum/MethodenVerfahren/Mikrozensus/Arbeitspapiere/Arbeitspapier12,property=file.pdf> (letzter Zugriff: 23. März 2008)
- Toutenburg, H.; Heumann, H.; Nittner, T. (2004): *Statistische Methoden bei unvollständigen Daten*. Discussion paper 380, Ludwig-Maximilian-Universität München. URL: <http://www.stat.uni-muenchen.de/sfb386/papers/dsp/paper380.ps> (letzter Zugriff: 23. März 2008)
- Zühlke, S. (2001): *Längsschnittdaten auf der Basis des Mikrozensus. Methodische Probleme und Lösungsansätze*, in: Statistische Analysen und Studien Nordrhein-Westfalen; **53**: 3-13. URL: <https://webshop.lds.nrw.de/webshop/gratis/Z089%20200154.pdf> (letzter Zugriff: 23. März 2008)

Kapitel 3

Das faktisch anonymisierte Mikrozensus-Panel 2001–2004

Wolf Bihler (Statistisches Bundesamt; Gruppe II A¹)

Robert Herter-Eschweiler (Statistisches Bundesamt, Gruppe VIII C)

Heike Wirth (Zentrum für Umfragen, Methoden und Analysen – ZUMA²)

Das Mikrozensus-Panel 2001–2004 wird für die Durchführung wissenschaftlicher Vorhaben unabhängigen inländischen Forschungseinrichtungen gemäß § 16, Abs. 6 des Bundesstatistikgesetzes (BStatG) als faktisch anonymisiertes Einzelmaterial zur Verfügung gestellt. Faktisch anonymisiert bedeutet in diesem Zusammenhang, dass die Einzelangaben *„nur mit einem unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft zugeordnet werden können“* (§ 16, Abs. 6 BStatG). Zur Einhaltung der faktischen Anonymisierung wurden zum einen allgemeine Schutzvorkehrungen getroffen und zum anderen spezifische Anonymisierungsmaßnahmen durchgeführt. Diese Vorkehrungen zur Erstellung des faktisch anonymisierten Mikrozensus-Panels 2001–2004 sind im nachfolgenden Kapitel 3.1 näher erläutert.

Wie bereits erwähnt enthält das faktisch anonymisierte Mikrozensus-Panel 2001–2004 nicht nur die plausibel zusammengeführten Datensätze, sondern umfasst alle Personensätze der ausgewählten Auswahlbezirke. Damit besteht die Möglichkeit mit dem faktisch anonymisierten Mikrozensus-Panel 2001–2004 auch Querschnittsanalysen bezüglich der einzelnen Erhebungsjahre durchzuführen.

Um eine möglichst gute Anpassung der Querschnitte des Mikrozensus-Panels an die jeweiligen Mikrozensus-Originalfiles der einzelnen Erhebungsjahre zu erreichen, werden daher neben den Original-Hochrechnungsfaktoren so genannte angepasste Hochrechnungsfaktoren bereitgestellt, sowohl für deskriptive Auswertungen, als auch für Inferenzstatistiken, mit denen Parameter der Grundgesamtheit geschätzt

¹Auswahlplan der Substichprobenziehung des faktisch anonymisierten Mikrozensus-Panels 2001–2004.

²Anonymisierungskonzept.

werden können oder mit denen Stichprobenbefunde zur statistischen Prüfung von Hypothesen herangezogen werden können. Die Bildung dieser angepasster Hochrechnungsfaktoren wird in Kapitel 3.2 erläutert.

Da das vorliegende faktisch anonymisierte Mikrozensus-Panel 2001–2004 nur Personen aus Privathaushalten beinhaltet, können die im Querschnitt hochgerechneten Randverteilungen des faktisch anonymisierten Mikrozensus-Panels 2001–2004 nicht mit veröffentlichten Tabellen der Fachserien des Mikrozensus verglichen werden, da in diesen zumeist auch die Personen aus Gemeinschaftsunterkünften enthalten sind. Für ausgewählte Merkmale und unterschiedliche Bevölkerungskonzepte wurden daher die Randverteilungen der ausgewählten Merkmale für die Erhebungsjahre 2001 bis 2004 (unter Berücksichtigung der jeweiligen Hochrechnung, in den Mikrozensus-Originalfiles sind dies die jeweiligen Original-Hochrechnungsfaktoren und im Mikrozensus-Panel die angepassten Hochrechnungsfaktoren für deskriptive Auswertungen) der Mikrozensus-Originalfiles sowie der Querschnitte des faktisch anonymisierten Mikrozensus-Panels 2001–2004 ausgewiesen. Aufgrund des umfangreichen Datenmaterials sind die Tabellen in einer gesonderten Datei hinterlegt. Allerdings enthält das Handbuch zur Illustration einige wenige Tabellen.

3.1 Maßnahmen zur Erstellung des faktisch anonymisierten Mikrozensus-Panels 2001–2004

3.1.1 Allgemeine Schutzvorkehrungen

Die allgemeinen Schutzvorkehrungen sehen vor:

1. Vertragliche Bindung des Empfängers/der Empfängerin hinsichtlich:
 - (a) geeigneter technisch-organisatorischer Maßnahmen zur Kontrolle der Datennutzung,
 - (b) des Verbots von Reidentifikationsversuchen,
 - (c) der Nutzungsbegrenzung auf das anzugebende wissenschaftliche Vorhaben,
 - (d) des Verbots der Weitergabe der Daten an Dritte,
 - (e) der Löschung bzw. Rückgabe der Daten nach Abschluss des wissenschaftlichen Vorhabens,
 - (f) der Behandlung von Datenausügen oder Datenduplikaten sowie
 - (g) des Verbots von Nachforschungen bezüglich der lokalen Umsetzung der Stichprobenpläne.
2. Geheimhaltung der lokalen Umsetzung der Stichprobenpläne.
3. Systemfreie Anordnung der Daten.

3.1.2 Spezifische Anonymisierungsmaßnahmen

An spezifischen Anonymisierungsmaßnahmen wurden durchgeführt:

1. Als einziges Regionalmerkmal wird das Bundesland weitergegeben.
2. Regionaltypen sind im faktisch anonymisierten Mikrozensus-Panel 2001–2004 nur vergrößert im Datensatz enthalten, und zwar dergestalt, dass
 - (a) keine einzelne Gemeinde mit weniger als 500 000 Einwohnern identifizierbar ist und
 - (b) ein Gemeindetyp, dem mehrere Gemeinden zugehören, in jedem Bundesland mindestens 400 000 Einwohner umfasst.
3. Es ist keine Nationalität oder Gruppe von Nationalitäten mit weniger als 50 000 Einwohnern in der Bundesrepublik Deutschland identifizierbar.
4. Bei allen übrigen Merkmalen wurde – soweit erforderlich – eine Ausprägungsvergrößerung vorgenommen, so dass in der univariaten Randverteilung jede ausgewiesene Merkmalsausprägung für die Bundesrepublik Deutschland mindestens 5 000 Fälle umfasst.
5. Es wird eine Substichprobe weitergegeben, deren Auswahlatz etwa 70 Prozent der Haushalte bzw. Personen beträgt. Die Beschreibung des Auswahlplans zur Generierung der Substichprobe enthält die nachfolgende Tabelle 3.1.
6. Es werden keine Merkmale der Unterstichprobe für die Erhebungsjahre 2001 und 2004 ausgewiesen. Denn auf Grund der Konzeption der Unterstichprobe werden einerseits wechselnde Teile eines Rotationsviertels in das Unterstichprobenprogramm einbezogen. Andererseits ergeben sich infolge der Genauigkeitsansprüche an die EU-Arbeitskräfteerhebung unterschiedliche Auswahlätze in den einzelnen Regierungsbezirken (siehe hierzu die Erläuterungen auf Seite 28). Damit wären Regierungsbezirke mit einem hohen Auswahlatz identifizierbar. Nach den Datenschutzregeln zur Umsetzung der faktischen Anonymisierung dürfen aber keine Regierungsbezirke als regionale Einheiten erkennbar sein, weshalb die Merkmale der Unterstichprobe für die Erhebungsjahre 2001 und 2004 gänzlich unterdrückt wurden.

Welche konkreten Merkmalsvergrößerungen zur Einhaltung der spezifischen Anonymisierungsmaßnahmen vorgenommen wurden, können Sie der Datensatzbeschreibung, dem so genannten Schlüsselverzeichnis entnehmen.

Tabelle 3.1: Auswahlplan der Substichprobenziehung zur Generierung eines faktisch anonymisierten Mikrozensus-Panels 2001–2004

Auswahlplan der Substichprobenziehung des faktisch anonymisierten Mikrozensus-Panels 2001–2004	
Auswahlgesamtheit	Personen des dritten Rotationsviertels der fünften Stichprobe.
Auswahleinheiten	Auswahlbezirke
Auswahltechnik	<p>Sortierung der Auswahlbezirke nach:</p> <ol style="list-style-type: none"> 1. Bundesland 2. Größenklasse (basierend auf Dezilen) der Auswahlbezirke 3. Kennung über Grundausswahl/Aktualisierung der Auswahlbezirke 4. Kennung über den Einbezug eines Auswahlbezirks in die Unterstichprobe 5. Regierungsbezirk 6. Regionale Anpassungsschicht 7. Regionale Schicht 8. Regionale Untergruppe 9. Gemeindegrößenklasse 10. Nummer des Auswahlbezirks <p>getrennt nach Auswahlbezirken mit Privathaushalten und Auswahlbezirken mit Gemeinschaftsunterkünften.</p> <p>Die Sortierung folgt damit weitgehend der Schichtung des Originalmaterials.</p> <p>Neu aufgenommen in die Sortierung wurde die Größenklasse der Auswahlbezirke sowie die Kennung eines Auswahlbezirks über dessen Einbezug in die Unterstichprobe.</p> <p>Zwar wurde bei der Grundausswahl bzw. wird bei der jährlichen Aktualisierung der Auswahlbezirke eine einheitliche Größe der Gebäudegrößenklassen angestrebt, sie unterscheiden sich jedoch fluktuationsbedingt erheblich. Insbesondere die Auswahlbezirke der Grundausswahl, deren Gebäudegrößenklassen auf der Grundlage der Volkszählungsdaten 1987 basieren, weisen mittlerweile zum Teil erhebliche Schwankungen auf. Um eine daraus resultierende Erhöhung der Fehlervarianz zu begrenzen, erfolgte die zusätzliche Aufnahme einer Sortierung nach Größenklassen der Auswahlbezirke.</p>

Fortsetzung der Tabelle auf nachfolgender Seite.

Fortsetzung von Tabelle 3.1

Die Messung der Größenklassen erfolgte durch die Zahl der Personen in Privathaushalten bzw. Gemeinschaftsunterkünften in den Auswahlbezirken, und zwar als Durchschnitt über diejenigen Erhebungsjahre, in denen der Auswahlbezirk in den Mikrozensus einbezogen wurde und Angaben von Personen in Privathaushalten vorliegen.

Mit der Aufnahme der Kennung über den Einbezug eines Auswahlbezirks in die Unterstichprobe soll die Unterstichprobeneigenschaft in der Substichprobenziehung kontrolliert werden.

Beide Merkmale (Größenklasse und Unterstichprobenkennung eines Auswahlbezirks) wurden für so wichtig gehalten, dass sie in der Sortierung nach dem Bundesland an zweiter bzw. dritter Stelle platziert wurden.

Die jeweils getrennte Schichtung von Auswahlbezirken mit Privathaushalten und Auswahlbezirken mit Gemeinschaftsunterkünften wurde vorgenommen, um die unterschiedliche Struktur der jeweiligen Personengemeinschaft berücksichtigen zu können.

Schichtung:

Jeweils zehn in der Reihenfolge der Sortierung aufeinander folgende Auswahlbezirke bilden eine Schicht.

Schichtungswechsel innerhalb einer Sortierung sind ungünstig, da sie tendenziell einen höheren Varianzbeitrag liefern. Im Hinblick auf regionale Analysemöglichkeiten nach Bundesländern und der weitgehenden Beibehaltung homogener Größenklassen der Auswahlbezirke, werden lediglich bei Sortierwechsel von Bundesland und Größenklassen der Auswahlbezirke "unvollständige" Schichten mit weniger als zehn Auswahlbezirke zugelassen. Die letzte Schicht kann ebenfalls weniger als zehn Auswahlbezirke umfassen.

Auswahlsatz:

Es wird ein einheitlicher Auswahlsatz von 7/10 in allen Schichten angestrebt. Dazu werden in jeder Schicht mit Hilfe eines "Pseudozufallszahlengenerators" sieben unterschiedliche Zahlen im Intervall a ($1 \leq a \leq 10, a \in \mathbb{N}$) gezogen. Jeder Auswahlbezirk innerhalb einer Schicht mit gezogener Zufallszahl gilt als ausgewählt. Bei den "unvollständigen" Schichten gelten diejenigen Auswahlbezirke als gezogen, deren laufende Nummern innerhalb der Schicht sich unter den sieben zufällig aus den Zahlen eins bis zehn gezogenen Zahlen befindet. Der Stichprobenumfang einer solchen Schicht hängt also vom Zufall ab, die Ziehungswahrscheinlichkeit von 7/10 bleibt aber erhalten.

3.2 Die Stichprobenentwicklung des faktisch anonymisierten Mikrozensus-Panels 2001–2004

Die Substichprobenziehung des faktisch anonymisierten Mikrozensus-Panels 2001–2004 umfasst 70 Prozent der Auswahleinheiten, in denen im Erhebungszeitraum 2001 bis 2004 mindestens eine Person aus einem Privathaushalt oder einer Gemeinschaftsunterkunft befragt wurde. Durch die bei der Substichprobenziehung kontrollierte Größe der Auswahleinheiten – gemessen an der durchschnittlichen Wohnbevölkerung in einer Auswahleinheit – müssten damit annähernd auch 70 % einer jeden, durch *perkl* bzw. *hherkl* definierten Teileinheit, in der Substichprobe enthalten sein.

Auf der Personenebene wird der angestrebte Auswahlsatz von 70 Prozent für die Personen insgesamt weitgehend eingehalten (siehe Tabelle 3.2). Allerdings ist hinsichtlich der einzelnen Teileinheiten eine systematische Verschiebung zu beobachten. Personen, die im Verlauf des Erhebungszeitraums von 2001 bis 2004 aus einer Auswahleinheit fortzogen oder verstarben bzw. zuzogen oder geboren wurden, weisen mit einem Auswahlsatz von rund 69,5 % eine geringfügige Untererfassung (gemessen am erwarteten Auswahlsatz von 70 %) auf. Demgegenüber ist bei den hoch mobilen Personen, also Personen, die nach 2001 in einen Auswahlbezirk zuzogen und den Auswahlbezirk vor 2004 wieder verlassen haben, mit einem Auswahlsatz von etwa 71 % eine leichte Überrepräsentation gegeben.

Eine nahezu identische Struktur wie bei den Personen insgesamt zeigt sich bei den Personen aus Privathaushalten. Der angestrebte Auswahlsatz von 70 % wird bei den Personen insgesamt als auch den immobilen Personen weitgehend eingehalten, wohingegen bei den mobilen Personen mit einem Auswahlsatz von etwa 69,5 % eine geringfügige Unterrepräsentation und bei den hoch mobilen Personen mit einem Auswahlsatz von rund 71 % eine gemäßigte Überrepräsentation beobachtet werden kann.

Bei den Personen aus Gemeinschaftsunterkünften kann der angestrebte Auswahlsatz von 70 % nicht erreicht werden. Ist in den beiden ersten Erhebungsjahren ein zu hoher Auswahlsatz festzustellen (73,5 %), liegt er in den beiden letzten Erhebungsjahren mit 68,5 % deutlich unter der 70 %-Marke. Merkwürdig ist der Auswahlsatz zudem in der Gruppe der immobilen sowie der mobilen Personen. In beiden Teilgruppen rund drei Prozentpunkte über dem angestrebten Auswahlsatz von 70 %. Eine Ursache dieses Sachverhalts dürfte unter anderem in den hohen Schwankungen hinsichtlich der Anzahl und Größe von Gemeinschaftsunterkünften zu finden sein.

Ein ähnliches Bild zeichnet sich auch auf der Ebene der Haushalte ab (siehe Tabelle 3.3), weshalb auf die kurze Beschreibung der Auswahlsätze verzichtet wird.

Anzumerken ist jedoch, dass die Prozentangaben bei den Gemeinschaftsunterkünften auf der Ebene der Haushalte aufgrund der geringen Fallzahlen nicht interpretierbar sind.

Tabelle 3.2: Die Stichprobenentwicklung des faktisch anonymisierten Mikrozensus-Panels 2001–2004 auf der Personenebene

Personen insgesamt	Anzahl der Personen insgesamt im MZ-Panel Originalfile (MZPO) und im faktisch anonymisierten MZ-Panel (fMZP) sowie des Auswahlstatus im Erhebungsjahr ...											
	2001			2002			2003			2004		
	Anz. MZPO	Anz. fMZP	Ausw. satz	Anz. MZPO	Anz. fMZP	Ausw. satz	Anz. MZPO	Anz. fMZP	Ausw. satz	Anz. MZPO	Anz. fMZP	Ausw. satz
Insgesamt	179 095	125 227	69,9	179 041	125 202	69,9	179 572	125 474	69,9	177 889	124 322	69,9
davon mit:												
- unplausiblen Angaben	5 681	3 974	70,0	5 631	3 914	69,5	5 611	3 866	68,9	5 442	3 756	69,0
- plausiblen Angaben	173 414	121 253	69,9	173 410	121 288	69,9	173 961	121 608	69,9	172 447	120 566	69,9
davon mit:												
- Angaben zum Erhebungsjahr ...												
2002 2002 2003 2004												
X - - - -	24 199	16 894	69,8	-	-	-	-	-	-	-	-	-
X X - - -	14 405	9 946	69,0	14 405	9 946	69,0	-	-	-	-	-	-
X X X - -	12 406	8 612	69,4	12 406	8 612	69,4	12 406	8 612	69,4	-	-	-
X X X X -	120 550	84 439	70,0	120 550	84 439	70,0	120 550	84 439	70,0	120 550	84 439	70,0
- X - - -	-	-	-	6 431	4 605	71,6	-	-	-	-	-	-
- X X - -	-	-	-	3 385	2 387	70,5	3 385	2 387	70,5	-	-	-
- X X X -	-	-	-	15 499	10 787	69,6	15 499	10 787	69,6	15 499	10 787	69,6
- - X - -	-	-	-	-	-	-	6 090	4 346	71,4	-	-	-
- - X X -	-	-	-	-	-	-	14 832	10 141	68,4	14 832	10 141	68,4
- - - X -	-	-	-	-	-	-	-	-	-	19 775	13 892	70,3
Summe	171 560	119 891	69,9	172 676	120 776	69,9	172 762	120 712	69,9	170 656	119 259	69,9
- einem Ausfall in mindestens einem der anderen Erhebungsjahre)	1 854	1 362	73,5	734	512	69,8	1 199	896	74,7	1 791	1 307	73,0

a) Es ist zu beachten, dass es auf der Personenebene keine temporären Ausfälle geben kann, da für alle Mitglieder eines Haushalts eine Auskunftspflicht besteht. Temporäre Ausfälle betreffen also stets alle Mitglieder eines Haushalts, weil der Haushalt im entsprechenden Erhebungsjahr nicht erreicht werden konnte.

Fortsetzung der Tabelle auf nächster Seite

Fortsetzung von Tabelle 3.2

	Anzahl der Personen in Privathaushalten im MZ-Panel Originalfilie (MZPO) und im faktisch anonymisierten MZ-Panel (fMZP) sowie des Auswahlsatzes im Erhebungsjahr											
	2001			2002			2003			2004		
	Anz. MZPO	Ausw. satz	Anz. fMZP	Anz. MZPO	Ausw. satz	Anz. fMZP	Anz. MZPO	Ausw. satz	Anz. fMZP	Anz. MZPO	Ausw. satz	Anz. fMZP
Insgesamt	177 314	69,9	123 919	177 322	69,9	123 927	177 893	69,9	124 309	176 124	69,9	123 095
davon mit:												
- unplausiblen Angaben	5 462	69,6	3 801	5 451	69,2	3 771	5 455	68,9	3 756	5 245	68,8	3 607
- plausiblen Angaben	171 852	69,9	120 118	171 871	69,9	120 156	172 438	69,9	120 553	170 879	69,9	119 488
davon mit:												
- Angaben zum Erhebungsjahr . . .												
2002 2002 2003 2004												
X - - - -	23 721	69,8	16 547	-	-	-	-	-	-	-	-	-
X X - - -	14 125	69,0	9 743	14 125	69,0	9 743	-	-	-	-	-	-
X X X - -	12 267	69,4	8 513	12 267	69,4	8 513	12 267	69,4	8 513	12 267	69,4	-
X X X X -	119 891	70,0	83 959	119 891	70,0	83 959	119 891	70,0	83 959	119 891	70,0	83 959
- X - - -	-	-	-	6 234	71,2	4 441	-	-	-	-	-	-
- X X - -	-	-	-	3 342	70,5	2 357	3 342	70,5	2 357	3 342	70,5	-
- X X X -	-	-	-	15 282	69,6	10 635	15 282	69,6	10 635	15 282	69,6	10 635
- - X X -	-	-	-	-	-	-	5 903	71,3	4 209	-	-	-
- - - X X	-	-	-	-	-	-	14 556	68,6	9 986	14 556	68,6	9 986
- - - - X	-	-	-	-	-	-	-	-	-	19 365	70,3	13 607
Summe	170 004	69,9	118 762	171 141	69,9	119 648	171 241	69,9	119 659	169 094	69,9	118 187
- einem Ausfall in mindestens einem der anderen Erhebungsjahre)	1 848	73,4	1 356	730	69,6	508	1 197	894	74,7	1 785	1 301	72,9

a) Es ist zu beachten, dass es auf der Personenebene keine temporären Ausfälle geben kann, da für alle Mitglieder eines Haushalts eine Auskunftspflicht besteht. Temporäre Ausfälle betreffen also stets alle Mitglieder eines Haushalts, weil der Haushalt im entsprechenden Erhebungsjahr nicht erreicht werden konnte.

Fortsetzung der Tabelle auf nächster Seite

Fortsetzung von Tabelle 3.2

Personen in Gemeinschaftsunterkünften	Anzahl der Personen in Gemeinschaftsunterkünften im MZ-Panel Originalfile (MZPO) und im faktisch anonymisierten MZ-Panel (fMZP) sowie des Auswahlsatzes im Erhebungsjahr											
	2001			2002			2003			2004		
	Anz. MZPO	Anz. fMZP	Ausw. satz	Anz. MZPO	Anz. fMZP	Ausw. satz	Anz. MZPO	Anz. fMZP	Ausw. satz	Anz. MZPO	Anz. fMZP	Ausw. satz
Insgesamt	1 781	1 312	73,7	1 719	1 264	73,5	1 679	1 150	68,5	1 765	1 213	68,7
davon mit:												
- unplausiblen Angaben	219	170	77,6	180	139	77,2	156	108	69,2	197	146	74,1
- plausiblen Angaben	1 562	1 142	73,1	1 539	1 125	73,1	1 523	1 042	68,4	1 568	1 067	68,0
davon mit:												
- Angaben zum Erhebungsjahr												
2002 2002 2003 2004												
X - - - -	478	350	73,2	-	-	-	-	-	-	-	-	-
X X - - -	280	202	72,1	280	202	72,1	-	-	-	-	-	-
X X X - -	139	103	74,1	139	103	74,1	139	103	74,1	-	-	-
X X X X	659	481	73,0	659	481	73,0	659	481	73,0	659	481	73,0
- X - - -	-	-	-	197	160	81,2	-	-	-	-	-	-
- X X - -	-	-	-	43	30	69,8	43	30	69,8	-	-	-
- X X X	-	-	-	217	145	66,8	217	145	66,8	217	145	66,8
- - X X	-	-	-	-	-	-	187	132	70,6	-	-	-
- - - X X	-	-	-	-	-	-	276	149	54,0	276	149	54,0
- - - - X	-	-	-	-	-	-	-	-	-	410	286	69,8
Summe	1 556	1 136	73,0	1 535	1 121	73,0	1 521	1 040	68,4	1 562	1 061	67,9
- einem Ausfall in mindestens einem der anderen Erhebungsjahre)	6	6	100	4	4	100	2	2	100	6	6	100

Tabelle 3.3: Die Stichprobenentwicklung des faktisch anonymisierten Mikrozensus-Panels 2001–2004 auf der Haushaltsebene

Haushalte insgesamt	Anzahl der Haushalte im MZ-Panel Originalfile (MZPO) und im faktisch anonymisierten MZ-Panel (fMZP) sowie des Auswahlstatus im Erhebungsjahr ...											
	2001			2002			2003			2004		
	Anz. MZPO	Anz. fMZP	Ausw. satz	Anz. MZPO	Anz. fMZP	Ausw. satz	Anz. MZPO	Anz. fMZP	Ausw. satz	Anz. MZPO	Anz. fMZP	Ausw. satz
Insgesamt	83 160	58 108	69,9	83 454	58 388	70,0	84 066	58 751	69,9	83 632	58 469	69,9
davon mit:												
- unplausiblen Angaben	4 802	3 315	69,0	6 076	4 269	70,3	6 246	4 408	70,6	6 841	4 768	69,7
- plausiblen Angaben	78 358	54 793	69,9	77 378	54 119	69,9	77 820	54 343	69,8	76 791	53 701	69,9
davon mit:												
- Angaben zum Erhebungsjahr ...												
2001 2002 2003 2004												
X - - - -	8 426	5 891	69,9	-	-	-	-	-	-	-	-	-
X X - - -	5 863	4 079	69,6	5 863	4 079	69,6	-	-	-	-	-	-
X X X - -	4 327	2 987	69,0	4 327	2 987	69,0	4 327	2 987	69,0	-	-	-
X X X X -	56 797	39 805	70,1	56 797	39 805	70,1	56 797	39 805	70,1	56 797	39 805	70,1
- X - - -	-	-	-	1 936	1 377	71,1	-	-	-	-	-	-
- X X - -	-	-	-	1 204	835	69,4	1 204	835	69,4	-	-	-
- X X X -	-	-	-	5 723	3 969	69,4	5 723	3 969	69,4	5 723	3 969	69,4
- - X - -	-	-	-	-	-	-	1 859	1 315	70,7	-	-	-
- - X X -	-	-	-	-	-	-	6 345	4 333	68,3	6 345	4 333	68,3
- - - X -	-	-	-	-	-	-	-	-	-	7 058	4 989	70,7
- - - - X	-	-	-	-	-	-	-	-	-	-	-	-
Summe	75 413	52 760	70,0	75 850	53 052	69,9	76 255	53 244	69,8	75 924	53 096	69,9
- einem Ausfall in mindestens einem der anderen Erhebungsjahre	2 945	2 033	69,0	1 528	1 067	69,8	1 565	1 100	70,3	867	605	69,8

Fortsetzung der Tabelle auf nächster Seite

Fortsetzung von Tabelle 3.3

Privathaushalte	Anzahl der Privathaushalte im MZ-Panel Originalfile (MZPO) und im faktisch anonymisierten MZ-Panel (fMZP) sowie des Auswahlstatus im Erhebungsjahr ...											
	2001			2002			2003			2004		
	Anz. MZPO	Ausw. fMZP	Ausw. satz	Anz. MZPO	Anz. fMZP	Ausw. satz	Anz. MZPO	Anz. fMZP	Ausw. satz	Anz. MZPO	Anz. fMZP	Ausw. satz
Insgesamt	82 917	57 968	69,9	83 269	58 253	70,0	83 875	58 627	69,9	83 424	58 335	69,9
davon mit:												
- unplausiblen Angaben	4 658	3 249	69,8	5 961	4 188	70,3	6 121	4 336	70,8	6 698	4 684	69,9
- plausiblen Angaben	78 259	54 719	69,9	77 308	54 065	69,9	77 754	54 291	69,8	76 726	53 651	69,9
davon mit:												
- Angaben zum Erhebungsjahr ...												
2001 2002 2003 2004												
X - - - -	8 397	5 868	69,9	-	-	-	-	-	-	-	-	-
X X - - -	5 857	4 075	69,6	5 857	4 075	69,6	-	-	-	-	-	-
X X X - -	4 326	2 986	69,0	4 326	2 986	69,0	4 326	2 986	69,0	-	-	-
X X X X -	56 739	39 761	70,1	56 739	39 761	70,1	56 739	39 761	70,1	56 739	39 761	70,1
- X - - -	-	-	-	1 936	1 377	71,1	-	-	-	-	-	-
- X X - -	-	-	-	1 204	835	69,4	1 204	835	69,4	-	-	-
- X X X -	-	-	-	5 719	3 965	69,3	5 719	3 965	69,3	5 719	3 965	69,3
- - X X -	-	-	-	-	-	-	1 858	1 314	70,7	-	-	-
- - - X X	-	-	-	-	-	-	6 345	4 332	68,3	6 345	4 332	68,3
- - - - X	-	-	-	-	-	-	-	-	-	7 056	4 988	70,7
Summe	75 319	52 690	70,0	75 781	52 999	69,9	76 191	53 193	69,8	75 859	53 046	69,9
- einem Ausfall in mindestens einem der anderen Erhebungsjahre	2 940	2 029	69,0	1 527	1 066	69,8	1 563	1 098	70,2	867	605	69,8

Fortsetzung der Tabelle auf nächster Seite

Fortsetzung von Tabelle 3.3

Gemeinschaftsunterkünfte	Anzahl der Gemeinschaftsunterkünfte im MZ-Panel Originalfile (MZPO) und im faktisch anonymisierten MZ-Panel (fMZP) sowie des Auswahlsatzes im Erhebungsjahr											
	2001			2002			2003			2004		
	Anz. MZPO	Anz. fMZP	Ausw. satz	Anz. MZPO	Anz. fMZP	Ausw. satz	Anz. MZPO	Anz. fMZP	Ausw. satz	Anz. MZPO	Anz. fMZP	Ausw. satz
Insgesamt	243	140	57,6	185	135	73,0	191	124	64,9	208	134	64,4
davon mit:												
- unplausiblen Angaben	144	66	45,8	115	81	70,4	125	72	57,6	143	84	58,7
- plausiblen Angaben	99	74	74,7	70	54	77,1	66	52	78,8	65	50	76,9
davon mit:												
- Angaben zum Erhebungsjahr . . .												
2001 2002 2003 2004												
X - - - -	29	21	72,4	-	-	-	-	-	-	-	-	-
X X - - -	6	4	66,7	6	4	66,7	-	-	-	-	-	-
X X X - -	1	1	100	1	1	100	1	1	100	-	-	-
X X X X	58	44	75,9	58	44	75,9	58	44	75,9	58	44	75,9
- X - - -	-	-	-	0	0	-	-	-	-	-	-	-
- X X - -	-	-	-	0	0	-	0	0	-	-	-	-
- X X X	-	-	-	4	4	100	4	4	100	4	4	100
- X X X	-	-	-	-	-	-	1	1	100	-	-	-
- - X X	-	-	-	-	-	-	1	1	100	1	1	100
- - - X	-	-	-	-	-	-	-	-	-	2	1	50,0
- - - - X	94	70	74,5	69	53	76,8	65	51	78,5	65	50	76,9
Summe												
- einem Ausfall in mindestens einem der anderen Erhebungsjahre	5	4	80,0	1	1	100	1	1	100	0	0	-

3.3 Das Variablenspektrum des faktisch anonymisierten Mikrozensus-Panels 2001–2004

Das Variablenspektrum des faktisch anonymisierten Mikrozensus-Panels 2001–2004 orientiert sich an der Variablenauswahl der faktisch anonymisierten Mikrozensus-Grundfiles der entsprechenden Erhebungsjahre. Der Datensatz enthält mit Ausnahme der Erhebungsjahre 2001 und 2004 auch die Merkmale des Ergänzungsprogramms, die mit einem reduzierten Auswahlsatz erhoben werden, als auch die Merkmale des Zusatzprogramms, die nicht jährlich, sondern in einem Abstand von vier Jahren und teilweise ebenfalls mit einem reduzierten Auswahlsatz erhoben werden.

Die Merkmale, die mit einem reduzierten Auswahlsatz erhoben werden, wurden für die Erhebungsjahre 2001 und 2004 nicht in den Datensatz aufgenommen, da der Auswahlsatz, der am Ergänzungsprogramm beteiligten Haushalte/Personen in der Regel innerhalb eines Rotationsviertels von Erhebungsjahr zu Erhebungsjahr variiert (siehe [Tabelle 1.2](#)). Für die Erhebungsjahre 2001 und 2004 hätte dies zur Folge, dass einzelne Regierungsbezirke identifizierbar gewesen wären und dies der Anonymisierungsmaßnahme widersprochen hätte, wonach als unmittelbare Regionalangabe in einem faktisch anonymisierten Grundfile nur das Bundesland enthalten sein darf.

Für die Merkmale des Zusatzprogramms liegen in einem Mikrozensus-Panel über vier Erhebungszeitpunkte zwar nur für jeweils ein Erhebungsjahr Angaben vor, die Merkmale wurden aber dennoch in das faktisch anonymisierte Mikrozensus-Panel aufgenommen. Damit haben Sie die Möglichkeit auch Querschnittsanalysen durchzuführen. Allerdings gegenüber den faktisch anonymisierten Mikrozensus-Grundfile mit einem deutlich geringeren Auswahlsatz³. Da als Auswahlseinheiten der Substichprobenziehung des faktisch anonymisierten Mikrozensus-Panels 2001–2004 die so genannten Auswahlbezirke herangezogen wurden, ergibt sich für die einzelnen Erhebungsjahre kein einheitlicher Auswahlsatz auf der Ebene der Haushalte/Personen. Die im faktisch anonymisierten Mikrozensus-Grundfile praktizierte Vorgehensweise, die "Original-Hochrechnungsfaktoren" des Mikrozensus multiplikativ mit dem Kehrwert der Auswahlwahrscheinlichkeit in die Substichprobe des faktisch anonymisierten Mikrozensus-Grundfile zu gelangen, zu verknüpfen, kann daher im Fall des Mikrozensus-Panels nicht angewendet werden. Zudem sei daran erinnert, dass die Merkmale des Ergänzungsprogramms mit einem, über die Erhebungsjahre hinweg, unterschiedlichen Auswahlsatz in einem Rotationsviertel enthalten sind. Es wurden daher für jedes Erhebungsjahr und jeden Hochrechnungsfaktor so genannte angepasste Hochrechnungsfaktoren ermittelt, mit denen Sie (gewichtet) Querschnittsanalysen durchführen können. Die Beschreibung der angepassten Hochrechnungsfaktoren finden Sie im nachfolgenden Kapitel.

³Die Auswahlwahrscheinlichkeit in einen Querschnitt des faktisch anonymisierten Mikrozensus-Panels 2001–2004 zu gelangen beträgt etwa 0.175

3.4 Die Bildung angepasster Hochrechnungsfaktoren für Querschnittsanalysen im Mikrozensus-Panel 2001–2004

Um eine möglichst gute Anpassung der Mikrozensus-Panel-Querschnitte an die jeweiligen Querschnitte der Mikrozensus der Erhebungsjahre 2001 bis 2004 zu erreichen, werden neben den Original-Hochrechnungsfaktoren so genannte angepasste Hochrechnungsfaktoren bereitgestellt. Bei diesen angepassten Hochrechnungsfaktoren wurde der jeweilige Original-Hochrechnungsfaktor des Erhebungsjahres t mit dem Kehrwert der Auswahlwahrscheinlichkeit multipliziert, dass eine Person aus dem zweiten Rotationsviertel der vierten Stichprobe in die Substichprobe des Mikrozensus-Panels 2001–2004 des t -ten Erhebungsjahres gelangt ($t = 2001, 2002, 2003, 2004$).

Bei der Bestimmung der Auswahlwahrscheinlichkeiten wurde nicht nur auf die hochgerechnete Bevölkerungszahl insgesamt abgestellt, sondern es wurden auch das Bundesland, das Geschlecht, die Staatsangehörigkeit, das Alter sowie der Erwerbsstatus einer Person mit berücksichtigt. Die Abgrenzung der Subpopulationen erfolgte getrennt für das frühere Bundesgebiet (einschließlich Berlin) und die neuen Bundesländer nach der Kreuzkombination folgender Variablen:

- (a) für das frühere Bundesgebiet (einschließlich Berlin):
 - Bundesland (11 Bundesländer)
 - × Geschlecht (männlich, weiblich)
 - × Altersgruppe (< 15 Jahre, $15 \leq 45$ Jahre, $45 \leq 65$ Jahre, ≥ 65 Jahre)
 - × deutsche Staatsangehörigkeit
 - × Erwerbsstatus (erwerbstätig, nicht erwerbstätig)
 - × nicht deutsche Staatsangehörigkeit
- (a) für die neuen Bundesländer:
 - Bundesland (5 Bundesländer)
 - × Geschlecht (männlich, weiblich)
 - × Altersgruppe (< 15 Jahre, $15 \leq 45$ Jahre, $45 \leq 65$ Jahre, ≥ 65 Jahre)
 - × Erwerbsstatus (erwerbstätig, nicht erwerbstätige).

Für einen angepassten Hochrechnungsfaktor einer Person i in der j -ten Subpopulation des t -ten Erhebungsjahres mit ($t = 2001, 2002, 2003, 2004$) ($HF_{j,t}^{Anp}$) ergibt sich somit:

$$HF_{i,j,t}^{Anp} = HF_{i,j,t} * \frac{n_{j,t}^{MZQ_g}}{n_{j,t}^{MZP_g}}$$

- mit: $HF_{i,j,t}^{Anp}$ = Angepasster Hochrechnungsfaktor der i-ten Person in der j-ten Subpopulation des t-ten Erhebungsjahres.
 $HF_{i,j,t}$ = Hochrechnungsfaktor der i-ten Person in der j-ten Subpopulation aus dem Mikrozensus-Grundfile des t-ten Erhebungsjahres.
 $n_{j,t}^{MQg}$ = Anzahl der hochgerechneten Personen im Mikrozensus-Originalfile in der j-ten Subpopulation des t-ten Erhebungsjahres.
 $n_{j,t}^{MPg}$ = Anzahl der hochgerechneten Personen in der faktisch anonymisierten Substichprobe des Mikrozensus-Panel-Querschnitts in der j-ten Subpopulation des t-ten Erhebungsjahres, wobei als Hochrechnungsfaktor der Hochrechnungsfaktor aus dem Mikrozensus-Originalfile übernommen wurde.

Der Quotient $n_{j,t}^{MQg} / n_{j,t}^{MPg}$ bezeichnet dabei den Korrekturfaktor.

Das Verfahren zur Ermittlung angepasster Hochrechnungsfaktoren weicht damit geringfügig von der Vorgehensweise der Bildung angepasster Hochrechnungsverfahren im Mikrozensus-Panel 1996-1999 ab. Die Modifikation wurde bewußt vorgenommen, um sicherzustellen, dass alle Kreuzkombinationen auch in der Substichprobe eine Zellenbesetzung aufweisen.

Da das faktisch anonymisierte Mikrozensus-Panel 2001–2004 nur auf einem Rotationsviertel basiert und eine 70-prozentige Substichprobe aus einem Rotationsviertel ist, ergibt sich eine Auswahlwahrscheinlichkeit, dass eine Person aus einem Mikrozensus-Originalfile in das faktisch anonymisierte Mikrozensus-Panel 2001–2004 gelangt, von $0,25 * 0,70 = 0,175$. Bei der Ermittlung der Auswahlwahrscheinlichkeit, dass eine Person auch in die Unterstichprobe einbezogen ist, ist zusätzlich der Auswahlsatz der Unterstichprobe, der in den einzelnen Erhebungsjahren in einem Rotationsviertel unterschiedlich ausfällt (siehe Tabelle 1.2 auf Seite 29), zu berücksichtigen. Für die einzelnen Erhebungsjahre ergeben sich für den Bundesdurchschnitt somit folgende Auswahlwahrscheinlichkeiten, dass eine Person aus der Unterstichprobe des Mikrozensus-Originalfiles in die Unterstichprobe des faktisch anonymisierten Mikrozensus-Panels 2001–2004 gelangt:

Tabelle 3.4: Auswahlwahrscheinlichkeit für eine Person aus der Unterstichprobe des Mikrozensus-Originalfiles in die Unterstichprobe des faktisch anonymisierten Mikrozensus-Panels 2001–2004 zu gelangen

Erhebungsjahr	Auswahlwahrscheinlichkeit
2002	$0,25 * 0,844 * 0,70 = 0,1477$
2003	$0,25 * 0,843 * 0,70 = 0,1475$

Die so vorgenommenen Korrekturen der Hochrechnungsfaktoren sind jedoch nur für deskriptive Analysen geeignet, da die Fallzahlen auf 1/100 der Gesamtbevölkerung hochgerechnet wurden. Für die Anwendung inferenzstatistischer Auswertungen werden daher des Weiteren Hochrechnungsfaktoren bereitgestellt, die zwar die Struktur der gewichteten Verteilungen wiedergeben, deren Fallzahlen aber auf die Stichprobengröße eines jeden Erhebungsjahres eines Mikrozensus-Panel-Querschnitts normiert wurden:

$$HF_{i,t,Stp} = HF_{i,t} * k_{i,j,t} * \frac{n_t^{MZP}}{n_t^{MZPg}}$$

- mit: $HF_{i,t,Stp}$ = Auf den Stichprobenumfang des faktisch anonymisierten Mikrozensus-Panel-Querschnitts normierter Hochrechnungsfaktor im t-ten Erhebungsjahr für die Person i .
- $HF_{i,t}$ = Aus dem Mikrozensus-Originalfile übernommener Hochrechnungsfaktor des t-ten Erhebungsjahres für die Person i .
- $k_{i,j,t}$ = Korrekturfaktor zur Bildung angepasster Hochrechnungsfaktoren an das jeweilige Mikrozensus-Originalfile für die Person i .
- n_t^{MZP} = Anzahl der befragten Personen in der faktisch anonymisierten Substichprobe des Mikrozensus-Panel-Querschnitts im t-ten Erhebungsjahr
- n_t^{MZPg} = Anzahl der hochgerechneten Personen in der faktisch anonymisierten Substichprobe des Mikrozensus-Panel-Querschnitts im t-ten Erhebungsjahr.

Für Querschnittsauswertungen stehen in den einzelnen Erhebungsjahren für Analysen auf der Basis des Einbezugs aller Personen und auf der Basis der Unterstichprobe sowie für die Personen- und Haushaltsebene jeweils zwei angepasste Hochrechnungsfaktoren zur Verfügung:

1. angepasste Hochrechnungsfaktoren für deskriptive Auswertungen, die die Fallzahlen des Mikrozensus-Panels eines Erhebungsjahres auf 1/100 der Gesamtbevölkerung hochrechnen und
2. angepasste Hochrechnungsfaktoren für inferenzstatistische Auswertungen, die unter Beibehaltung der Struktur der auf die Gesamtbevölkerung hochgerechneten Verteilungen die Fallzahlen auf die Anzahl der tatsächlich befragten Personen eines Erhebungsjahres "normiert" sind.

Die Variablennamen und -labels der einzelnen angepassten Hochrechnungsfaktoren können Sie nachfolgender Tabelle entnehmen:

Tabelle 3.5: Variablennamen und -labels der angepassten Hochrechnungsfaktoren in den einzelnen Erhebungsjahren des Mikrozensus-Panels 2001–2004

Angepasster Hochrechnungsfaktor . . .		
für das Erhebungsjahr	Variablenname	Variablenlabel
2001	ff750d	(EF750) MZ01: Personen-Hochrechnungsfaktor - (korrigiert für deskriptive Statistiken)
	ff750i	(EF750) MZ01: Personen-Hochrechnungsfaktor - (korrigiert für Inferenzstatistiken)
	ff751d	(EF751) MZ01: HH- und Familien-Hochrechnungsfaktor - (korrigiert für deskriptive Statistiken)
	ff751i	(EF751) MZ01: HH- und Familien-Hochrechnungsfaktor - (korrigiert für Inferenzstatistiken)
2002	gf750d	(EF750) MZ02: Personen-Hochrechnungsfaktor - (korrigiert für deskriptive Statistiken)
	gf750i	(EF750) MZ02: Personen-Hochrechnungsfaktor - (korrigiert für Inferenzstatistiken)
	gf751d	(EF751) MZ02: HH- und Familien-Hochrechnungsfaktor - (korrigiert für deskriptive Statistiken)
	gf751i	(EF751) MZ02: HH- und Familien-Hochrechnungsfaktor - (korrigiert für Inferenzstatistiken)
	gf755d	(EF755) MZ02: Personen-Hochrechnungsfaktor - UStp (korrigiert für deskriptive Statistiken)
	gf755i	(EF755) MZ02: Personen-Hochrechnungsfaktor - UStp (korrigiert für Inferenzstatistiken)
	gf756d	(EF756) MZ02: HH- und Familien-Hochrechnungsfaktor - UStp (korrigiert für deskriptive Statistiken)
	gf756i	(EF756) MZ02: HH- und Familien-Hochrechnungsfaktor - UStp (korrigiert für Inferenzstatistiken)
	gf761d	(EF761) MZ02: Wohnungsfaktor - (korrigiert für deskriptive Statistiken)
	gf761i	(EF761) MZ02: Wohnungsfaktor - (korrigiert für Inferenzstatistiken)
2003	hf750d	(EF750) MZ03: Personen-Hochrechnungsfaktor - (korrigiert für deskriptive Statistiken)
	hf750i	(EF750) MZ03: Personen-Hochrechnungsfaktor - (korrigiert für Inferenzstatistiken)
	hf751d	(EF751) MZ03: HH- und Familien-Hochrechnungsfaktor - (korrigiert für deskriptive Statistiken)

Fortsetzung der Tabelle auf nächster Seite

Angepasster Hochrechnungsfaktor . . .		
für das Erhebungs-jahr	Variable-name	Variablenlabel
2004	hf751i	(EF751) MZ03: HH- und Familien-Hochrechnungsfaktor - (korrigiert für Inferenzstatistiken)
	hf755d	(EF755) MZ03: Personen-Hochrechnungsfaktor - UStp (korrigiert für deskriptive Statistiken)
	hf755i	(EF755) MZ03: Personen-Hochrechnungsfaktor - UStp (korrigiert für Inferenzstatistiken)
	hf756d	(EF756) MZ03: HH- und Familien-Hochrechnungsfaktor - UStp (korrigiert für deskriptive Statistiken)
	hf756i	(EF756) MZ03: HH- und Familien-Hochrechnungsfaktor - UStp (korrigiert für Inferenzstatistiken)
	if750d	(EF750) MZ04: Personen-Hochrechnungsfaktor - (korrigiert für deskriptive Statistiken)
	if750i	(EF750) MZ04: Personen-Hochrechnungsfaktor - (korrigiert für Inferenzstatistiken)
	if751d	(EF751) MZ04: HH- und Familien-Hochrechnungsfaktor - (korrigiert für deskriptive Statistiken)
	if751i	(EF751) MZ04: HH- und Familien-Hochrechnungsfaktor - (korrigiert für Inferenzstatistiken)

3.5 Randverteilungen ausgewählter Merkmale in den Mikrozensus-Originalfiles und den Querschnitten des faktisch anonymisierten Mikrozensus-Panels 2001-2004

Das faktisch anonymisierte Mikrozensus-Panel 2001-2004 enthält zwar Personen aus Privathaushalten und Gemeinschaftsunterkünften, dennoch wird an der Praxis festgehalten für ausgewählte Merkmale (zur Liste der ausgewählten Merkmale siehe unten stehende Auflistung) und unterschiedliche Bevölkerungskonzepte

- Bevölkerung am Ort ihrer Hauptwohnung,
- erwerbstätige Bevölkerung am Ort ihrer Hauptwohnung und
- Bevölkerung in Privathaushalten am Familienwohnsitz (nach dem neuen Konzept der Lebensformen)

die Randverteilungen der ausgewählten Merkmale für die Erhebungsjahre 2001 bis 2004 (unter Berücksichtigung der jeweiligen Hochrechnung, in den Mikrozensus-Originalfiles sind dies die jeweiligen Original-Hochrechnungsfaktoren und im

Mikrozensus-Panel die angepassten Hochrechnungsfaktoren für deskriptive Auswertungen) der Mikrozensus-Originalfiles sowie der Querschnitte des faktisch anonymisierten Mikrozensus-Panels 2001-2004 auszuweisen. Die entsprechenden Tabellen finden Sie in einem gesonderten Skript.

Zur Illustration finden Sie nachfolgend die Tabelle der Randverteilungen der Bevölkerung am Ort der Hauptwohnung nach ausgewählten Merkmalen im Mikrozensus-Originalfile und dem Querschnitt des Mikrozensus-Panels mit Personengewichtung für das Erhebungsjahr 2002 (im Anhangband ist dies die Tabelle 2a).

Liste der ausgewiesenen Merkmale des Vergleichs zwischen Mikrozensus-Originalfile und Mikrozensus-Panel-Querschnitt:

- Bundesland,
- Geschlecht,
- Familienstand,
- Alter (in 5-Jahresschritten),
- Staatsangehörigkeit (3er Kategorisierung),
- Erwerbsstatus,
- Stellung im Beruf in der gegenwärtigen Tätigkeit,
- Voll-/Teilzeit,
- überwiegender Lebensunterhalt,
- individuelles Nettoeinkommen,
- höchster allgemeinbildender Schulabschluss,
- höchster beruflicher Ausbildungsabschluss,
- Stellung im Beruf in der früheren Tätigkeit,
- Art des Gebäudes sowie
- Jahr, in dem die Wohnung gebaut wurde.

Tabelle 3.6: Bevölkerung am Ort der Hauptwohnung nach ausgewählten Merkmalen im MZ-Originalfile und dem Querschnitt des MZ-Panels im Erhebungsjahr 2002

Bevölkerung am Ort der Hauptwohnung im Erhebungsjahr 2002						
Gegenstand der Nachweisung	MZ-Originalfile		Querschnitt des MZ-Panels		Differenz	
	Anzahl	Prozent	Anzahl	Prozent	absolut	relativ
Insgesamt	82 456	100,0	82 456	100,0	0	0,00
davon						
im Bundesland						
Schleswig-Holstein	2 808	3,4	2 808	3,4	0	0,00
Hamburg	1 727	2,1	1 727	2,1	0	0,00
Niedersachsen	7 964	9,7	7 964	9,7	0	0,00
Bremen	669	0,8	669	0,8	0	0,00
Nordrhein-Westfalen	18 055	21,9	18 055	21,9	0	0,00
Hessen	6 080	7,4	6 080	7,4	0	0,00
Rheinland-Pfalz	4 048	4,9	4 048	4,9	0	0,00
Baden-Württemberg	10 614	12,9	10 614	12,9	0	0,00
Bayern	12 354	15,0	12 354	15,0	0	0,00
Saarland	1 065	1,3	1 065	1,3	0	0,00
Berlin	3 389	4,1	3 389	4,1	0	0,00
Brandenburg	2 588	3,1	2 588	3,1	0	0,00
Mecklenburg-Vorpommern ..	1 755	2,1	1 809	2,1	0	0,00
Sachsen	4 373	5,3	4 549	5,3	0	0,00
Sachsen-Anhalt	2 572	3,1	2 719	3,1	0	0,00
Thüringen	2 407	2,9	2 484	2,9	0	0,00
nach dem Geschlecht						
männlich	40 283	48,9	40 283	48,9	0	0,00
weiblich	42 174	51,1	42 174	51,1	0	0,00
nach dem Familienstand						
ledig	31 618	38,3	31 630	38,4	12	0,04
verheiratet	40 098	48,6	39 955	48,5	-143	-0,36
verwitwet	6 376	7,7	6 497	7,9	121	1,90
geschieden	4 364	5,3	4 373	5,3	9	1,22

Fortsetzung der Tabelle auf nächster Seite

Bevölkerung am Ort der Hauptwohnung im Erhebungsjahr 2002						
Gegenstand der Nachweisung	MZ-Originalfile		Querschnitt des MZ-Panels		Differenz	
	Anzahl	Prozent	Anzahl	Prozent	absolut	relativ
nach Altersgruppen						
0 < 5 Jahre	3 772	4,6	3 774	4,6	2	0,05
5 < 10 Jahre	3 957	4,8	3 940	4,8	-17	-0,43
10 < 15 Jahre	4 601	5,6	4 616	5,6	15	0,33
15 < 20 Jahre	4 620	5,6	4 577	5,6	-43	-0,94
20 < 25 Jahre	4 608	5,6	4 675	5,7	67	1,45
25 < 30 Jahre	4 379	5,3	4 385	5,3	7	0,15
30 < 35 Jahre	5 952	7,2	5 938	7,2	-14	-0,23
35 < 40 Jahre	6 924	8,4	6 930	8,4	6	0,09
40 < 45 Jahre	6 509	7,9	6 487	7,9	-22	-1,34
45 < 50 Jahre	5 860	7,1	5 780	7,0	-79	-1,36
50 < 55 Jahre	5 518	6,7	5 527	6,7	8	0,15
55 < 60 Jahre	4 723	5,7	4 794	5,8	71	1,51
60 < 65 Jahre	6 137	7,4	6 211	7,5	74	1,20
65 < 70 Jahre	4 773	5,8	4 829	5,9	56	1,18
70 < 75 Jahre	3 833	4,6	3 737	4,5	-96	-2,50
75 < 80 Jahre	3 001	3,6	2 953	3,6	-48	-1,59
≥ 80 Jahre	3 290	4,0	3 303	4,0	13	0,40
nach der Staatsangehörigkeit						
Deutsche(r)	77 328	91,4	75 316	91,3	-12	-0,02
Ausländer(in) aus einem EU-Staat	1 836	2,2	1 793	2,2	-43	-2,36
Ausländer(in) aus einem Nicht-EU-Staat	5 292	6,4	5 347	6,5	55	1,04
nach Art der Erwerbsbeteiligung						
erwerbstätig	36 536	44,3	36 510	44,3	-26	-0,07
erwerbslos nach ILO	3 486	4,2	3 534	4,3	48	1,39
sonstig erwerbslos	585	0,7	562	0,7	-23	-3,89
Nichterwerbsperson	41 850	50,8	41 850	50,8	0	0,00

Fortsetzung der Tabelle auf nächster Seite

Bevölkerung am Ort der Hauptwohnung im Erhebungsjahr 2002						
Gegenstand der Nachweisung	MZ-Originalfile		Querschnitt des MZ-Panels		Differenz	
	Anzahl	Prozent	Anzahl	Prozent	absolut	relativ
nach der Stellung im Beruf in der gegenwärtigen Tätigkeit						
Selbstständig ohne Beschäftigte	1 858	2,3	1 881	2,3	23	1,23
Selbstständig mit Beschäftigten	1 796	2,2	1 804	2,2	8	0,43
Mithelfende(r) Familienangehörige(r)	414	0,5	407	0,5	-6	-1,57
Beamten/Beamtinnen, Richter(in)	1 914	2,3	1 938	2,3	24	1,25
Angestellte(r)	17 779	21,6	17 627	21,4	-153	-,86
(Heim-)Arbeiter(in)	10 806	13,1	10 899	13,2	93	0,86
Kaufm./techn. Auszubildende(r)	795	1,0	801	1,0	7	0,83
Gewerbl. Auszubildende(r)	770	0,9	746	0,9	-24	-3,09
Zeit-/Berufssoldat(in) (einschl. BGS und Bereitschaftspolizei)	218	0,3	234	0,3	16	7,34
Grundwehrdienstleistender	93	0,1	79	0,1	-14	-14,66
Zivildienstleistender	94	0,1	95	0,1	1	0,76
trifft nicht zu (Person < 15 Jahre, nicht erwerbstätig)	45 920	55,7	45 946	55,7	26	0,06
nach Voll-/Teilzeittätigkeit						
Vollzeittätigkeit	28 893	35,0	28 925	35,1	32	-0,11
Teilzeittätigkeit	7 643	9,3	7 585	9,2	-58	-0,75
trifft nicht zu (Person < 15 Jahre, nicht erwerbstätig)	45 920	55,7	45 946	55,7	26	0,06

Fortsetzung der Tabelle auf nächster Seite

Bevölkerung am Ort der Hauptwohnung im Erhebungsjahr 2002						
Gegenstand der Nachweisung	MZ-Originalfile		Querschnitt des MZ-Panels		Differenz	
	Anzahl	Prozent	Anzahl	Prozent	absolut	relativ
nach dem überwiegender Lebensunterhalt						
Erwerbs-/Berufstätigkeit ..	33 254	40,3	33 283	40,4	29	0.09
Arbeitslosengeld/-hilfe	3 112	3,8	3 155	3,8	43	1.37
Rente, Pension	18 642	22,6	18 638	22,6	-4	-0.02
Unterhalt durch Eltern, Ehemann/-frau oder andere Angehörige	24 512	29,7	24 489	29,7	-23	-0.09
eigenes Vermögen, Vermie- tung, Zinsen, Altenteil ..	367	0,4	381	0,5	14	3.83
Sozialhilfe	1 578	1,9	1 489	1,8	-88	-5.60
Leistungen aus einer Pflegeversicherung	130	0,2	123	0,1	-7	-5.77
sonstige Unterstützungen	861	1,0	898	1,1	37	4.34
nach den Individualeinkommensklassen						
1 < 150 Euro	2 037	2,5	1 983	2,4	-54	-2,64
150 < 300 Euro	3 516	4,3	3 477	4,2	-38	-1,08
300 < 500 Euro	5 526	6,7	5 482	6,6	-44	-0,80
500 < 700 Euro	6 263	7,6	6 267	7,6	4	0,06
700 < 900 Euro	5 913	7,2	6 017	7,3	104	1,76
900 < 1 100 Euro ...	6 800	8,2	6 721	8,2	-79	-1,16
1 100 < 1 300 Euro ...	6 875	8,3	6 914	8,4	40	0,57
1 300 < 1 500 Euro ...	5 686	6,9	5 676	6,9	-11	-0,19
1 500 < 1 700 Euro ...	4 307	5,2	4 379	5,3	72	1,68
1 700 < 2 000 Euro ...	3 807	4,6	3 724	4,5	-83	-2,18
2 000 < 2 300 Euro ...	2 724	3,3	2 726	3,3	2	0,08
2 300 < 2 600 Euro ...	1 994	2,4	1 980	2,4	-14	-0,69
2 600 < 2 900 Euro ...	1 015	1,2	1 004	1,2	-11	-1,08
2 900 < 3 200 Euro ...	919	1,1	918	1,1	-2	-0,17
3 200 < 3 600 Euro ...	723	0,9	723	0,9	0	-0,03
3 600 < 4 000 Euro ...	465	0,6	430	0,5	-35	-7,57
4 000 < 4 500 Euro ...	374	0,5	380	0,5	6	1,67
4 500 < 5 000 Euro ...	255	0,3	235	0,3	-20	-7,74
5 000 < 5 500 Euro ...	167	0,2	166	0,2	-1	-0,88
5 500 < 6 000 Euro ...	131	0,2	138	0,2	7	5,12
6 000 < 7 500 Euro ...	139	0,2	148	0,2	9	6,58
7 500 < 10 000 Euro	119	0,1	133	0,2	15	12,30
10 000 < 18 000 Euro	82	0,1	78	0,1	-5	-5,50
≥ 18 000 Euro	55	0,1	47	0,1	-9	-15,45
selbstständige(r) Landwirt- wirt(in) in der Haupt- tätigkeit	267	0,3	251	0,3	-16	-6,04
kein Einkommen	18 558	22,9	18 869	22,9	-17	-0,09
keine Angabe	3 410	4,1	3 589	4,4	179	5,25

Fortsetzung der Tabelle auf nächster Seite

Bevölkerung am Ort der Hauptwohnung im Erhebungsjahr 2002						
Gegenstand der Nachweisung	MZ-Originalfile		Querschnitt des MZ-Panels		Differenz	
	Anzahl	Prozent	Anzahl	Prozent	absolut	relativ
nach dem höchsten allgemeinbildenden Schulabschluss						
Volks-/Hauptschulabschluss	30 110	36,5	30 436	36,9	326	1,08
Abschluss der allgemeinbildenden POS der ehemaligen DDR	4 979	6,0	4 924	6,0	-55	-1,10
Realschul- oder gleichwertiger Abschluss	12 793	15,5	12 866	15,6	72	0,56
Fachhochschulreife	2 614	3,2	2 588	3,1	-27	-1,01
Allgemeine oder fachgebundene Hochschulreife	10 748	13,0	10 544	12,8	-203	-1,89
keine Angabe	572	0,7	617	0,7	45	7,84
trifft nicht zu (Person < 15 Jahre, kein allgemeinbildender Abschluss bzw. keine Angabe hierzu)	20 640	25,0	20 482	24,8	-158	-0,77
nach dem höchsten beruflichen Ausbildungsabschluss						
Anlernausbildung, berufliches Praktikum	1 075	1,3	1 088	1,3	13	1,20
Berufsvorbereitungsjahr	137	0,2	141	0,2	4	2,96
Abschluss einer Lehrausbildung oder gleichwertigem Berufsfachschulabschluss	30 466	36,9	30 652	37,2	206	0,68
Abschluss an einer Berufsfach- oder Kollegschule	2 043	2,5	2 218	2,6	84	4,12
Meister-/Technikerabschluss oder gleichwertiger Fachschulabschluss	3 884	4,7	3 920	4,8	36	0,92
Abschluss der Fachschule der ehemaligen DDR	1 018	1,2	1 004	1,2	-13	-1,30
Abschluss an einer Verwaltungsfachhochschule	520	0,6	531	0,6	11	2,11
Fachhochschulabschluss (auch Ingenieurschulabschluss)	2 620	3,2	2 592	3,1	-28	-1,05
Hochschulabschluss	3 616	4,4	3 532	4,3	-83	-2,30
Promotion	574	0,7	561	0,7	-13	-2,35
keine Angabe	1 085	1,3	1 074	1,3	-11	-1,02
trifft nicht zu (Person < 15 Jahre, kein beruflicher Abschluss bzw. keine Angabe hierzu)	35 440	43,0	35 234	42,7	-205	-0,58

3.6 Metainformationen zum Mikrozensus-Panel 2001–2004

Neben dem vorliegenden Handbuch und der Randverteilung ausgewählter Merkmale zum Mikrozensus-Panel 2001–2004 werden noch weitere Metainformationen bereitgestellt. Es handelt sich hierbei zum einen um die Fragebogen für Selbstausfüller der einzelnen Erhebungsjahre, das Schlüsselverzeichnis, in dem die konkrete Datensatzbeschreibung des Mikrozensus-Panels 2001–2004 festgehalten ist und um Dokumente zu den Klassifikationen der Wirtschaftszweige, der Berufe und des Bildungswesens.

Welche einzelnen Metainformationen insgesamt vorliegen, können Sie nachfolgender Tabelle entnehmen. Eine kurze Beschreibung zu den Metainformationen Schlüsselverzeichnis und Klassifikationen wird in den nachfolgenden Unterabschnitten gegeben.

Tabelle 3.7: Metainformationen zum Mikrozensus-Panel 2001–2004

Dateiname	Inhaltsbeschreibung
Randverteilungen.pdf	Randverteilungen ausgewählter Merkmale aus den Mikrozensus-Originalfiles und den Querschnitten des Mikrozensus-Panels 2001–2004 für die Erhebungsjahre 2001 bis 2004
FB01_oEP.pdf	Fragebogen für Selbstausfüller ohne Ergänzungsprogramm für das Erhebungsjahr 2001
FB01_mEP.pdf	Fragebogen für Selbstausfüller mit Ergänzungsprogramm für das Erhebungsjahr 2001
FB02_oEP.pdf	Fragebogen für Selbstausfüller ohne Ergänzungsprogramm für das Erhebungsjahr 2002
FB02_mEP.pdf	Fragebogen für Selbstausfüller mit Ergänzungsprogramm für das Erhebungsjahr 2002
FB03_oEP.pdf	Fragebogen für Selbstausfüller ohne Ergänzungsprogramm für das Erhebungsjahr 2003
FB03_mEP.pdf	Fragebogen für Selbstausfüller mit Ergänzungsprogramm für das Erhebungsjahr 2003
FB04_oEP.pdf	Fragebogen für Selbstausfüller ohne Ergänzungsprogramm für das Erhebungsjahr 2004
FB04_mEP.pdf	Fragebogen für Selbstausfüller mit Ergänzungsprogramm für das Erhebungsjahr 2004
FMZP0104_SV.pdf	Schlüsselverzeichnis des faktisch anonymisierten Mikrozensus-Panels 2001–2004 mit der konkreten Datensatzbeschreibung

Fortsetzung der Tabelle auf nächster Seite

Dateiname	Inhaltsbeschreibung
WZ93_MZ.pdf	Systematisches Verzeichnis der Wirtschaftszweige des Statistischen Bundesamtes in der Fassung für den Mikrozensus – Ausgabe 1993 – (WZ 93)
WZ2003_MZ.pdf	Systematisches Verzeichnis der Wirtschaftszweige des Statistischen Bundesamtes in der Fassung für den Mikrozensus – Ausgabe 2003 – (WZ 2003) einschließlich einer Kennzeichnung gruppenspezifischer bzw. gruppenübergreifender Veränderungen gegenüber der Wirtschaftszweigeklassifikation in der Ausgabe von 1993
KldB92_MZ.pdf	Klassifikation der Berufe des Statistischen Bundesamtes in der Fassung für den Mikrozensus – Ausgabe 1992 – (KldB 92)
ISCO88COM_MZ.pdf	Internationale Standardklassifikation der Berufe zur Verwendung innerhalb der Europäischen Gemeinschaft – Ausgabe 1988 – (ISCO-88 COM)
HFR2003_MZ.pdf	Gliederung der Hauptfachrichtung des Abschlusses an einer (Fach-)Hochschule bzw. eines beruflichen Abschlusses – Ausgabe 2003 – (HFR 2003) einschließlich einer Gegenüberstellung mit den Ausbildungsfeldern der ISCED 97

Bitte beachten Sie, dass die WZ 93 nicht den aktuellen Stand der Systematik wiedergibt, hier aber auf sie Bezug genommen wird, weil sie für den Erhebungszeitraum des Mikrozensus bis einschließlich 2002 Gültigkeit hatte. Seit dem Jahr 2003 liegt eine neue Systematik der Wirtschaftszweige (WZ2003) vor.

3.6.1 Das Schlüsselverzeichnis des faktisch anonymisierten Mikrozensus-Panels 2001–2004

Das Schlüsselverzeichnis beinhaltet die Beschreibung des faktisch anonymisierten Mikrozensus-Panels 2001–2004 hinsichtlich

- der Variablennamen,
- der Etikettierung der Variablen und deren Werteausprägungen,
- der vorgenommenen Zusammenfassungen zur Einhaltung der Anonymisierungsrichtlinien,
- der, einer Variable zugrunde liegenden Fragennummer im Fragebogen für Selbstausfüller/innen bzw. der Kennzeichnung für organisatorisch-technische oder typisierte Merkmale sowie
- der Spaltenposition und des Formats der Variablen in der ASCII-Datei famzp_0104v1.dat.

Die Variablennamen in einem Mikrozensus-Originalfile setzen sich in der Regel aus zwei Teilen zusammen. Aus einem zweistelligen Buchstabenkode (ef) gefolgt von der Spaltenposition der Variable im Datensatz des Mikrozensus-Originalfiles. In einigen wenigen Fällen, wenn der Wert einer Merkmalsausprägung sich nochmals unterteilen lässt, folgt im Variablennamen der Spaltenposition noch eine Kennzeichnung des Unterfeldes.

Um die Umbenennungen der Variablennamen im Mikrozensus-Panel gegenüber dem jeweiligen Mikrozensus-Originalfile so gering wie möglich zu halten und unter der Vorgabe, dass die Maximallänge eines Variablennamens acht Zeichen nicht überschreiten sollte, wurde von dem ursprünglichen Variablennamen in jedem Erhebungsjahr das erste Zeichen ersetzt. Bei Variablen des Erhebungsjahres 2001 beginnt der Variablenname mit einem f, bei Variablen des Erhebungsjahres 2002 beginnt der Variablenname mit einem g usw. Die erste Stelle eines Variablennamens kennzeichnet also den Erhebungszeitpunkt. Nur bei Variablen, die längsschnittspezi-fische bzw. längsschnittkonsistente Informationen beinhalten, wird von dieser Regel abgewichen.

Mit der Vergabe der neuen Variablennamen wird also an die Vergabe der Variablennamen im Mikrozensus-Panel 1996–1999 angeknüpft, wonach ein Variablenname in einem Mikrozensus-Panel für das erste Erhebungsjahr eines neuen Mikrozensusgesetzes mit dem Buchstaben a beginnt, der Variablenname des zweiten Erhebungsjahres mit dem Buchstaben b usw.

Wenn also in der Spalte "Variablenname" nur eine Ziffer steht, so muss diese stets mit einem entsprechenden Präfix ergänzt werden, um den vollständigen Variablennamen zu erhalten.

In der Spalte "Variablenname" ist zudem dokumentiert, ob eine Variable dem Ergänzungsprogramm angehört. Denn wie in Kapitel 1.3 kurz erläutert wurde, können die im Mikrozensus erhobenen Tatbestände in ein Grund-, ein Ergänzungs- und ein Zusatzprogramm untergliedert werden, wobei das Ergänzungsprogramm mit einem verringerten Auswahlatz erhoben wird. Die Zugehörigkeit einer Variable zum Ergänzungsprogramm, das einer Unterstichprobe entspricht, wird in der Spalte "Variablenname" mit UStp (Unterstichprobe) gekennzeichnet.

Die Etikettierung der Variablen und deren Wertausprägungen können Sie der zweiten Spalte des Schlüsselverzeichnisses "Variablelabel und Wertelabel in den einzelnen Querschnittserhebungen des Mikrozensus-Panels" entnehmen.

Bei der Vergabe der Variablenlabels wurde wie bei den Variablennamen eine allgemeine Kennung vergeben. Jedes Variablenlabel einer Variablen aus einem Mikrozensus-Originalfiles beginnt mit dem, in eckiger Klammer eingegrenztem Variablenname des entsprechenden Mikrozensus-Originalfiles gefolgt mit dem entsprechenden Erhebungsjahr. Danach folgt eine kurze Beschreibung des Variableninhalts. Die Beschreibung der Variable mit der Kennung des Bundeslandes aus dem Erhebungsjahres 2002, ist somit wie folgt gegeben:

[EF1] MZ02: Bundesland

Analog wurde bei der Vergabe der Wertelabels verfahren. Jede inhaltliche Beschreibung wird eingeleitet durch die in eckiger Klammer gesetzte Angabe des entsprechenden Wertes. Für das obige Beispiel, der Variablen mit der Kennung für das Bundesland, wurden demnach die Wertelabels wie folgt definiert:

Variablenwert	I	Wertelabel
1	I [1]	Schleswig-Holstein
2	I [2]	Hamburg
3	I [3]	Niedersachsen
.	I	.
.	I	.
15	I [15]	Sachsen-Anhalt
16	I [16]	Thüringen

Die Weitergabe eines faktisch anonymisierten Datenfiles mit Einzelangaben, wie dies für das Mikrozensus-Panels zutrifft, bedeutet, dass die Einzelangaben so zu anonymisieren sind, dass sie den Auskunftspflichtigen nur mit einem unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft zugeordnet werden können.

Zur Einhaltung dieser Maßgaben ist es unter anderem erforderlich, dass Merkmalsvergrößerungen vorgenommen werden müssen. Diese betreffen insbesondere die Angaben zum gegenwärtigen (früheren) Beruf sowie dem Wirtschaftszweig des Betriebs, in dem die befragte Person gegenwärtig (früher) beschäftigt ist (war).

In der Spalte "Zusammenfassungen zur Einhaltung der Anonymisierungsrichtlinien Bemerkungen" ist dokumentiert, welche einzelnen Merkmalsausprägungen zusammengelegt wurden.

Des Weiteren finden Sie in dieser Spalte Anmerkungen darüber, ob für eine Variable keine bzw. nur in eingeschränktem Umfang Wertelabels vergeben wurden.

In der Spalte "Quelle" wird für jedes Erhebungsjahr angegeben, ob die Variable aus der organisatorisch-technischen Durchführung entnommen wurde, aus einer Typisierung hervorgegangen ist oder auf einer Frage basiert. Variablen, die der organisatorisch-technischen Durchführung entnommen wurden, sind mit einem "O" gekennzeichnet, Variablen, die aus einer Typisierung hervorgegangen sind mit einem "T" und Variablen, die auf einer Frage basieren mit der Fragennummer aus dem zugehörigen Fragebogen für Selbstausfüller. Ein Minus in der Quellenangabe bedeutet, dass dieses Merkmal in dem entsprechenden Erhebungsjahr nicht ausgewiesen ist bzw. nur im Ergänzungsprogramm erhoben wurde.

Das faktisch anonymisierte Mikrozensus-Panels 2001–2004 wird der Wissenschaft in drei verschiedenen Datenformaten bereitgestellt. Als SPSS-Datenfile, als SAS-Datenfile und in einem ASCII-Format mit fester Spaltenposition der Variablenwerte.

Die Spaltenposition der Variablenwerte und deren Format sind in der letzten Spalte "Spaltenposition in der Datei: famzp_0104v1.dat" angegeben, wobei famzp_0104v1.dat der Name der ASCII-Datei ist.

Hinzuweisen ist noch auf vier weitere Aspekte:

1. Variablen, die auf Fragen basieren, deren Beantwortung für alle Personen freiwillig sind und die aus diesen Fragen abgeleiteten Variablen sind im Schlüsselverzeichnis mit "Grau-10%" hinterlegt.
2. Variablen, die nur für einen eingeschränkten Personenkreis freiwillig sind bzw. deren Beantwortung nicht in allen Erhebungsjahren freiwillig ist und die aus diesen Fragen abgeleiteten Variablen sind im Schlüsselverzeichnis mit "Grau-5%" hinterlegt.
3. Variablen, bei denen aufgrund unterschiedlicher Merkmalsausprägungen im Zeitablauf, Brüche im Einbezug eines Merkmals in das Unterstichprobenprogramm, der Neuaufnahme eines Merkmals, Zensierungen etc. vorliegen — also ein intertemporaler Bruch gegeben ist — werden durch einen schwarzen Balken an der linken Seite des betreffenden "Variablenblocks" gekennzeichnet.
4. In den Mikrozensus-Originalfiles werden fehlende Angaben aufgrund von "trifft nicht zu" (Filterführung) mit keinem speziellen Kode versehen und werden in Statistikprogrammssystemen als "leer" oder "system missing" ausgewiesen. Im Mikrozensus-Paneldatensatz 2001–2004 dagegen wurden alle Variablen, die einem Ausfall aufweisen mit einem – je nach Ausfallart – generellen Wert versehen. Ausfallart und zugehöriger Variablenwert können Sie nachfolgender Tabelle entnehmen.

Tabelle 3.8: Ausfallart sowie zugehöriger Variablenwert und Wertelabel im Mikrozensus-Paneldatensatz 2001–2004

Ausfallart	Variablenwert	Wertelabel
Person/Haushalt wurde nicht befragt, weil die Befragungseinheit(en) noch nicht bzw. nicht mehr Element der Stichprobe ist (sind) bzw. der Haushalt ausgefallen ist.	-1	[-1] tnz: Pers/HH ist nicht Element der Stp.; HH ausgefallen
Haushalt ist nicht Element der Unterstichprobe	-2	[-2] tnz: HH ist nicht Element der UStp
Trifft nicht zu aufgrund der Filterführung	-3	[-3] tnz: aufgrund der Filterführung
Trifft nicht zu, weil Person in einer Gemeinschaftsunterkunft wohnt	-4	[-4] tnz: Gemeinschaftsunterkunft

Bitte beachten Sie, dass die Werte in den Datendateien nicht als missing Werte deklariert wurden.

Weiterhin ist zu beachten, dass es aufgrund der Umsetzungen der "entfällt"-Kategorie zu einer Zunahme in der Kategorie "keine Angabe" kommen kann. So werden beispielsweise in der Originalvariable EF40 (Verwandtschaftliche Beziehung zum/zur Lebenspartner/in der ersten Person) nur Personen in der Kategorie "keine Angabe" ausgewiesen, die hierzu explizit keine Angaben machten. Verneint hingegen eine Person die Frage nach dem verwandtschaftlichen Verhältnis zu ersten Person (EF37: Mit der ersten Person verheiratet, verwandt, verschwägert = 8 (Nein)) und verneint die Frage, ob Sie Lebenspartner/in der ersten Person ist (EF39: Lebenspartner/in der ersten Person = 8 (Nein)), führt diese Datenkonstellation zumeist in eine "entfällt"-Kategorie in der Variable EF40. Infolge der Umsetzungen der "entfällt"-Kategorien wird die zuvor beschriebene Datenkonstellation umgesetzt in "keine Angabe".

3.6.2 Das systematische Verzeichnis der Wirtschaftszweige des Statistischen Bundesamtes in der Fassung für den Mikrozensus – Ausgabe 1993

Das systematische Verzeichnis der Wirtschaftszweige des Statistischen Bundesamtes in der Ausgabe von 1993 (WZ 93) dient dazu, die wirtschaftlichen Tätigkeiten von Unternehmen und Betrieben in einheitlicher Weise zu erfassen und darzustellen. Das Klassifikationssystem ist hierarchisch aufgebaut und unterscheidet sechs Gliederungsebenen:

Tabelle 3.9: Die Gliederungsebenen der Klassifikation der Wirtschaftszweige des Statistischen Bundesamtes in der Ausgabe von 1993

Gliederungsebene	Anzahl der Einheiten	Schlüssel
Abschnitte	17	A – Q
Unterabschnitte	31	AA – QA
Abteilungen	60	01 – 99
Gruppen	222	01.1 – 99.0
Klassen	503	01.11 – 99.00
Unterklassen	1 062	01.11.1 – 99.00.0

Bis zur fünften Gliederungsebene, den Klassen, ist die WZ 93 identisch mit der Wirtschaftszweigeklassifikation der Europäischen Union in der ersten Revision (NACE Rev. 1; NACE = Nomenclature statistique de activités économiques dans la Communauté Européenne). Die sechste Gliederungsebene wurde in der WZ 93 für nationale Belange neu geschaffen. Darüber hinaus wurden für Analysezwecke die Wirtschaftsabschnitte nochmals zu so genannten Bereichen und Unterbereichen

aggregiert. Unterschieden wird hierbei zwischen vier Bereichen und zehn Unterbereichen, denen jedoch kein eigener Schlüssel zugewiesen wurde.

In der Fassung für den Mikrozensus wird die WZ 93 nicht in ihrem vollen Umfang genutzt, sondern nur bis zur Gliederungsebene der Gruppen. Weiterhin ist darauf hinzuweisen, dass

- die Schlüssel ohne abgrenzende Punkte vergeben wurden, so dass ein numerischer Kode vorliegt und
- nur die Gliederungsebene der Gruppen verkodet vorliegt. Soll also eine Auswertung für eine übergeordnete Hierarchieebene vorgenommen werden, so muss diese generiert werden.

3.6.3 Das systematische Verzeichnis der Wirtschaftszweige des Statistischen Bundesamtes in der Fassung für den Mikrozensus – Ausgabe 2003

Das systematische Verzeichnis der Wirtschaftszweige in der Ausgabe von 2003 ist *"das Ergebnis einer behutsamen Aktualisierung der Klassifikation der Wirtschaftszweige, Ausgabe 1993 (WZ 93), deren Struktur weitgehend beibehalten werden konnte. Nur dort, wo sich seit Einführung der WZ 93 gravierende technische und wirtschaftliche Veränderungen gezeigt haben, z.B. im Zusammenhang mit einer fortschreitenden Deregulierung auf verschiedenen Märkten, waren Modifikationen der Gliederung erforderlich. Außerdem wurde die Gelegenheit genutzt, die inhaltliche Beschreibung der Unterteilungen der Klassifikation der Wirtschaftszweige zu erweitern und zu aktualisieren"* (Statistisches Bundesamt 2003: 3).

Ausgehend von der aktualisierten EU-Wirtschaftszweigklassifikation NACE in der Revision 1.1, die ihrerseits aus der ISIC in der Revision 3.1 abgeleitet ist, wurde die Klassifikation der Wirtschaftszweige für nationale Zwecke wie folgt entwickelt:

- Die erste Ebene (Abschnitte, die mit einem Buchstabenkode (A-Q) gekennzeichnet sind) der ISIC Rev. 3.1 wird in die NACE Rev. 1.1 unverändert übernommen.
- In einigen Bereichen wurden die Abschnitte zusätzlich in Unterabschnitte untergliedert und mit einem zweistelligen Buchstabenkode versehen.
- Die zweite Ebene (Abteilungen, die mit einem Ziffernkode (01-99) gekennzeichnet sind) der ISIC Rev. 3.1 wird in die NACE Rev. 1.1 unverändert übernommen.
- Die dritte und vierte Ebene (Gruppen und Klassen) der ISIC Rev. 3.1 wird in der NACE Rev. 1.1 den europäischen Erfordernissen entsprechend unterteilt,

können jedoch stets zu den Gruppen und Klassen aus denen sie abgeleitet wurden, aggregiert werden. Die Unterteilung ist mit einem x hinter dem jeweiligen Kode der ISIC Rev. 3.1 gekennzeichnet.

- Um die Unterschiede zwischen der ISIC Rev. 3.1 und NACE Rev. 1.1 auch in der Kodierung deutlich zu machen, wurde bei der NACE Rev. 1.1 zwischen die zweite und dritte Stelle des Nummernkodes ein Punkt gesetzt.
- Die Ziffer "9" bedeutet in der ISIC Rev. 3.1 immer "Sonstige", wohingegen sie in der NACE Rev. 1.1 wie die übrigen Ziffern verwendet wird.
- Mit drei Ausnahmen wurde die Kodierung der NACE Rev. 1.1 in die WZ 2003 unverändert übernommen. Zwei Klassen der NACE Rev. 1.1 sind in der WZ 2003 abweichend kodiert bzw. weiter aufgliedert worden:

Tabelle 3.10: Differenzen zwischen der Kodierung und Aufgliederung von NACE Rev. 1.1 und WZ 2003

Klasse der NACE Rev. 1.1		Klasse der WZ 2003	
52.48	Sonstiger Facheinzelhandel a.n.g. (in Verkaufsräumen)	52.48	Einzelhandel mit Tapeten, Bodenbelägen, Kunstgegenständen, Briefmarken, Münzen, Geschenkartikeln, Uhren, Schmuck und Spielwaren
		52.49	Facheinzelhandel a.n.g. (in Verkaufsräumen)
64.20	Fernmeldedienste	64.30	Fernmeldedienste

Mit dieser Maßnahme sollen zusätzliche Möglichkeiten zur Einführung nationaler Unterteilungen im Bereich des Einzelhandels geschaffen und außerdem im Interesse größtmöglicher Klarheit in diesen wichtigen Bereichen der Wirtschaft die Wiederverwendung von Codes der WZ 93 mit geänderten Inhalten in der WZ 2003 vermieden werden.

- Um für bestimmte wirtschaftsanalytische Fragestellungen auch die Produktivitätstätigkeiten von privaten Haushalten für den Eigenbedarf erfassen zu können, wurden in der ISIC Rev. 3.1 zwei Abteilungen geschaffen, die in die NACE Rev. 1.1 übernommen wurden:
 - 96 Herstellung von Waren durch private Haushalte für den Eigenbedarf ohne ausgeprägten Schwerpunkt
 - 97 Erbringung von Dienstleistungen durch private Haushalte für den Eigenbedarf ohne ausgeprägten Schwerpunkt

Für nationale Zwecke besteht an diesen Abteilungen kein Bedarf. Deshalb wurden sie – mit Billigung von Eurostat – in der WZ 2003 nicht berücksichtigt. Entsprechendes gilt für die beiden Gruppen mit gleichem Inhalt. Die beiden Gruppen wurden ohne Kodierung in der WZ 2003 übernommen und zudem mit roter Schrift hervorgehoben.

- Die WZ 2003 stimmt zwar mit Ausnahme der vorstehend angeführten Abweichungen inhaltlich mit der NACE Rev. 1.1 überein, in der WZ 2003 weichen aber die Bezeichnungen der Wirtschaftszweige teilweise von den entsprechenden Bezeichnungen der NACE Rev. 1.1 ab. Dadurch soll den Anwender(n)/innen der Klassifikation auf nationaler Ebene der gewohnte, übliche Sprachgebrauch geboten werden. Abweichende Bezeichnungen in der NACE Rev. 1.1 gegenüber der WZ 2003 sind im Klassifikationsverzeichnis den Bezeichnungen in der WZ 2003 nachgeordnet und in kursiv-blauer Schrift hervorgehoben.
- Um die nationalen Bedürfnisse zu berücksichtigen, wurden die Klassen zusätzlich in Unterklassen untergliedert.
Zur formalen Unterscheidung erfolgte wiederum eine Trennung zwischen der vierten und fünften Stelle des Nummernkodes durch einen Punkt:

Tabelle 3.11: Die Gliederungsebenen der Klassifikation der Wirtschaftszweige von ISIC Rev. 3.1, NACE Rev. 1.1 und WZ 2003

Gliederungsebene	ISIC Rev. 3.1		NACE Rev. 1.1		WZ 2003	
	Anzahl	Kode	Anzahl	Kode	Anzahl	Kode
Abschnitte	17	A – Q	17	A – Q	17	A – Q
Unterabschnitte	—		31	AA – QA	31	AA – QA
Abteilungen	62	01 – 99	62	01 – 99	60	01 – 99
Gruppen	161	011 – 990	224	01.1 – 99.0	222	01.1 – 99.0
Klassen	298	0111 – 9900	515	01.11 – 99.00	513	01.11 – 99.00
Unterklassen	—		—		1 041	01.11.1 – 99.00.0

In der Fassung für den Mikrozensus wird die Klassifikation der Wirtschaftszweige des Statistischen Bundesamtes in der Ausgabe von 2003 (WZ 2003) nicht in ihrem vollem Umfang genutzt, sondern nur bis zur Gliederungsebene der Gruppen. Diese Gliederung ist in der Datei WZ2003_MZ.pdf wiedergegeben, wobei eine zusätzliche Kennzeichnung bezüglich der Änderungen gegenüber der Klassifikation der Wirtschaftszweige in der Ausgabe von 1993 eingearbeitet wurde, sofern sie die Gliederungsebene der Gruppen betreffen. Die Änderungen und zugehörige Kennzeichnungen sind nachstehender Tabelle zu entnehmen:

Tabelle 3.12: Art der Änderungen und zugehörige Kennzeichnungen zwischen den Wirtschaftszweckklassifikationen WZ 2003 und WZ 93

Kennzeichnung	Art der Änderung der WZ 2003 gegenüber der WZ 93
B	Änderung in der Bezeichnung
K	Änderung in der Zuweisung der Kennziffer
Z	Gruppenübergreifende Neuordnung einer wirtschaftlichen Tätigkeit aus der Unterklasse
A	Aufteilung einer bestehenden Gruppe

3.6.4 Klassifikationen beruflicher Tätigkeiten

Zur Ordnung von Berufsinformationen und darauf aufbauend zur Beschreibung von Struktur und Verteilung beruflicher Tätigkeiten liegen im Mikrozensus zwei Klassifikationssysteme vor. Zum einen die nationale Klassifizierung der Berufe des Statistischen Bundesamtes in der Ausgabe von 1992 (KldB 92) und zum anderen die Internationale Standardklassifikation der Berufe zur Verwendung innerhalb der Europäischen Gemeinschaft (ISCO-88 COM). Die KldB 92 umfasst fünf hierarchisch geordnete Ebenen:

Tabelle 3.13: Die Gliederungsebenen der Klassifikation der Berufe des Statistischen Bundesamtes in der Ausgabe von 1992

Gliederungsebene	Anzahl der Einheiten	Schlüssel
Berufsbereiche	6	I – VI
Berufsabschnitte	33	Ia – VIa
Berufsgruppen	88	01 – 99
Berufsordnungen	639	011 – 997
Berufsklassen	2 287	0110 – 9971

Auf der obersten Ebene befinden sich sechs Berufsbereiche, deren Einteilung und Anordnung *"weitgehend der traditionellen Unterscheidung in Urproduktions-, Fertigungs- und Dienstleistungsberufe"* folgt (Statistisches Bundesamt, 1992: 13).

Auf der nächsten Ebene werden 33 Berufsabschnitte unterschieden. *"Sie umfassen jeweils solche Berufe, die sich im Wesen der Berufsaufgabe, in der Berufstätigkeit, in der Art des verarbeiteten Materials oder in anderer Hinsicht ähneln"* (Statistisches Bundesamt, 1992: 13).

Es folgt die Ebene der Berufsgruppen. Sie *"fassen die fachlich näher zueinander gehörenden, dem Wesen ihrer Berufsaufgabe und Tätigkeit nach verwandten Berufe zusammen, wobei in den Fertigungsberufen vielfach das verwendete Material als Leitprinzip dient, weil es die Art der Berufstätigkeit entscheidend prägt"* (Statistisches Bundesamt, 1992: 13).

Die auf der Ebene der Berufsordnungen zusammengefassten Berufe sind nach dem Wesen ihrer Berufsaufgabe und Tätigkeit als gleichartig anzusehen.

Die einzelnen Berufe oder Berufsarten werden schließlich in den Berufsklassen ausgewiesen.

Die ISCO-88 COM ist ebenfalls hierarchisch aufgebaut, umfasst jedoch nur vier Gliederungsebenen:

Tabelle 3.14: Die Gliederungsebenen der Internationalen Standardklassifikation der Berufe zur Verwendung innerhalb der Europäischen Gemeinschaft in der Ausgabe von 1988

Gliederungsebene	Anzahl der Einheiten	Schlüssel
Berufshauptgruppen	10	1 – 0
Berufsgruppen	28	11 – 01
Berufsuntergruppen	116	111 – 011
Berufsgattungen	390	1110 – 0110

Beiden Klassifikationssystemen – der Berufsklassifikation des Statistischen Bundesamtes (KlDB 92) und der Internationalen Standardklassifikation der Berufe (ISCO-88 COM) – ist gemeinsam, dass sie eine hierarchische Gliederung beruflicher Tätigkeiten vornehmen, sie unterscheiden sich aber im Hinblick auf die Berücksichtigung des Anforderungsniveaus, das mit einer beruflichen Tätigkeit verbunden ist. Während bei der KlDB 92 das ausschlaggebende Kriterium der Systematik die Art der ausgeübten Tätigkeit ist und Abhängigkeitsverhältnisse, Status oder Rangordnung innerhalb des Beschäftigungsbetriebes keine Bestimmungsmerkmale der Klassifikation sind, werden diese bei dem Ordnungssystem der ISCO-88 COM explizit berücksichtigt. Dazu wird das Konzept des *skill* (der Fertigkeiten) eingeführt und zwischen *skill level* und *skill specialisation* unterschieden. Das *skill level* definiert den Grad der Komplexität einer Tätigkeit im Sinne eines Anforderungsniveaus und die *skill specialisation* beschreibt innerhalb durch das *skill level* bestimmte Qualifikationsniveau das "Gebiet, auf dem Kenntnisse erforderlich sind, damit die jeweiligen Aufgaben kompetent ausgeführt werden können" (Elias und Birch 1993: 11).

Bei den *skill levels* werden in Anlehnung an das System der Internationalen Standardklassifikation des Bildungswesens (ISCED)⁴ vier Anforderungsstufen unterschieden, die die Grundstruktur der ISCO-88 COM auf ihrer obersten Ebene – der Berufshauptgruppen – weitgehend festlegen⁵. Zu beachten ist hierbei, dass

⁴ISCED = International Standard Classification of Education.

⁵Zwei der zehn Berufshauptgruppen weisen keine Verbindung zu den *skill levels* auf. Es sind dies die Berufshauptgruppen 1 (Angehörige gesetzgebender Körperschaften, leitende Verwaltungsbedienstete und Führungskräfte in der Privatwirtschaft) und 0 (Soldaten), bei denen andere Aspekte der Tätigkeitsart für wichtiger erachtet werden als deren Ähnlichkeitskriterien hinsichtlich der mit ihnen verbundenen Aufgaben und Pflichten; nämlich Leitungsfunktionen im politischen und wirtschaftlichen Bereich bzw. militärische Aufgaben (vgl. Elias und Birch 1993: 15f.)

trotz Bezug zur ISCED die *skill levels* nicht durch eine formale Ausbildung erworben sein müssen. Sie können auch "durch Berufserfahrung und informelle berufliche Ausbildung erworben werden" (Elias und Birch 1993: 14).

Bitte beachten Sie, dass in der Fassung für den Mikrozensus die KldB 92 nur bis zur Gliederungsebene der Berufsordnungen und die ISCO-88 COM nur bis zur Gliederungsebene der Berufsuntergruppen ausgewiesen werden. Sollen Auswertungen für andere Gliederungsebenen vorgenommen werden, müssen diese zuvor selbst generiert werden.

Weiterhin ist zu beachten, dass aufgrund der unterschiedlich ausgerichteten Systematiken ein jeweils spezifisches Abfragesystem zur Einordnung einer beruflichen Tätigkeit notwendig ist. Im Rahmen des Mikrozensus werden die beruflichen Tätigkeiten jedoch nur auf die KldB 92 ausgerichtet erhoben. Die Zuordnung in die Systematik der ISCO 88 (COM) wird aus der KldB 92 abgeleitet⁶. Bei einigen Berufsangaben ist jedoch keine eindeutige Ableitung möglich. D. h. eine Berufsangabe aus der KldB 92 lässt sich zwei verschiedene Berufsklassen in der ISCO 88 (COM) zuordnen. In diesen Fällen werden neben der Berufsangabe zusätzliche Informationen benötigt. Es sind dies das Abhängigkeitsverhältnis (Selbstständige versus abhängig Beschäftigte) sowie die Betriebsgröße (Betriebe mit bis zu 10 Beschäftigten versus Betriebe mit 11 oder mehr Beschäftigten). Da die Betriebsgröße nur für die gegenwärtig ausgeübte Tätigkeit, nicht aber für die frühere Tätigkeit erhoben wird, kann für die frühere Tätigkeit in den Fällen keine Zuordnung von der KldB 92 auf die ISCO 88 (COM) erfolgen, in denen die Betriebsgröße als Zusatzinformation benötigt wird.

Um in diesen Fällen dennoch eine Zuordnung vornehmen zu können, könnte der strukturelle Zusammenhang zwischen Betriebsgröße und Berufsklasse durch ein stochastisches Verfahren der Umschlüsselung auf die frühere Tätigkeit übertragen werden. Hierzu könnte die jeweils gegenwärtige Struktur zwischen Betriebsgröße und den betreffenden Berufsklassen des Jahres, in dem die befragte Person ihre frühere Tätigkeit beendet hat, auf die frühere Tätigkeit übertragen werden. Aufgrund der geringen Fallzahlen in den betreffenden Berufsklassen und der Tatsache, dass die für den Beobachtungszeitraum geltenden Berufsklassifikation erst 1992 eingesetzt wurde, kann ein solches Verfahren nicht eingesetzt werden. Wird jedoch angenommen, dass die strukturellen Zusammenhänge zwischen Betriebsgröße und Berufsklasse weitestgehend stabil sind, können zeitnahe Rekodierungsmuster auch auf länger zurückliegende Rekodierungsmuster übertragen werden. Für die Umkodierungen in der früheren Tätigkeit wurden in den Fällen, in denen die Betriebsgröße benötigt wird ein stochastisches Verfahren der Umschlüsselung angewendet. Dazu wurden auf der Basis der kumulierten Datensätze von 1992 bis t_i (t_i = Jahr, in der eine Umschlüsselung vorgenommen wird) die Anteilwerte der Berufe in der ISCO-88 (COM) Klassifikation nach der Umschlüsselung ermittelt, bei denen die Betriebsgröße als Zusatzinformation herangezogen wurde. Des Weiteren wurde jedem Per-

⁶Zur Qualität einer solchen Ableitung siehe Geis und Hoffmeyer-Zlotnik 2001.

sonensatz eine gleichverteilte Zufallszahl z_i im Intervall $(0, 1)$ zugeordnet. Weist nun in einem Umschlüsselverfahren zur früheren Tätigkeit, bei dem die Betriebsgröße als Zusatzinformation benötigt wird, der Zufallswert z_i einen Wert auf, der kleiner-gleich dem Anteilwert von B_{a1} ist (B_{a1} = Anteilwert der ISCO-88 (COM) Berufsklasse im kumulierten Datensatz für dessen Umstieg von der KldB 92 auf die ISCO-88 (COM) eine Betriebsgröße von bis zu 10 Beschäftigten gegeben ist), wird der Person in ihrer früheren Tätigkeit die ISCO-88 (COM) Berufsklasse zugeordnet, die eine Betriebsgröße von bis zu 10 Beschäftigten vorausgesetzt. Ist $z_i > B_{a1}$ wird die ISCO-88 (COM) Berufsklasse zugeordnet, die eine Betriebsgröße von 11 oder mehr Beschäftigten vorausgesetzt.

Das Vorgehen soll an einem Beispiel demonstriert werden. Die Berufsklasse 0311 (Verwalter/innen in der Landwirtschaft) der KldB 92 wird bei einer Betriebsgröße von bis zu 10 Beschäftigten in der ISCO-88 (COM) der Berufsuntergruppe 131 (Leiter/innen kleiner Unternehmen) und bei einer Betriebsgröße von 11 oder mehr Beschäftigten der Berufsuntergruppe 121 (Produktions- und Operationsleiter/innen) zugeordnet. Aus dem kumulierten Datensatz wurden für die Umverschlüsselungen von der KldB 92 auf die ISCO-88 (COM) für die betreffenden Berufsuntergruppen in der gegenwärtigen Tätigkeit Anteilwerte von 0.341 (Berufsuntergruppe 121) bzw. 0.659 (Berufsuntergruppe 131) ermittelt. Ist für eine Person für Ihre frühere Tätigkeit nach der KldB 92 die Berufsklasse 0311 gegeben und weist der Zufallswert z_i einen Wert von 0.881 auf, wird der Person die Berufsuntergruppe 131 der ISCO-88 (COM) zugeordnet.

3.6.5 Gliederung der Hauptfachrichtung des Abschlusses an einer (Fach-)Hochschule bzw. eines beruflichen Abschlusses – Ausgabe 2003

Für eine systematische Erfassung und Darstellung der Hauptfachrichtung des höchsten Abschlusses an einer Fachhochschule oder Hochschule bzw. eines beruflichen Abschlusses im Mikrozensus, liegt in der Ausgabe von 2003 ein hierarchisch, zweigliedriges Ordnungssystem vor, das auf der untersten, der zweiten Gliederungsebene 98 Fachrichtungen ausweist und auf der ersten Gliederungsebene neun breite Felder. Bitte beachten Sie, dass im Mikrozensus nur die Gliederung auf der zweiten Ebene ausgewiesen ist. Sollen Auswertungen für die erste Ebene vorgenommen werden, muss diese erst generiert werden.

Weiterhin ist zu beachten, dass das Gliederungsschema auch zur Klassifikation der Inhalte von Lehrveranstaltungen der allgemeinen und beruflichen Weiterbildung Anwendung findet.

3.7 Weiterführende Literatur

- Bihler, W. (2005): *Verfahren der Unterstichprobenziehung für das faktisch anonymisierte Längsschnittfile 1996-1999*. Wiesbaden: Statistisches Bundesamt, Gruppe II A (unveröffentlichtes Manuskript).
- Elias, P.; Birch, M. (1993): *ISCO 88 (COM). Fassung der Internationalen Standardklassifikation der Berufe 1988 zur Verwendung innerhalb der Europäischen Gemeinschaft*. Institut für Beschäftigungsforschung: Universität Warwick.
- Geis, A. J.; Hoffmeyer-Zlotnik, J. H. P. (2001): *Kompatibilität von ISCO-68, ISCO-88 und KldB-92*, in: ZUMA-Nachrichten Nr. 48, S. 117-138. URL: http://www.geis.org/fileadmin/upload/forschung/publikationen/zeitschriften/zuma_nachrichten/zn_48.pdf (letzter Zugriff: 08. November 2008)
- Heidenreich, H.-J. (1994): *Hochrechnung des Mikrozensus ab 1990*; S. 112-123, in: Gabler, S.; Hoffmeyer-Zlotnik, J.; Krebs, D. (Hrsg.): *Gewichtung in der Umfragepraxis*. Opladen: Westdeutscher Verlag.
- Schimpl-Neimanns, B. (2006): *Filekonzept zum Mikrozensuspanel*. Arbeitspapier #12, Methodenverbund "Aufbereitung und Bereitstellung des Mikrozensus als Panelstichprobe". URL: <http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Content/Wissenschaftsforum/MethodenVerfahren/Mikrozensus/Arbeitspapiere/Arbeitspapier12,property=file.pdf> (letzter Zugriff: 23. März 2008)
- Statistisches Bundesamt (1992): *Klassifizierung der Berufe – Systematisches und alphabetisches Verzeichnis der Berufsbenennungen*. Ausgabe 1992, Stuttgart: Metzler-Poeschel.
- Statistisches Bundesamt (1995): *Klassifikation der Wirtschaftszweige mit Erläuterungen*. Ausgabe 1993, Stuttgart: Metzler-Poeschel (3. unveränderter Nachdruck).
- Statistisches Bundesamt (2003): *Klassifikation der Wirtschaftszweige mit Erläuterungen*. Ausgabe 2003, Wiesbaden: Statistisches Bundesamt.
- Statistisches Bundesamt (2004): *Arbeitsunterlagen zum Mikrozensus. Das Erhebungsprogramm des Mikrozensus seit 1957*. Wiesbaden, Zweigstelle Bonn.
- Wirth, H. (2006): *Anonymisierung des Mikrozensuspanels im Kontext der Bereitstellung als Scientific-Use-File*. Arbeitspapier#11, Methodenverbund "Aufbereitung und Bereitstellung des Mikrozensus als Panelstichprobe". URL: <http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Content/Wissenschaftsforum/MethodenVerfahren/Mikrozensus/Arbeitspapiere/Arbeitspapier11,property=file.pdf> (letzter Zugriff: 23. März 2008)

Kapitel 4

Das Hochrechnungsverfahren für Längsschnittauswertungen im Mikrozensus-Panel 2001–2004

Edin Basic (Freie Universität Berlin¹)

Ivo Marek (Freie Universität Berlin²)

Sandra Rohloff (Statistisches Bundesamt, Gruppe II A³)

4.1 Die Grundidee der Hochrechnung

Ziel der Hochrechnung des Mikrozensus-Panels ist es, Aussagen über Veränderungen in den zu analysierenden Variablen von der Stichprobe auf die Gesamtbevölkerung Deutschlands zwischen 2001 und 2004 übertragen zu können. Die resultierende Schätzfunktion soll – neben einer möglichst kleinen Varianz – in erster Linie einen möglichst kleinen systematischen Fehler (Bias) aufweisen. Ein Bias kann – zusätzlich zu den Nonresponse-Fällen im Querschnitt, welche im Mikrozensus keine bedeutsame Rolle spielen – auch durch Ausfälle im Längsschnitt aufgrund von Mobilität verursacht werden, da fortziehende Haushalte und Personen nicht weiterverfolgt werden. Diese Selektion in einem Mikrozensus-Panel stellt, sofern der Ausfall infolge räumlicher Mobilität mit keinem zu analysierenden Merkmal in Verbindung steht, kein großes Problem dar. Bei Ausfällen aufgrund räumlicher Mobilität ist dies in der Regel eine abzulehnende Annahme. Wenn zum Beispiel ein Wohnortwechsel mit der Aufnahme einer Tätigkeit nach einer Zeit der Erwerbslosigkeit einhergeht,

¹Mobilitätsgewichtung.

²Mobilitätsgewichtung.

³Hochrechnung.

würde das Merkmal "Aufnahme einer Tätigkeit nach Erwerbslosigkeit" auf Basis einer Auswertung nur der immobilen Personen unterschätzt.

Eine Alternative, Verzerrungen infolge räumlicher Mobilität zu beheben, ist der Einsatz einer so genannten gebundenen Hochrechnung. Hierbei erfolgt nicht nur eine Gewichtung mit dem Kehrwert der Auswahlwahrscheinlichkeit, sondern es wird zusätzlich eine nachträgliche Anpassung ausgewählter Merkmalsverteilungen der Stichprobe an die entsprechenden Verteilungen in der Grundgesamtheit vorgenommen. Die Annahme hierbei ist, dass mit der nachträglichen Anpassung an Verteilungen der Grundgesamtheit auch andere, durch die räumlich Mobilität bedingte Verteilungsabweichungen zwischen Stichprobe Grundgesamtheit korrigiert werden.

Die gebundene Hochrechnung verwendet also zusätzliche Informationen über die Beziehung zwischen Hilfs- und Untersuchungsvariablen. Je nach Art der unterstellten Beziehung zwischen den Hilfs- und Untersuchungsvariablen kann zur Gewichtung eine Differenz-, eine Verhältnis- oder eine allgemeine Regressionsschätzung herangezogen werden. Deville und Särndal (1992) haben diese verschiedenen Schätzverfahren unter Berücksichtigung von vor der Auswahl feststehender Designgewichte bzw. unter Berücksichtigung von Poststratifizierungsgewichte zu einem allgemeinen Gewichtungsansatz unter dem Titel "calibration estimation" zusammengefasst. Die zentrale Annahme hierbei ist, dass die Ausfälle innerhalb der durch die Hilfsvariablen definierten Teilgruppen von den Untersuchungsvariablen unabhängig sind.

Aus der Anwendung ausgewählter Fragestellungen im Rahmen des Verbundprojektes ist zu empfehlen, die Ausfallannahmen zu überprüfen und gegebenenfalls die Zusammenhänge zwischen dem Ausfall und den Untersuchungsvariablen explizit zu modellieren. Besteht nämlich innerhalb der durch die Hilfsvariablen definierten Teilgruppen ein Zusammenhang zwischen Ausfallmechanismus und den Untersuchungsvariablen, führt die Gewichtung zu keiner wirksamen Korrektur von Selektionsprozessen.

Die für die Kalibrierung herangezogenen Hilfsvariablen müssen also für die Grundgesamtheit bekannt sein und in ausreichender Qualität in der Grundgesamtheit als auch in der Stichprobe gemessen werden können und sie sollten sich bei der Auswahl im Hinblick auf die durch die Gewichtung vorgenommene Korrektur gemäß Lundström und Särndal (2002) an folgenden Prinzipien orientieren:

1. die Hilfsvariablen erklären unterschiedliche Antwortwahrscheinlichkeiten,
2. die Hilfsvariablen erklären die Variabilität der wichtigsten Untersuchungsvariablen und
3. die Hilfsvariablen identifizieren die wichtigsten Untersuchungsgebiete bzw. Teilpopulationen.

Zur Behandlung des Selektionsbias werden drei Ansätze verfolgt. Der erste schätzt die Wahrscheinlichkeit, am Wohnort zu verbleiben durch eine verallgemeinerte

Schätzgleichungen⁴ wie sie von Liang und Zeger (1986) und Zeger und Liang (1986) zur Analyse korrelierter Daten eingeführt wurden. Er wird im nächsten Kapitel vorgestellt. Der zweite Ansatz verfolgt eine Kalibrierung unter Verwendung von bekannten Gesamtwerten der Grundgesamtheit, während in einem dritten Ansatz die beiden ersten Ansätze kombiniert werden, indem die geschätzte Wahrscheinlichkeit für Immobilität im Eingangsgewicht für die Kalibrierung verwendet wird. Der zweite und dritte Ansatz sind im Kapitel 4.4 näher ausgeführt. Für jeden dieser Ansätze steht im Scientific Use File ein Gewichtungsfaktor zur Verfügung.

4.2 Verallgemeinerte Schätzgleichungen

Das Mikrozensus-Panel enthält räumlich mobile sowie immobile Personen. Mobile Personen sind dadurch gekennzeichnet, daß nach Auszug aus einem Befragungshaushalt bzw. vor Einzug in einen Befragungshaushalt Informationen über diese Personengruppe bis zum Zeitpunkt des Umzuges nicht vorhanden sind. Somit sind für Längsschnittauswertungen für einen vorgegebenen Zeitraum nur Personen auswertbar, die in diesem Zeitraum ihren Wohnort nicht gewechselt haben, also von Beginn bis Ende des Zeitraums in einem Befragungshaushalt leben. Eine Auswertung der immobilen Personen ohne Berücksichtigung der Mobilität kann zu systematischen Abweichungen vom wahren Wert einer Kenngröße führen, wenn sich die Immobilen von den Mobilien in Bezug auf ein Untersuchungsmerkmal systematisch voneinander unterscheiden. Ein solcher systematischer Fehler wird als Selektionsbias bezeichnet.

Um unverzerrte, d.h. erwartungstreue und konsistente Schätzwerte für Längsschnittmerkmale zu erhalten, ist es notwendig die Design-Gewichte wegen den Ausfällen zu modifizieren. Angenommen, die Wahrscheinlichkeit am Wohnort zu verbleiben ist für jede Person bekannt. Nach Multiplikation des vorhandenen Design-Gewichts mit der inversen Wahrscheinlichkeit für Immobilität erhält man ein modifiziertes Design-Gewicht. Die Verwendung dieses modifizierten Design-Gewichts bei Längsschnittauswertungen auf Basis der immobilen Personen führt dann, unter der Annahme, dass die fehlenden Angaben aufgrund räumlicher Mobilität in keinem Zusammenhang mit unbeobachteten Merkmalen stehen, zu unverzerrten Schätzwerten, da in diesem Fall für die unbekannte Population der mobilen Personen kompensiert wird. Jedoch sind in der Realität die Wahrscheinlichkeiten für Immobilität nicht bekannt. Allerdings lassen sich diese Wahrscheinlichkeiten durch ein geeignetes Modell – hier verallgemeinerte Schätzgleichungen – für Immobilität berechnen.

Verallgemeinerte Schätzgleichungen sind Regressionsmodelle, die auf einer Erweiterung generalisierter linearer Modelle mit korrelierten Daten basieren. Sie finden damit in Situationen Anwendung, in denen Beobachtungen einer Zielvariable mehrfach auftreten und die Messungen korreliert sind.

⁴Im englischen als General Estimating Equations bezeichnet.

Bei den generalisierten linearen Modellen wird im Gegensatz zum klassischen linearen Modell keine Normalverteilung für die Verbindungsfunktion zwischen der Zielvariable und den Kovariablen unterstellt, sondern es kann eine beliebige Verteilung aus der Klasse der exponentiellen Familien eingesetzt werden (zur konkreten Spezifikation und Erläuterung des Modells siehe Basic 2008). Bei der Modellierung des Modells wurden folgende Merkmale einbezogen, wobei die Merkmalsauswahl auf den Ergebnissen bisheriger Untersuchungen zum Mobilitätsverhalten basiert.

Tabelle 4.1: Kovariablen in den verallgemeinerten Schätzgleichungen zur Ermittlung der Bleibewahrscheinlichkeiten im Mikrozensus-Panel 2001–2004

Variable	Merkmalsausprägung
Altersgruppe	17–25 Jahre 26–44 Jahre 45–64 Jahre 65 Jahre oder älter
Geschlecht	männlich weiblich
Familienstand	verheiratet nicht verheiratet
Bildungsniveau	noch in Ausbildung primäres Bildungsniveau tertiäres Bildungsniveau
Nationalität	deutsch nicht deutsch
Haushaltsgröße	Haushalte mit ... Person(en) 1 2 3 oder mehr
Erwerbstätigkeit	erwerbstätig nicht erwerbstätig
Region	Früheres Bundesgebiet Neue Länder

Neben den hier aufgelisteten Merkmalen wurde zusätzlich ein Interaktionseffekt in den Modellen berücksichtigt, der sich zusammensetzte aus der Altersgruppe der 17–25-jährigen Frauen.

Im Scientific Use File stehen die Kehrwerte der geschätzten Bleibewahrscheinlichkeiten für die einzelnen Zeiträume als Variablen zu Verfügung. Zur genauen Bezeichnung der Variablennamen und für welche Zeiträume Korrekturgewichte bereitgestellt sind, können Sie der Tabelle 4.2 in Kapitel 4.5 entnehmen.

4.3 Grundgesamtheit und Stichprobe des Mikrozensus-Panels 2001–2004

Die Grundgesamtheit, auf die mit dem Mikrozensus-Panel 2001–2004 hochgerechnet wird, kann auf unterschiedliche Zeithorizonte bezogen werden und besteht aus allen Personen, die zumindest in einem der beiden Erhebungszeitpunkte t_i und t_j (mit $i=2001$ und – je nach zu analysierendem Längsschnitt – $j=2002, 2003$ oder 2004) mit Hauptwohnsitz in einem Privathaushalt in Deutschland lebten. Personen am Nebenwohnsitz zählen nicht zur Grundgesamtheit, da sie bereits über den Hauptwohnsitz erfasst werden und Mehrfachzählungen auszuschließen sind.

Im Gegensatz zur Grundgesamtheit wird die Stichprobe des Mikrozensus-Panels für die Hochrechnung deutlich differenzierter abgegrenzt. Zwar zählen zur Stichprobe des Längsschnitts $t_i - t_j$ alle Personen, die zumindest in einem der beiden Erhebungszeitpunkte mit Hauptwohnsitz in einem Privathaushalt in Deutschland lebten, doch werden in der Stichprobe zur Hochrechnung des Mikrozensus-Panels nur Personen berücksichtigt,

- die zu allen Erhebungszeitpunkten zwischen den Erhebungszeitpunkten t_i und t_j mit Hauptwohnsitz in einem Privathaushalt in einem Stichprobenauswahlbezirk lebten,
- die im Erhebungszeitpunkt t_i mit Hauptwohnsitz in einem Privathaushalt in einem Stichprobenauswahlbezirk lebten und vor dem Erhebungszeitpunkt t_j verstarben,
- die im Erhebungszeitpunkt t_i mit Hauptwohnsitz in einem Privathaushalt in einem Stichprobenauswahlbezirk lebten und vor dem Erhebungszeitpunkt t_j ins Ausland verzogen,
- die nach dem Erhebungszeitpunkt t_i geboren wurden und zum Erhebungszeitpunkt t_j mit Hauptwohnsitz in einem Privathaushalt in einem Stichprobenauswahlbezirk lebten,
- die nach dem Erhebungszeitpunkt t_i aus dem Ausland in einen Stichprobenauswahlbezirk zuzogen und in diesem zum Erhebungszeitpunkt t_j mit Hauptwohnsitz in einem Privathaushalt erfasst wurden.

Die Beschränkung auf Personen mit Hauptwohnsitz in einem Privathaushalt ist notwendig, da im Mikrozensus alle Personen zu befragen sind, die nach den §§ 12 bis 16 des Melderechtsrahmengesetzes meldepflichtig sind. Personen mit mehreren meldepflichtigen Wohnsitzen werden dementsprechend an jedem ihrer Wohnsitze befragt, sofern der Wohnsitz sich in einem Auswahlbezirk des Mikrozensus befindet. Für die Hochrechnung müssen also Personen, die an ihrem Nebenwohnsitz befragt wurden ausgeschlossen werden, weil sie auch am Ort ihres Hauptwohnsitzes erfasst werden und somit Mehrfachzählungen vermieden werden.

Die Abgrenzung der Stichprobe erfolgt über die Variablen zur Kennzeichnung von Geburten, Sterbefälle, Zu- und Fortzüge, dem Wohnsitz ein Jahr vor der Erhebung sowie der Kennzeichnung über die Plausibilität der Zusammenführung.

4.4 Die Kalibrierung des Mikrozensus-Panels 2001–2004

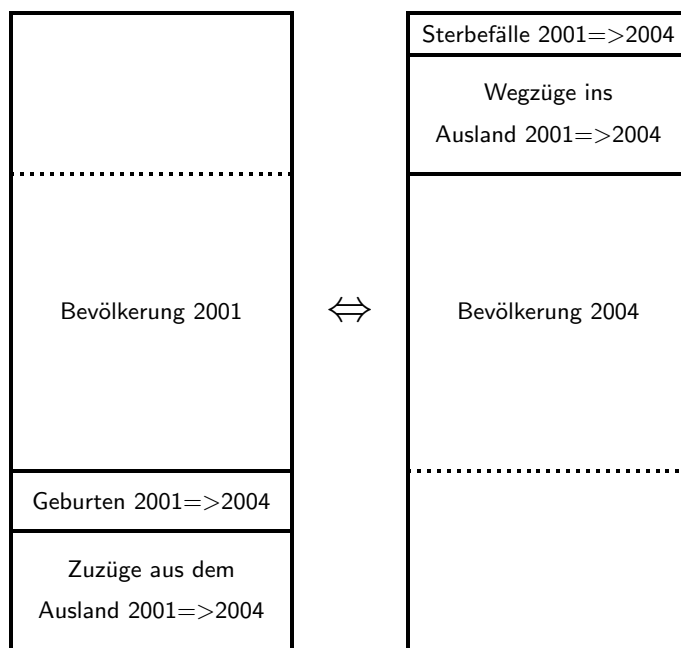
Um Aussagen über die Gesamtbevölkerung Deutschlands in den Jahren 2001 bis 2004 und die Veränderungen im Zeitverlauf treffen zu können, wird die Stichprobe an die Grundgesamtheit angepasst. Es erfolgt eine gleichzeitige Anpassung an Daten aus der laufenden Bevölkerungsfortschreibung und an Veränderungsdaten zur Bevölkerung, wie Geburten, Zuzüge aus dem Ausland und Sterbefälle und Fortzüge ins Ausland. Gleichzeitig wird zur Reduktion der gegebenenfalls durch räumliche Mobilität entstehenden, Verzerrungen an Variablen angepasst, die einen statistischen Zusammenhang zu dem Ereignis "Umzug innerhalb Deutschlands" aufweisen. Um diese Veränderungen zu berücksichtigen, wird auch an Variablen wie Alter, Eheschließungen und Ehescheidungen angepasst⁵.

Die Anpassung des Längsschnitts erfolgt parallel an die Bevölkerung von 2001 (t_i) und an die Bevölkerung des im Längsschnitt zuletzt betrachteten Jahres (t_j). Bei Verwendung der Gewichte ist lediglich darauf zu achten, dass die Summe der betrachteten Eckwertvektoren gleich ist. Hierzu werden zu den Angaben der immobilen Personen die Zahlen der Geburten, Zuzüge aus dem Ausland, Sterbefälle und Fortzüge ins Ausland addiert.

Betrachtet man den Längsschnitt 2001–2004, so werden Personen, die nach der Erhebung 2001 sterben oder ins Ausland ziehen, zwar 2001 befragt, doch sie fehlen in der Menge der befragten Personen von 2004. Um dieses Defizit auszugleichen, wird diese Menge zu derjenigen der immobilen Personen von 2004 hinzugefügt. Ebenso wurden Geburten und Zuzüge aus dem Ausland zwischen 2001 und 2004 nicht in der Stichprobenmenge 2001 erfasst. Diese Personen werden daher zur Menge der Personen von 2001 dazugezählt. Jede Person der betrachteten Stichprobe muss in der Menge der Personen von 2001 und in der Menge der Personen von 2004 vertreten sein. Beide Mengen sind in Abbildung 4.1 abgebildet.

⁵Diese parallele Randanpassung von Stichprobenergebnisse an bekannte bzw. geschätzte Randverteilungen wird als Kalibrierung bezeichnet.

Abbildung 4.1: Definition der beiden betrachteten Mengen für die Randanpassung des Längsschnitts 2001–2004



Personen, die während des gesamten betrachteten Längsschnitts im Mikrozensus erhoben wurden, sind in den Mengen Bevölkerung 2001 und Bevölkerung 2004 vertreten. Geburten sind unter Geburten 2001–2004 und in der Menge Bevölkerung 2004 zu finden. Analog gilt dies für Zuzüge aus dem Ausland. Sterbefälle und Fortzüge ins Ausland sind in der Menge Bevölkerung 2001 und 2004 in der Menge Sterbefälle 2001–2004 bzw. Fortzüge ins Ausland 2001–2004 vertreten. Angepasst wird nun an jeweils zwei Vektoren gleichzeitig, die jedoch während der Hochrechnung zu einem Vektor \mathbf{t}_x (siehe hierzu nachfolgende Ausführungen) zusammengefasst werden.

Um den Zufallsfehler und den möglichen durch Nonresponse entstehenden Bias zu verringern, wird der Kehrwert des Produkts aus Ziehungswahrscheinlichkeit π_k und Wahrscheinlichkeit am Wohnort zu verbleiben $\hat{\theta}_k$ mit einem Korrekturfaktor g_k multipliziert, um so für jede Person k den Hochrechnungsfaktor w_k zu erhalten.

$$w_k = \frac{g_k}{\pi_k \hat{\theta}_k}$$

Die Berechnung von g_k für jede Person k geschieht über die Anpassung folgender acht Eckwertvektoren der Bevölkerung durch eine verallgemeinerte Regressionschätzung, siehe Särndal et al. (1992):

Eckwertvektor 1:

- Bundesland * Alter * Geschlecht zum Zeitpunkt t_i
- Geburten zwischen t_i und t_j * Geschlecht
- Zuzüge aus dem Ausland zwischen t_i und t_j * Geschlecht

Eckwertvektor 2:

- Bundesland * Alter * Geschlecht zum Zeitpunkt t_j
- Sterbefälle und Fortzüge ins Ausland zwischen t_i und t_j

Eckwertvektor 3:

- Regierungsbezirk * Geschlecht zum Zeitpunkt t_i
- Geburten zwischen t_i und t_j * Geschlecht
- Zuzüge aus dem Ausland zwischen t_i und t_j * Geschlecht

Eckwertvektor 4:

- Regierungsbezirk * Geschlecht zum Zeitpunkt t_j
- Sterbefälle und Fortzüge ins Ausland zwischen t_i und t_j

Eckwertvektor 5:

- Staatsangehörigkeit (deutsch, nicht deutsch) zum Zeitpunkt t_i
- Geburten zwischen t_i und t_j * Geschlecht
- Zuzüge aus dem Ausland zwischen t_i und t_j * Geschlecht

Eckwertvektor 6:

- Staatsangehörigkeit (deutsch, nicht deutsch) zum Zeitpunkt t_j
- Sterbefälle und Fortzüge ins Ausland zwischen t_i und t_j

Eckwertvektor 7:

- Eheschließungen (mal 2) zwischen t_i und t_j

Eckwertvektor 8:

- Ehescheidungen (mal 2) zwischen t_i und t_j

Der *-Operator beschreibt die Kreuzkombination aus den angegebenen Variablen. Die Eheschließungen und -scheidungen müssen mit zwei multipliziert werden, da die Ereignisse fallbezogen vorliegen, die Gewichtungsfaktoren aber für die Personenebene ermittelt werden.

In die Berechnung von g_k geht dann die Differenz des Vektors t_x der Gesamtwerte der Hilfsmerkmale und des Horwitz-Thompson-Schätzers⁶ der Gesamtwerte

⁶Der Horwitz-Thompson-Schätzer ist ein allgemeiner Mittelwertschätzer für Stichproben mit beliebigen Ziehungswahrscheinlichkeiten der einzelnen Stichprobenelemente, bei denen es aber nicht zulässig ist, dass sie mehrfach in die Stichprobe gelangen.

$(\hat{\mathbf{t}}_{x,HT})$ mit in die Berechnung ein. \mathbf{x}_k bezeichnet für jede Person k den Vektor aller Ausprägungen der Hilfsmerkmale für die Person k .

$$g_k = 1 + (\mathbf{t}_x - \hat{\mathbf{t}}_{x,HT})' \left(\sum_{k=1}^n \frac{\mathbf{x}_k \mathbf{x}_k'}{\pi_k \hat{\theta}_k} \right)^{-1} \mathbf{x}_k$$

Bei der Berechnung des Horwitz-Thompson-Schätzers $(\hat{\mathbf{t}}_{x,HT})$ ist hier lediglich zu beachten, dass neben der Ziehungswahrscheinlichkeit auch der Mobilitätsfaktor der Person k mit in die Berechnung eingeht:

$$\hat{\mathbf{t}}_{x,HT} = \sum_{k=1}^n \frac{\mathbf{x}_k}{\pi_k \hat{\theta}_k}$$

Für die Berechnung des Regressionsschätzers \hat{t}_y ergeben sich somit folgende Beziehungen:

$$\begin{aligned} \hat{t}_y &= \hat{t}_{y,HT} + \hat{\mathbf{B}}(\mathbf{t}_x - \hat{\mathbf{t}}_{x,HT}) \\ &= \sum_{k=1}^n \left(1 + (\mathbf{t}_x - \hat{\mathbf{t}}_{x,HT})' \left(\frac{\mathbf{x}_k \mathbf{x}_k'}{\pi_k \hat{\theta}_k} \right)^{-1} \mathbf{x}_k \right) \frac{1}{\pi_k \hat{\theta}_k} y_k \\ &= \sum_{k=1}^n w_k y_k \end{aligned}$$

Bei der Konstruktion der Eckwerte ist zu bedenken, dass nicht alle Daten monatlich ermittelt werden. Teilweise stehen, wie zum Beispiel für Ehescheidungen, nur jährliche Angaben aus der Bevölkerungsfortschreibung zur Verfügung. In diesen Fällen wird eine lineare Interpolation vorgenommen, um näherungsweise Aprilwerte zu erhalten. So werden die Angaben zum vorherigen Jahr mit dem Faktor $2/3$ gewichtet, die aus dem aktuellen Jahr mit $1/3$. Liegen monatliche Angaben vor, wie beispielsweise bei der Zahl der Eheschließungen und Geburten, so werden diese jeweils von Mai des vorherigen Jahres bis April des aktuellen Jahres verwendet. Die lineare Interpolationen eines Teils der Eckwerte können daher zu Ungenauigkeiten in den Eckwerten führen.

Gemeinsam mit der Ziehungswahrscheinlichkeit π_k für jede Person k geht der geschätzte Mobilitätsfaktor $\hat{\theta}_k$ in die Kalibrierung der Stichprobe an die Bevölkerungsdaten ein. Analog wurde mit denselben Eckwerten ein zusätzlicher Hochrechnungsfaktor berechnet, der den Mobilitätsfaktor nicht berücksichtigt, indem in den entsprechenden Formeln $\hat{\theta}_k = 1$ gesetzt wurde. Im Scientific Use File stehen die durch die Kalibrierung erzeugten Hochrechnungsfaktoren als Variablen mit und ohne Berücksichtigung der Mobilitätsfaktoren zur Verfügung. Die genaue Bezeichnung der Variablennamen und für welche Übergänge Hochrechnungsfaktoren zur Verfügung stehen, können Sie der [Tabelle 4.2](#) in Kapitel 4.5 entnehmen.

4.5 Die im Mikrozensus-Panel 2001–2004 bereitgestellten Hochrechnungs- und Mobilitätsfaktoren

Insgesamt stehen im Mikrozensus-Panel 2001–2004 zwei unterschiedliche Längsschnittgewichte und ein Mobilitätsgewicht für jeweils unterschiedliche Zeiträume zur Verfügung:

1. Längsschnittgewichte zur Hochrechnung ohne Berücksichtigung der Mobilitätsgewichte,
2. Längsschnittgewichte zur Hochrechnung mit Berücksichtigung der Mobilitätsgewichte und
3. die Mobilitätsgewichte selbst.

Zur besseren Kennzeichnung der einzelnen Gewichtungsarten setzt sich der jeweilige Variablenname eines Gewichtungsfaktors aus zwei Kennungen zusammen. Die erste Kennung bezeichnet die Gewichtungsart und die zweite Kennung die Übergangsdauer. Die genauen Variablenamen, die jeweils zugehörigen Gewichtungsfaktoren und Übergangsdauer können Sie nachfolgender Tabelle entnehmen.

Tabelle 4.2: Variablenamen der einzelnen Längsschnitt- und Mobilitätsgewichte im Mikrozensus-Panel 2001–2004

Variablenname	Gewichtungsfaktor	Übergang
HG_0102	Längsschnittgewicht zur Hochrechnung ohne Berücksichtigung der Mobilitätsgewichte für den Übergang ...	2001 ⇒ 2002
HG_0103		2001 ⇒ 2003
HG_0104		2001 ⇒ 2004
HGM_0102	Längsschnittgewicht zur Hochrechnung mit Berücksichtigung der Mobilitätsgewichte für den Übergang ...	2001 ⇒ 2002
HGM_0103		2001 ⇒ 2003
HGM_0104		2001 ⇒ 2004
MG_0102	Mobilitätsgewicht (Kehrwert der Bleibewahrscheinlichkeit) für den Übergang ...	2001 ⇒ 2002
MG_0103		2001 ⇒ 2003
MG_0104		2001 ⇒ 2004
MG_0203		2002 ⇒ 2003
MG_0204		2002 ⇒ 2004
MG_0304		2003 ⇒ 2003

Wie Sie vorstehender Tabelle entnehmen können, stehen derzeit die Längsschnittgewichte zur Hochrechnung ohne und mit Berücksichtigung der Mobilitätsgewichte nur mit Ausgangsbasis des Erhebungsjahres 2001 zur Verfügung, wohingegen für die Mobilitätsgewichte auch die Erhebungsjahre 2002 und 2003 als Ausgangsbasis vorliegen.

4.6 Weiterführende Literatur

- Basic, E. (2008): *The Problem of Missing Residential Mobility Information in the German Microcensus. An Evaluation of Two Statistical Approaches with the Socio-Economic Panel*. Hamburg: Verlag Dr. Kovac
- Basic, E.; Marek, I.; Rendtel, U. (2005): *The German Microcensus as a tool for longitudinal data analysis: An evaluation using SOEP data*. Arbeitspapier #3, Methodenverbund "Aufbereitung und Bereitstellung des Mikrozensus als Panelstichprobe". URL: <http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Content/Wissenschaftsforum/MethodenVerfahren/Mikrozensus/Arbeitspapiere/Arbeitspapier3,property=file.pdf>
- Clarke, P.S.; Chambers, R.L. (1998): *Estimating Labour Force Gross Flows From Surveys Subject to Household-Level Nonignorable Nonresponse*, in: *Survey Methodology*; **24**: 123-129.
- Deville, J.C.; Särndal, C.-E. (1992): *Calibration estimation in survey sampling*, in: *Journal of the American Statistical Association*, **87**: 376-382.
- Liang, K. Y.; Zeger, S. L. (1986): *Longitudinal Data Analysis Using Generalized Linear Models*, in *Biometrika*, **73**: 13-22.
- Lundström, S.; Särndal, C.-E. (2002): *Estimation in the presence of Nonresponse and Frame Imperfections*. Stockholm: Statistics Sweden.
- McCullagh, P. ; Nelder, J. A. (1989): *Generalized Linear Models*. Chapman & Hall
- Rohloff, S. (2005): *Das Hochrechnungsverfahren für Längsschnittauswertungen aus dem Mikrozensus*. Arbeitspapier #6, Methodenverbund "Aufbereitung und Bereitstellung des Mikrozensus als Panelstichprobe". URL: <http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Content/Wissenschaftsforum/MethodenVerfahren/Mikrozensus/Arbeitspapiere/Arbeitspapier6,property=file.pdf>
- Särndal, C.-E.; Swensson, B.; Wretman, J. (1992) *Model Assisted Survey Sampling*. New-York: Springer-Verlag.
- Zeger, S. L.; Liang, K. Y. (1986): *Longitudinal Data Analysis for Discrete and Continuous Outcomes*, in: *Biometrics*, **42**: 121-130.

Kapitel 5

Möglichkeiten zur Korrektur von Verzerrungen aufgrund räumlicher Immobilität

Edin Basic (Freie Universität Berlin)

5.1 Das Auftreten von räumlicher Mobilität

Räumliche Mobilität, d.h. der Zu- und Fortzug von Personen und Haushalten in einen bzw. aus einem Auswahlbezirk wird im Rahmen der Feldarbeit zum Mikrozensus erfasst und ist im MZ-Panel als Längsschnittinformation dokumentiert (siehe hierzu die Erläuterungen in Abschnitt 2.5). Dies ist für eine Längsschnittnutzung des Mikrozensus von großem Vorteil.

Pro Jahr ergibt sich ein Mobilitätsrisiko von ca. 13 Prozent, vgl. Basic et al. (2005). Räumliche Mobilität tritt selektiv auf. Daher kann ihre Nicht-Berücksichtigung bei Längsschnittanalysen zu verzerrten Ergebnissen führen. Die Aufdeckung eines Mobilitätsbias und dessen Korrektur ist das zentrale Problem des MZ-Panels. Dabei lassen sich folgende Situationen unterscheiden:

- a) Junge Leute verlassen das Elternhaus wegen Ausbildung oder Familiengründung.
- b) Die Geburt von Kindern erzwingt einen Umzug in eine größere Wohnung.
- c) Zwei Lebenspartner ziehen zusammen.
- d) Die Ehescheidung ist an eine vorherige Trennung der Wohnung der Ehepartner gebunden.
- e) Bei Tod eines Lebenspartners ist die alte Wohnung zu groß.

- f) Ein Wechsel ins Altersheim tritt im höheren Lebensalter verstärkt auf.
- g) Ein Wechsel des Arbeitgebers kann auch einen Wechsel des Wohnorts nach sich ziehen. In vielen Fällen wird jedoch zunächst einmal zwischen Arbeitsplatz und Wohnort gependelt.

Die meisten dieser Ereignisse sind relativ stark mit dem Lebensalter verknüpft: a), b) und c) treffen für junge Leute zu, während e) und f) für ältere Personen von Bedeutung sind. Daher verspricht die Kontrolle über die Variable Alter eine Verringerung eines Mobilitätsbias.

5.2 Methoden zur Biasaufdeckung

Ein wesentliches Hilfsmittel zur Aufdeckung des Mobilitätsbias im MZ-Panel sind Ergebnisse des Sozio-ökonomischen Panels (SOEP). Da das SOEP Umzüge erfasst, wird empfohlen, die jeweilige Analyse einem künstlichen Non-Coverage Effekt zu unterziehen. Non-Coverage ist ein Begriff aus der Stichprobentheorie und besagt, das zu untersuchende Elemente zwar zur Zielpopulation gehören, aufgrund des Auswahlplans sie aber keine Chance haben in die Stichprobe zu gelangen.

Im Rahmen der Untersuchungen zur Biasaufdeckung mit dem SOEP werden dementsprechend zwei Stichproben gebildet und deren Schätzwerte miteinander verglichen. Der Schätzwert der ersten Stichprobe $\hat{\vartheta}_{FULL}$ basiert auf den Angaben aller Personen und der Schätzwert der zweiten Stichprobe $\hat{\vartheta}_{IMMO}$ nur auf den Angaben der immobilen Personen. In der zweiten Stichprobe werden die mobilen Personen also systematisch ausgeschlossen. Die Differenz dieser beiden Schätzwerte kann dann als Mass für den Mobilitätsbias genommen werden.¹ Dieser Strategie liegen zwei Annahmen zugrunde:

1. Der Mikrozensus und das SOEP haben unterschiedliche Auswahldesigns, vgl. z.B. Heidenreich (2002) für den Mikrozensus und Haisken-DeNew/Frick (2005) für das SOEP. Wenn die Mobilität hinsichtlich dieser Designmerkmale unterschiedlich stark ausfällt, so kann auch der Mobilitätsbias durch das unterschiedlich starke Auftreten der Untergruppen unterschiedlich groß sein. Beispielweise sind Haushalte in Ost-Deutschland und Haushalte mit Ausländern – beides Gruppen, die im SOEP überrepräsentiert sind – mobiler als Haushalte in West-Deutschland, vgl. Basic et al. (2005). Hier müssen wir annehmen, dass der Schätzer $\hat{\vartheta}$ die Effekte des Stichprobendesigns entweder durch Benutzung der Designgewichte oder durch Wahl von Kontrollvariablen berücksichtigt.² Es ist jedoch beruhigend zu sehen, dass beim Vergleich der

¹Es wird also nicht verlangt, dass $\hat{\vartheta}_{IMMO}$ für MZ-Panel und SOEP übereinstimmen.

²Diese Annahme wird vernünftigerweise für jeden Schätzer $\hat{\vartheta}$ gemacht, da sonst der Schätzer nicht mehr konsistent wäre.

globalen Mobilitätsziffern von SOEP (mit und ohne Design-Gewichtung) und Mikrozensus nur geringfügige Unterschiede auftreten, vgl. Basic et al. (2005, Tabelle 1).

2. Auch das SOEP hat Probleme mit der Weiterverfolgung mobiler Haushalte, vgl. Rendtel (1995). Hier sind es in erster Linie die neuen Haushalte, also Haushaltsabsplaltungen, wenn beispielsweise Kinder den elterlichen Haushalt verlassen oder Paare sich trennen, die das Ausfallrisiko von ca. 3 Prozent auf 15 bis 20 Prozent vergrößern. Wenn diese ausgefallene Personen sich anders verhalten als die 80 bis 85 Prozent der neuen Haushalte, die im SOEP verbleiben, ergibt sich das Problem, dass die Schätzung $\hat{\vartheta}_{FULL}$ wiederum inkonsistent ist.³ In der Regel lässt sich die Annahme, dass sich die ausgefallenen Personen und die in der Erhebung verbleibenden Personen gleich verhalten, nicht statistisch überprüfen, da die hierfür notwendigen Daten fehlen. Beispielsweise könnte man annehmen, dass Personen, die in eine andere Stadt gezogen sind und dort einen neuen Arbeitsplatz haben, weniger Zeit für Interviews haben und aus diesem Grund eine geringere Interview-Neigung haben. In einem solchen Fall würde auch das SOEP den Anteil der Jobwechsler unterschätzen. Allerdings zeigt eine einfache Überschlagsrechnung, dass die hieraus resultierende Unterschätzung gering sein muss, wenn man von einer Mobilitätsrate von 15 Prozent und einer Ausfallrate unter Mobilien von ebenfalls 15 Prozent ausgeht. Damit wird (unter der Annahme, dass Mobilität und Ausfall statistisch unabhängige Prozesse sind) bei lediglich 2.25 Prozent (entspricht dem Produkt von Mobilitäts- und Ausfallrate) der Stichprobe ein Anteil, etwa der Anteil der Jobwechsler, eventuell etwas unterschätzt.

Generell werden zwei Strategien zur Biasaufdeckung vorgeschlagen:

- a) Die Differenz $\Delta = \hat{\vartheta}_{FULL} - \hat{\vartheta}_{IMMO}$ kann für unterschiedlich lange Zeitintervalle berechnet werden. Da die Zahl der mobilen Personen in etwa linear wächst, müsste sich ein Mobilitätsbias in einem Trend über die drei Messzeitpunkte zeigen, vgl. Basic et al. (2005).
- b) Möchte man durch einen formalen statistischen Test untersuchen, ob $\hat{\vartheta}_{IMMO}$ ein konsistenter Schätzer ist, so führt dies bei ungewichteten Daten auf einen Hausman-Test,⁴ wenn $\hat{\vartheta}_{FULL}$ ein effizienter Schätzer ist und $\hat{\vartheta}_{IMMO}$ ein unter der Null-Hypothese konsistenter Schätzer ist. In diesem Fall ist die Differenz unter der Null-Hypothese (kein Mobilitätsbias) approximativ multivariat normalverteilt mit Erwartungswertvektor 0 und einer Kovarianzmatrix

³In diesem Fall ist die Panelmortalität informativ hinsichtlich des zu schätzenden Modells.

⁴Bei dem üblichen Einsatz in der Ökonometrie ist $\hat{\vartheta}_1 = \hat{\vartheta}_{FULL}$ unter H_0 effizient und unter der Alternative inkonsistent. $\hat{\vartheta}_2 = \hat{\vartheta}_{IMMO}$ ist dort sowohl unter H_0 als auch der Alternative konsistent, vgl. Greene (2000: 576). Betrachtet wird jedoch nur die Differenz $\hat{\vartheta}_1 - \hat{\vartheta}_2$ und ihre Verteilung unter H_0 .

Σ , die sich aus der Differenz der beiden einzelnen Kovarianzmatrizen von $\hat{\vartheta}_{IMMO}$ und $\hat{\vartheta}_{FULL}$ ergibt. Hieraus lässt sich wieder eine χ^2 -Teststatistik T mit $T = (\hat{\vartheta}_{FULL} - \hat{\vartheta}_{IMMO})' \Sigma^{-1} (\hat{\vartheta}_{FULL} - \hat{\vartheta}_{IMMO})$ herleiten, vgl. Hausman (1978).

5.2.1 Beispiel

Im Folgenden wird die Methodik zur Biasaufdeckung beispielhaft an der Schätzung von Übergängen zwischen verschiedenen Erwerbszuständen vorgestellt. Da im Mikrozensus in erster Linie diskrete Merkmale auftreten, werden Übergänge zwischen zwei diskreten Merkmalen A und B analysiert, wobei der Zeitraum zwischen den Merkmalen ein, zwei oder drei Jahre betragen kann. Hinzu kommt das Merkmal R (Response), das anzeigt, ob im betrachteten Zeitintervall regionale Mobilität auftritt ($R = 0$) oder nicht ($R = 1$). Mit A als Ausgangszustand, B als Endzustand und $P(B|A)$ als Übergangswahrscheinlichkeit zwischen diesen Zuständen können mit dem SOEP zwei Schätzungen dieser Übergangswahrscheinlichkeiten ermittelt werden: $\hat{P}(B|A)$, die Schätzung auf Basis aller Personen und $\hat{P}(B|A, R = 1)$, die Schätzung auf Basis immobiler Personen. Bei der Übertragung der SOEP Ergebnisse auf den Mikrozensus ist zu beachten, dass die gemessenen Konzepte in den beiden Erhebungen in einer vergleichbaren Form aufbereitet werden müssen.

Es werden drei Erwerbszustände unterschieden: Erwerbstätig (E); Arbeitslos (U) und Nichterwerbsperson (N). Die Kriterien nach welchen die Einteilung des Erwerbszustandes in diese drei Kategorien erfolgte, befindet sich bei Basic (2005). Hier lassen sich a-priori zwei Bias-Hypothesen aufstellen:

- i) Die Übergangswahrscheinlichkeit von U nach E wird durch die ausschließliche Betrachtung von immobilen Personen unterschätzt, da die Aufnahme einer Erwerbstätigkeit in einigen Fällen mit einem Ortswechsel verbunden ist.
- ii) Die Übergangswahrscheinlichkeit von N nach E wird ebenfalls unterschätzt, da die Aufnahme einer Erwerbstätigkeit von Jugendlichen häufig mit dem Auszug aus der elterlichen Wohnung verbunden ist.

Beide Hypothesen werden anhand der SOEP-Daten klar bestätigt, vgl. Tabelle 5.1. Die erste Spalte in der Tabelle 5.1 (FULL) zeigt die Ergebnisse der Schätzung auf der Basis aller Personen, die zweite Spalte (IMMO) die Ergebnisse der Schätzung nur auf der Basis von immobilen Personen und die dritte Spalte (Δ) die Differenz dieser beiden Werte. Hier zeigt sich mit zunehmender Länge des Auswertungsintervalls eine immer grössere Unterschätzung der Übergangswahrscheinlichkeiten zwischen diesen beiden Erwerbszuständen. Für die anderen Übergänge war der Mobilitätsbias relativ gering, vgl. Basic et al. (2005). Die Anwendung des Hausman-Tests zeigt, dass alle Abweichungen in der Tabelle auf dem 5 Prozent Niveau signifikant sind.

Tabelle 5.1: Übergangswahrscheinlichkeiten zwischen Arbeitslosigkeit bzw. Nichterwerbsperson (1996) und Erwerbstätigkeit (t=1997, 1998, 1999). Datenbasis: SOEP

t	E			
	FULL	IMMO	Δ	
U	1997	32.84	30.85	1.99
	1998	34.92	31.79	3.13
	1999	41.37	37.46	3.89
N	1997	12.74	11.64	1.10
	1998	19.66	16.07	3.59
	1999	25.89	21.13	4.76

E Erwerbstätig

U Arbeitslos

N Nichterwerbsperson

FULL Übergangswahrscheinlichkeit (in Prozent) auf der Basis aller Personen

IMMO Übergangswahrscheinlichkeit (in Prozent) auf der Basis der immobilen Personen

$\Delta = FULL - IMMO$

5.2.2 Die Größenordnung des Mobilitätsbias

Für die Obergrenze des Mobilitätsbias läßt sich eine einfache, freilich etwas grobe und pessimistische Abschätzung finden. Dies sei beispielhaft an den Übergängen zwischen den Erwerbszuständen verdeutlicht. Wir nehmen an, dass der Anteil der mobilen Personen bei 20 Prozent liegt. Der Anteil der Personen mit Wechsel von Arbeitslosigkeit in die Erwerbstätigkeit unter den Immobilen betrage r . Für die mobilen Personen werde der ungünstigste Fall angenommen, nämlich eine Übergangsrate von 1. Der relative Bias⁵ B beträgt in diesem Fall:

$$B = \frac{0.8r + 0.2 - r}{0.8r + 0.2} = \frac{1 - r}{4r + 1}$$

Für $r = 0.30$ erhält man in diesem Fall $B = 0.32$. Dieser Wert ist jedoch nur eine obere Schranke und ein Vergleich mit dem empirischen Wert in Tabelle 5.1 ergibt für 1999 einen relativen Bias von etwa 10 Prozent. Das heißt, die Annahme, alle umgezogenen Arbeitslosen hätten eine neue Stelle gefunden, war zu extrem. Bei unseren empirischen Untersuchungen zum Thema Arbeitsmarkt wies der relative Bias nur in Ausnahmefällen einen Wert von etwa 20 Prozent auf, meist lag der Wert jedoch deutlich darunter.

⁵Der relative Bias ist definiert als $B = \frac{P(B|A) - P(B|A, R=1)}{P(B|A)}$

5.2.3 Makro hausman.sas

Das Ziel dieses SAS-Makros ist die simultane Schätzung der Übergangswahrscheinlichkeiten zwischen zwei diskreten Merkmalen für zwei Gruppen von Personen⁶. Die erste Schätzung basiert auf der gesamten Stichprobe unter Einschluss der mobilen Personen, während die zweite Schätzung nur auf der Basis von immobilien Personen basiert. Das Makro liefert auch die Differenz dieser beiden Schätzer sowie die Teststatistik des Hausman-Tests. Es ist auf beliebige Datensätze anwendbar. Einzige Voraussetzung ist die binäre Datenstruktur (0/1) der Mobilitätsvariablen, d.h. die Variablen, die anzeigen, ob die Person im betrachteten Zeitraum mobil oder immobil war.

Aufruf

Die Syntax des Makros ist wie folgt:

- dataset = Datensatz
- v1 = Merkmal zum Startzeitpunkt
- v2 = Merkmal zum Endzeitpunkt
- mob = Mobilitätsindikator mit Ausprägungen 0 für Immobiler und 1 für Mobile

5.3 Methoden zur Biasreduktion

Es werden drei Methoden zur Biasreduktion vorgeschlagen. Welche Methode dabei angewandt werden soll, hängt von dem Prozess ab, der fehlende Daten generiert. Nach Rubin (1976) kann zwischen folgenden Ausfallprozessen unterschieden werden:

- Die fehlenden Daten sind missing completely at random (MCAR). In diesem Fall ist der Ausfallprozess sowohl von den beobachteten als auch von den unbeobachteten Merkmalen unabhängig.
- Die fehlenden Daten sind missing at random (MAR). In diesem Fall hängt der Ausfallprozess von den beobachteten Merkmalen ab, ist aber unabhängig von unbeobachteten Merkmalen. Es handelt sich um bedingt zufälligen Datenausfall; der Ausfall lässt sich vollständig über Variablen erklären, die beobachtet wurden.
- Die fehlenden Daten sind not missing at random (NMAR). In diesem Fall hängt der Ausfallprozess zusätzlich von den unbeobachteten Merkmalen ab.

⁶Das Makro hausman.sas können Sie von der Internetseite des Verbundprojektes beim Statistischen Bundesamt herunterladen.

Im Falle, dass die fehlenden Daten MAR sind, unterscheiden sich die Übergangsmatrizen von mobilen und immobilen Personen nicht systematisch. In diesem Fall können mittels einer Gewichtung der Beobachtungen mit dem Kehrwert der Beobachtungswahrscheinlichkeit, also mit dem Kehrwert der Wahrscheinlichkeit für räumliche Immobilität, konsistente Schätzergebnisse erzielt werden, vgl. Robins/Rotnitzky (1995); Basic/Rendtel (2006). Dieser Ansatz wird weiter unten diskutiert, siehe Abschnitt 5.3.2.

Falls die fehlenden Daten NMAR sind, muss ein gemeinsames Modell für fehlende und beobachtete Fälle spezifiziert werden. Allerdings ist das Modell nur unter zusätzlichen Restriktionen schätzbar. Dieser Ansatz wird in Abschnitt 5.3.3 diskutiert.

Bevor auf den Gewichtungs- und Modellierungsansatz eingegangen wird, wird zuerst ein sehr einfacher Ansatz vorgestellt, bei dem versucht wird, den Grund für die räumliche Mobilität durch weitere Variablen, sogenannte Kontrollvariablen, in der Analyse zu erklären.

5.3.1 Verwendung von Kontrollvariablen

Die Idee dieser Methode besteht darin, den Zusammenhang zwischen Ausfallprozess und den Werten der fehlenden Variablen durch die beobachteten Daten zu erklären. Besteht beispielsweise ein Zusammenhang zwischen Alter und Mobilität, aber in keiner Altersgruppe ein Zusammenhang zwischen Erwerbsbeteiligung und Mobilität, dann führen unter Kontrolle der Altersgruppen die Analysen zur Erwerbsbeteiligung nicht zu Ergebnisverzerrungen.

Sehr einfach durchzuführen ist die Verwendung von Kontrollvariablen bei der Erstellung von Tabellen. Diese Kontrollvariablen sollten eng mit der räumlichen Mobilität in Verbindung stehen und gleichzeitig für die analysierte Tabelle Sinn machen. Für die Mobilitätsproblematik erweist sich das Lebensalter als eine zentrale Kontrollvariable. Hier wird die analysierte Kontingenztabelle nach Altersgruppen unterteilt. Allein die Verwendung dieser einen Kontrollvariablen reduziert den Mobilitätsbias wesentlich, vgl. Tabelle 5.2. Für den Übergang 1996-1999 reduziert sich der Bias auf weniger als die Hälfte.

Tabelle 5.2: Übergangswahrscheinlichkeiten zwischen Arbeitslosigkeit bzw. Nichterwerbsperson (1996) und Erwerbstätigkeit (t=1997, 1998, 1999) bei Kontrolle nach Altersgruppe. Datenbasis: SOEP

t	UE			NE		
	FULL	IMMO	Δ	FULL	IMMO	Δ
	Alter ≤ 30					
96→97	52.43	52.12	0.31	25.98	24.16	1.82
96→98	55.09	56.02	-0.93	37.86	33.33	4.53
96→99	65.69	64.05	1.64	50.07	46.28	3.79
	Alter > 30					
96→97	24.02	22.04	1.98	6.36	6.13	0.23
96→98	25.90	23.25	2.75	10.13	8.81	1.32
96→99	30.28	28.78	1.50	12.72	11.28	1.44
	Insgesamt					
96→97	32.84	30.85	1.99	12.74	11.64	1.10
96→98	34.92	31.79	3.13	19.66	16.07	3.59
96→99	41.37	37.46	3.89	25.89	21.13	4.76

UE Übergang Arbeitslos → Erwerbstätig
 NE Übergang Nichterwerbsperson → Erwerbstätig
 FULL Übergangswahrscheinlichkeit (in Prozent) auf der Basis aller Personen
 IMMO Übergangswahrscheinlichkeit (in Prozent) auf der Basis der immobilen Personen
 $\Delta = FULL - IMMO$

Werden noch weitere Kontrollvariablen benutzt und soll ihr Einfluss auf bestimmte Übergänge, z.B. den Wechsel von Arbeitslosigkeit in die Erwerbstätigkeit, modelliert werden, so wird man diese Analyse in der Regel über ein Logit-Modell durchführen. Das Logit-Modell für die Modellierung des Übergangs von Arbeitslosigkeit in die Erwerbstätigkeit lautet:

$$\ln \frac{P(Y = 1|x)}{1 - P(Y = 1|x)} = x'\beta \quad (5.1)$$

Die abhängige Variable Y ist ein Indikator für den Übergang von Arbeitslosigkeit in die Erwerbstätigkeit und x sind Variablen, die diesen Übergang erklären, wie z.B. Alter, Schulbildung, Dauer der Arbeitslosigkeit etc. Tabelle 5.3 zeigt die Ergebnisse der Logit-Schätzung für den Übergang aus der Arbeitslosigkeit in die Erwerbstätigkeit und Tabelle 5.4 die Ergebnisse für den Übergang aus der Nichterwerbsperson in die Erwerbstätigkeit.

Tabelle 5.3: Logistische Regression: Übergang von Arbeitslosigkeit (1996) nach Erwerbstätigkeit (1998). Datenbasis: SOEP

Variable	FULL	IMMO	Δ
Konstante	-0.5989 (0.0003)	-0.6084 (0.0009)	0.0095 (0.9058)
Alter ≤ 30	1.2521 (0.0001)	1.3920 (0.0001)	-0.1399 (0.1915)
Bildung keine Ausbildung	-0.2351 (0.2272)	-0.1707 (0.4530)	-0.0644 (0.5840)
Hochschule	0.4872 (0.0931)	0.3447 (0.3114)	0.1425 (0.4231)
Region West-Deutschland	-0.1240 (0.4651)	-0.1467 (0.4592)	0.0277 (0.8246)
Schule kein Schulabschluss	0.0785 (0.7860)	-0.0689 (0.8384)	0.1474 (0.3990)
Abitur	0.2165 (0.4150)	0.1205 (0.6972)	0.0960 (0.5463)
Geschlecht männlich	0.3841 (0.0135)	0.2558 (0.1566)	0.1283 (0.1617)
Dauer der U > 1 Jahr	-1.2364 (0.0001)	-1.3065 (0.0001)	0.0701 (0.4479)
Beobachtungen	899	715	

Abhängige Variable: Indikator für Übergang $U \rightarrow E$, p-Werte in Klammern

U Arbeitslos

$U \rightarrow E$ Übergang Arbeitslos \rightarrow Erwerbstätig

FULL Übergangswahrscheinlichkeit (in Prozent) auf der Basis aller Personen

IMMO Übergangswahrscheinlichkeit (in Prozent) auf der Basis der immobilen Personen

$\Delta = FULL - IMMO$

Die erste Spalte (FULL) zeigt die Ergebnisse der Schätzung auf Basis der Gesamtstichprobe, d.h. der Immobilen unter Einschluss der mobilen Personen. In der zweiten Spalte (IMMO) befinden sich die Ergebnisse der Schätzung lediglich auf der Basis der immobilen Personen und in der dritten Spalte (Δ) die Differenz dieser beiden Werte und die p-Werte der Teststatistik. Empirisch zeigt sich eine sehr gute Übereinstimmung der geschätzten Koeffizienten für die Gesamtstichprobe und die Teilstichprobe der Immobilen. Dies betrifft sowohl die Größe als auch die Signifikanz der geschätzten Koeffizienten, vgl. Tabelle 5.3 und Tabelle 5.4.

Tabelle 5.4: Logistische Regression: Übergang von Nichterwerbsperson (1996) nach Erwerbstätigkeit (1998). Datenbasis: SOEP

Variable	FULL	IMMO	Δ
Konstante	-2.2905 (0.0001)	-2.3496 (0.0001)	0.0591 (0.4781)
Alter ≤ 30	1.6378 (0.0001)	1.6214 (0.0001)	0.0164 (0.8636)
Bildung keine Ausbildung	-0.1076 (0.4733)	-0.0781 (0.6636)	-0.0295 (0.7644)
Bildung Hochschule	-0.1008 (0.7214)	0.0544 (0.8639)	-0.1552 (0.2805)
Region West-Deutschland	-0.2893 (0.0427)	-0.3935 (0.0163)	0.1042 (0.1950)
Schule kein Schulabschluss	-0.1560 (0.3706)	-0.1498 (0.4643)	-0.0061 (0.9544)
Schule Abitur	0.3096 (0.0671)	0.2850 (0.1627)	0.0246 (0.8300)
Geschlecht männlich	0.4939 (0.0001)	0.4521 (0.0025)	0.0418 (0.5876)
Anzahl der Kinder im HH	0.2895 (0.0001)	0.2626 (0.0001)	0.0269 (0.4197)
Beobachtungen	2 079	1 743	

Abhängige Variable: Indikator für Übergang $N \rightarrow E$, p-Werte in Klammern

$N \rightarrow E$ Übergang Nichterwerbsperson \rightarrow Erwerbstätig

FULL Übergangswahrscheinlichkeit (in Prozent) auf der Basis aller Personen

IMMO Übergangswahrscheinlichkeit (in Prozent) auf der Basis der immobilen Personen

$\Delta = FULL - IMMO$

5.3.1.1 Makro logit.sas

Im Gegensatz zum Makro `hausman.sas` wird dieses Makro zur Schätzung der Koeffizienten eines Logit-Modells verwendet⁷, vgl. Tabelle 5.3 und 5.4. Das Makro liefert die geschätzten Koeffizienten beider Logit-Modelle für Immobiler und Gesamter (Immobilier einschließlich räumlich Mobiler) sowie deren Differenzen und die Teststatistik des Hausman-Tests. Das Makro ist auf beliebige Datensätze anwendbar. Die Voraussetzungen dafür sind die binäre Datenstruktur (0/1) der Mobilitätsvariable, d.h. der Variable, die anzeigt, ob die Person im betrachteten Zeitraum mobil oder immobil war und die binäre Datenstruktur der Zielgröße.

⁷Das Makro `logit.sas` können Sie von der Internetseite des Verbundprojektes beim Statistischen Bundesamt herunterladen.

Aufruf

Die Syntax des Makros ist wie folgt:

- dataset = Datensatz
- v1 = Zielgrösse
- var1 = kategorische Variable(n)
- var2 = metrische Variable(n)
- mob = Mobilitätsindikator mit Ausprägungen 0 für Immobile und 1 für Mobile

5.3.2 Gewichtung mit dem Kehrwert der Immobilitätswahrscheinlichkeit

Bei diesem Ansatz werden lediglich immobile Personen betrachtet. Die wesentliche Idee dieses Ansatzes besteht darin, die Beobachtungen der immobilen Personen hochzugewichten, um für die fehlenden mobilen Personen, über die man keine Informationen hat, zu kompensieren. Dabei bekommen Personen mit höherer Mobilitätswahrscheinlichkeit ein höheres Gewicht und Personen mit geringerer Mobilitätswahrscheinlichkeit ein kleineres Gewicht. In einem ersten Schritt werden die (Im-)mobilitätswahrscheinlichkeiten ermittelt, z.B. über ein Logit-Modell mit dem Indikator für Mobilität als abhängiger Variable. Als unabhängige Variablen benutzt man beispielsweise sozio-demographische Merkmale aus der letzten Welle, die sowohl für mobile als auch für immobile Personen bekannt sind und mit der räumlichen Mobilität in Verbindung stehen. Im zweiten Schritt werden dann die Beobachtungen der immobilen Personen mit dem Kehrwert der Immobilitätswahrscheinlichkeit gewichtet. Diese Methode wird anhand der Modellierung der Übergangswahrscheinlichkeiten zwischen zwei diskreten Merkmalen veranschaulicht.

Die Wahrscheinlichkeit in einen Zustand $B = b$ zum Zeitpunkt t zu wechseln hängt von dem Zustand A zum Zeitpunkt $t - 1$ ab - dies entspricht genau einer Spalte der Übergangsmatrix. Formal lässt sich der Zusammenhang als Logit-Modell darstellen, wobei als Kovariate der Vektor a_i der Merkmalsausprägungen für Person i in Welle $t - 1$ gewählt wird und Y_i der Indikator für den betrachteten Übergang in einen bestimmten Zustand b ist. Das Logit-Modell sieht dann folgendermaßen aus:

$$\ln \frac{P(Y_i = 1|a_i)}{1 - P(Y_i = 1|a_i)} = a_i' \beta \quad (5.2)$$

Die Score-Gleichung des Logit-Modells in (2) lautet:

$$\sum_{i \in s} a_i (Y_i - P(Y_i = 1|a_i)) = 0 \quad (5.3)$$

Hierbei ist s die Stichprobe aller Personen, immobile und mobile Personen. Der Indikator für immobile Personen sei $R_i = 1$. Der Gewichtungsansatz basiert auf:

$$\sum_{i \in s} \frac{R_i}{P(R_i = 1|Y_i, a_i, x_i)} a_i (Y_i - P(Y_i = 1|a_i)) = 0 \quad (5.4)$$

Hier ist $P(R_i = 1|Y_i, a_i, x_i)$ die Wahrscheinlichkeit, dass Person i immobil bleibt, wobei x_i ein Vektor von immer beobachtbaren Kovariaten ist wie z.B. Alter, Geschlecht etc.

Die Schätzung unter (5.4) ist konsistent, falls ein konsistenter Schätzer für $P(R_i = 1|Y_i, a_i, x_i)$ gefunden wird, vgl. Basic/Rendtel (2006). Im vorliegenden Fall bedeutet dies, dass die Mobilitätswahrscheinlichkeit von den Kovariaten zum Zeitpunkt $t - 1$, also a_i und x_i , und nicht zum Zeitpunkt t , also Y_i , abhängen darf, da Y_i für die mobilen Personen nicht beobachtet wird. Es muss also gelten: $P(R_i = 1|Y_i, a_i, x_i) = P(R_i = 1|a_i, x_i)$ – dies entspricht der MAR-Annahme.

5.3.2.1 Beispiel

Die Schätzung der (Im-)mobilitätswahrscheinlichkeiten erfolgt über ein Logit-Modell, wobei als Kovariate Alter, Haushaltsgröße, Haushaltseinkommen, Region, Schulbildung, Berufsbildung, Arbeitsmarktzustände und Nationalität benutzt wurden, vgl. Basic/Rendtel (2006). Jede Person i bekommt dann ein Gewicht in Höhe von $(P(R_i = 1|a_i, x_i))^{-1}$.

Tabelle 5.5 gibt die Ergebnisse der Schätzung eines (Im-)mobilitätsmodells wieder. Bei dem Modell in der Tabelle 5.5 handelt es sich um ein Haushaltsmodell, da die Umzugsentscheidung oft auf der Haushaltsbasis getroffen wird, vgl. Clarke/Tate (2002). Die erste Spalte in Tabelle 5.5 (MZ) zeigt die Ergebnisse für den Mikrozensus, die zweite Spalte die Ergebnisse auf Basis des SOEP und die dritte Spalte die Differenz dieser beiden Werte. In den Klammern sind die Standardabweichungen ausgewiesen.

Tabelle 5.5: (Im-)mobilitätsmodell: Zeitintervall 1996-1998

Variable		MZ	SOEP	Diff.
Konstante		1.5548 (0.1147)	1.8182 (0.3018)	-0.2635 (0.3229)
HH-Grösse	1 Person	-1.2111 (0.0614)	-1.1797 (0.1608)	-0.0313 (0.1721)
	3 Personen	0.7824 (0.0612)	0.6776 (0.1527)	0.1048 (0.1645)
	4 Personen	1.3701 (0.0962)	1.4752 (0.2442)	-0.1051 (0.2624)
	≥ 5 Personen	1.4375 (0.1126)	1.6101 (0.2792)	-0.1727 (0.3011)
Alter ≤ 30	1 Person	-1.4298 (0.0320)	-1.3827 (0.0935)	-0.0471 (0.0988)
	≥ 2 Personen	-2.2189 (0.0496)	-2.1525 (0.1258)	-0.0665 (0.1352)
Alter > 60	1 Person	0.6612 (0.0436)	0.8264 (0.1472)	-0.1653 (0.1535)
	≥ 2 Personen	0.4538 (0.0781)	0.6598 (0.2586)	-0.2059 (0.2701)
HH-Einkommen	n/a	-0.1584 (0.0449)	-0.0132 (0.1692)	-0.1452 (0.1750)
	< 2 200 DM	0.2453 (0.0389)	0.2656 (0.1106)	-0.0203 (0.1172)
	2 200 < 3 000 DM	0.2157 (0.0403)	0.2519 (0.1093)	-0.0362 (0.1165)
	4 000 < 5 500 DM	-0.1540 (0.0415)	-0.1040 (0.0954)	-0.0500 (0.1040)
	≥ 5 500 DM	-0.2532 (0.0463)	-0.1944 (0.1113)	-0.0588 (0.1206)
Region	Ost-Deutschland	-0.3224 (0.0279)	-0.3881 (0.0794)	0.0657 (0.0842)
Anzahl der Personen im HH mit ...	mittlerer Reife	0.0171 (0.0334)	0.0906 (0.0905)	-0.0735 (0.0965)
	Abitur	0.1250 (0.0255)	0.1528 (0.0758)	-0.0278 (0.0800)
Anzahl der Personen im HH mit ...	Lehre	0.1535 (0.0221)	0.1690 (0.0620)	-0.0155 (0.0658)
	Hochschulabschluss	0.1618 (0.0366)	0.0723 (0.0937)	0.0895 (0.1006)
Nationalität	≥ 1 Ausländer	-0.7096 (0.0431)	-0.2816 (0.0942)	-0.4280 (0.1037)
Erwerbstätig	1 Person	0.3964 (0.0583)	0.0769 (0.1522)	0.3195 (0.1630)
	≥ 2 Personen	0.6222 (0.0990)	0.2395 (0.2460)	0.3827 (0.2651)
Arbeitslos	1 Person	0.0908 (0.0616)	-0.0020 (0.1460)	0.0928 (0.1585)
	≥ 2 Personen	-0.2181 (0.1530)	-0.5496 (0.3458)	0.3315 (0.3781)
Nichterwerbsperson	1 Person	0.1943 (0.0564)	0.0430 (0.1438)	0.1513 (0.1545)
	≥ 2 Personen	0.2429 (0.1044)	0.0296 (0.2597)	0.2132 (0.2799)
Beobachtungen		53 821	6 777	

Abhängige Variable: Indikator für Mobilität, Standardabweichung in Klammern
Koeffizienten für $\log(P(\text{Immo})/P(\text{Mob}))$

Hinsichtlich der Parameterinterpretation zeigen die Ergebnisse in Tabelle 5.5, dass Haushaltsgröße, Alter und Nationalität den größten Einfluss auf die (Im-)mobilitätsneigung haben. Je weniger Personen im Haushalt, desto mobiler ist der Haushalt bzw. je mehr ältere Personen in einem Haushalt leben, desto immobiler ist der Haushalt. Auch Haushalte mit ausländischen Mitgliedern haben eine höhere (Im-)mobilitätswahrscheinlichkeit. Die restlichen Kovariaten haben einen geringeren numerischen Einfluss auf die (Im-)mobilitätswahrscheinlichkeit. Dies gilt sowohl für den Mikrozensus als auch für das SOEP. Nebenbei wurde auch ein formaler Test in Bezug auf die Gleichheit der Koeffizienten zwischen den beiden Erhebungen durchgeführt. Der Test führte nur bei zwei Merkmalen (Anzahl der Ausländer im Haushalt ist größer gleich 1 und Anzahl der erwerbstätigen Personen im Haushalt ist größer gleich 1) zu einer signifikanten Abweichung von Mikrozensus und SOEP. Aus diesem Grund scheint auch die Abschätzung des Noncoverage-Bias durch das SOEP eine vernünftige Näherung für den Mikrozensus zu sein.

Um die Reduzierung des Bias durch die Gewichtung der Beobachtungen mit dem Kehrwert der Immobilitätswahrscheinlichkeit abzuschätzen, wird jedem Mitglied eines immobilen Haushalts das gleiche Gewicht $(P(R_i = 1|a_i, x_i))^{-1}$ und jedem Mitglied eines mobilen Haushalts ein Gewicht von Null zugewiesen. Tabelle 5.6 zeigt die Auswirkung des Gewichtungsansatzes auf die Biasreduktion am Beispiel der Übergänge Arbeitslosigkeit/Nichterwerbstätigkeit (1996) nach Erwerbstätigkeit (t=1997, 1998, 1999). Um die Auswirkung der Gewichtungsfaktoren auf die Biasreduktion im MZ-Panel abschätzen zu können, wird das Verhältnis von Bias zu Biaskorrektur berechnet. Hier wird angenommen, dass der Bias im Mikrozensus und im SOEP dieselbe Größe hat.

Tabelle 5.6: Verhältnis von Biaskorrektur zu Bias für die Benutzung der Gewichtungsvariablen über die Immobilitätswahrscheinlichkeit. Datenbasis: SOEP und MZ-Panel

t	Bias	Biaskorrektur/Bias	
		SOEP	MZ
Übergang Arbeitslos→Erwerbstätig			
1997	1.98	0.31	0.47
1998	3.13	0.35	0.68
1999	3.91	0.81	0.49
Übergang Nichterwerbsperson→Erwerbstätig			
1997	1.10	0.40	0.31
1998	3.59	0.31	0.31
1999	4.56	0.56	0.35

Die Spalte Bias in Tabelle 5.6 zeigt den absoluten Bias, $\hat{B} = \hat{P}(B|A)_{\text{SOEP}} - \hat{P}(B|A, R = 1)_{\text{SOEP}}$, berechnet anhand der SOEP Daten, vgl. Tabelle 5.1. Die Spalten SOEP bzw. MZ zeigen das Verhältnis von Bias zu Biaskorrektur im SOEP und im Mikrozensus. Hinsichtlich der Interpretation des Verhältnisses von Bias zu Biaskorrektur unterscheidet man folgende Fälle:

$$\frac{\hat{P}(B|A)_{\text{Gew,MZ}} - \hat{P}(B|A)_{\text{Imm,MZ}}}{\hat{B}} = \begin{cases} < 1 & \text{teilweise Korrektur bis Verschlechterung} \\ 1 & \text{vollständige Korrektur} \\ > 1 & \text{Überkorrektur} \end{cases}$$

Die Ergebnisse in Tabelle 5.6 zeigen, dass für beide Übergänge und für alle Zeitpunkte eine Reduktion des Bias stattfindet.

5.3.2.2 Makro

Für den Gewichtungsansatz wurde kein Makro implementiert, da bei den meisten Programmen die Zuweisung der Gewichte mit einem einzigen Statement erfolgt. Bei SAS und SPSS lautet dieser Befehl WEIGHT und wird üblicherweise vor dem MODEL Statement geschrieben. Allerdings ist hier zu beachten, dass diese Gewichte sowohl bei SAS als auch bei SPSS als konstant gelten, obwohl sie durch ein Logit-Modell geschätzt wurden und selbst zufällig sind. Die Folge ist, dass bei dem analysierten Modell die Varianz der geschätzten Koeffizienten etwas unterschätzt wird.

Weiterhin ist zu beachten, dass bei Nutzung des Datenanalysesystems SPSS die Gewichte auf die Anzahl der befragten Personen "normiert" werden muss, da SPSS für die Ermittlung von Teststatistiken generell die gewichteten Fallzahlen heranzieht.

Für die Normierung muss hierzu lediglich jeder Gewichtungsfaktor eines jeden i -ten Datensatzes (GF_i) mit dem Verhältnis von ungewichteter (n) und gewichteter (n_{GF}) Fallzahl multipliziert werden. Damit erhält man einen neuen Gewichtungsfaktor (nGF_i), der zwar die Struktur der Gewichtung beibehält, die Fallzahl aber auf die Anzahl der befragten Personen abstellt:

$$nGF_i = GF_i * \frac{n}{n_{GF}}$$

5.3.3 Schätzung eines Modells für Kontingenztabellen mit fehlenden Werten

Für den Fall, dass die räumliche Mobilität von den nicht beobachteten Veränderungsdaten abhängt, werden spezielle loglineare Modelle geschätzt, vgl. Basic/Rendtel (2006) sowie Baker/Laird (1988). Es wird zunächst die Situation von Übergängen zwischen zwei diskreten Merkmalen A (=Zustand zum Startzeitpunkt)

und B (=Zustand zum Endzeitpunkt) analysiert. Zu diesen beiden Merkmalen gibt es zwei Respondenzindikatoren R und S , die anzeigen, ob A beobachtet wurde ($R = 1$) oder B beobachtet wurde ($S = 1$). Für die Immobilien ($R = 1, S = 1$) wird die Tabelle $A \times B$ beobachtet, während bei den Fortgezogenen ($R = 1, S = 0$) nur die Randtabelle A beobachtet wird. Zugezogene ($R = 0, S = 1$) liefern lediglich eine Randtabelle für B . Die Spezifikation des loglinearen Modells erstreckt sich entsprechend sowohl auf die inhaltlichen Variablen, A und B , als auch auf die Respondenzindikatoren. Bei der loglinearen Analyse wird ein Modell der Form

$$P(A, B, R, S) = P(A, B)P(R, S|A, B) \quad (5.5)$$

spezifiziert. Allerdings ist das Modell in (5.5) überparametrisiert, da die Zahl der Modellparameter größer ist als die Zahl der beobachteten Zellen. Ein unrestringiertes Modell für $P(A, B, R, S)$ enthält $3K^2 - 1$ Parameter, wenn K die Anzahl der Ausprägungen von A bzw. B ist.⁸ Dem stehen jedoch nur $K^2 + 2K$ Beobachtungszellen gegenüber. Da es sich bei A und B um die wiederholten Messungen desselben Merkmals handelt, soll $P(A, B)$ nicht restringiert werden. Es sollen Restriktionen auf die Mobilitätswahrscheinlichkeiten $P(R, S|A, B)$ gesetzt werden. Die Restriktion

$$P(R = 1, S = 0|A, B) = P(R = 0, S = 1|A, B) \quad (5.6)$$

besagt, dass die Wahrscheinlichkeit von Fortzügen und Zuzügen gleich ist. Diese Annahme erscheint gerechtfertigt, da jeder Fortzug auch ein Zuzug ist und die Unterscheidung nur nach der zufälligen Auswahl des Auswahlbezirks im Mikrozensus getroffen wird. Dies senkt die Anzahl der Modellparameter auf $2K^2 - 1$. Für $K = 3$ sind damit noch mindestens zwei weitere Restriktionen zu setzen, um das Modell identifizieren zu können. Hierbei kann man wiederum das SOEP nutzen. Allerdings werden im SOEP die Zuzüge nicht vollständig erfasst, so dass nur Restriktionen hinsichtlich der Fortgezogenen empirisch überprüft werden können, d.h. im SOEP wird das Modell $P(R|A, B)$ spezifiziert, wobei jetzt A immer beobachtet wird und R anzeigt, ob B beobachtet wurde ($R = 1$) oder nicht⁹. Hinsichtlich des Modells $P(R|A, B)$ ist zu spezifizieren, ob es sich bei der Mobilitätsursache um eine ignorierbare¹⁰ oder nicht-ignorierbare¹¹ Ursache handelt. Die Mobilität wäre ignorierbar, falls kein statistischer Zusammenhang zwischen R und B existiert.

Da im SOEP die Analysevariablen A und B sowohl für Immobiler als auch für Fortgezogene vorliegen, ist es möglich, solche Hypothesen zu testen. Die dabei erzielten Ergebnisse können dann als Restriktion bei der Analyse des Mikrozensus-Panels übernommen werden.

Der beschriebene Ansatz kann auf die Betrachtung von mehreren Wellen erweitert werden. Dies führt auf die Schätzung von Verläufe. Hierbei vergrößert sich

⁸Der Fall $R = S = 0$ wird ignoriert.

⁹Im SOEP werden nur Zuzüge in bereits existierende Haushalte erfasst und keine Zuzüge der gesamten Haushalte.

¹⁰Im Sinne von MAR

¹¹Im Sinne von NMAR

die Zahl der beobachteten unvollständigen Kontingenztabelle. Gleichzeitig steigt die Zahl der Modellparameter. Da es sich bei dem MZ um ein Vier-Jahres-Panel handelt, wird die Spezifikation eines loglinearen Modells für vier Zeitpunkte beschrieben.

Bezeichnet man als A , B , C und D die Messungen eines Merkmals in vier aufeinander folgenden Jahren und als R , S , U und V die zugehörigen Respondenzindikatoren, dann wird das folgende loglineare Modell formuliert:

$$P(A, B, C, D, R, S, U, V) = P(A, B, C, D)P(R, S, U, V|A, B, C, D) \quad (5.7)$$

Um das Modell in (5.7) schätzen zu können, werden Restriktionen benötigt. Hier kann man die Markov-Annahme erster Ordnung, d.h. der Zustand zum Zeitpunkt t ist nur vom Zustand zum Zeitpunkt $t - 1$ abhängig

$$P(A, B, C, D) = P(A)P(B|A)P(C|B)P(D|C) \quad (5.8)$$

und Annahme

$$P(R, S, U, V|A, B, C, D) = P(R|A)P(S|A, B, R)P(U|B, C, S)P(V|C, D, U) \quad (5.9)$$

sowie die Homogenitätsannahme

$$P(S|A, B, R = 0) = P(U|B, C, S = 0) = P(V|C, D, U = 0) \quad (5.10)$$

nutzen. Hierbei besagt die Restriktion auf der rechten Seite von Gleichung (5.9), dass die Wahrscheinlichkeit für einen Umzug in Welle t von Mobilitätsverhalten und Merkmalen in den Wellen $t - 1$ und t abhängt. Für bereits umgezogene Haushalte ist diese Wahrscheinlichkeit gleich 1. Das Modell berücksichtigt die Möglichkeit, dass Mobilität in Abhängigkeit von Merkmalsübergängen auftreten kann.

5.3.3.1 Beispiel

Im Falle der Modellierung von Übergängen zwischen verschiedenen Erwerbszuständen deuten die Ergebnisse aus Tabelle 5.6, dass die fehlenden Daten nicht zufällig, also NMAR sind. Diese Vermutung lässt sich durch den Rückgriff auf die eben vorgestellte loglineare Modelle relativ einfach überprüfen. Es wird zuerst ein Modell der Form $P(A, B)P(R|A)$ spezifiziert. Bei diesem Modell handelt es sich um ein MAR Modell, da die Mobilität nur von den beobachteten Daten abhängt, d.h. R hängt nur von A ab. Allerdings erreicht das Modell keine ausreichende Modellanpassung. Diese Fehlanpassung beruht dabei speziell auf einer Fehlspezifikation der Mobilitätsgründe. Es zeigte sich, dass die Mobilität sowohl von A als auch von B abhängt, d.h. $P(R|AB)$. Bei diesem Modell handelt es sich um ein NMAR Modell. Weiterhin konnten noch drei Gruppen von Personen mit der gleichen Mobilitätswahrscheinlichkeit identifiziert werden. Es wurden folgende Restriktionen gesetzt:

- a) Die Mobilitätswahrscheinlichkeit aus den Erwerbszuständen E und U ist für alle Personen gleich (mit Ausnahme des Übergangs $U \rightarrow E$). Diese Gruppe umfasst alle Fälle mit niedriger Mobilität.
- b) Die Mobilitätswahrscheinlichkeit für die Übergänge $U \rightarrow E$, $N \rightarrow E$ und $N \rightarrow U$ ist für alle Personen gleich. Diese Gruppe umfasst alle Fälle mit hoher Mobilität.
- c) Die Mobilitätswahrscheinlichkeit $N \rightarrow N$ wird nicht restringiert.

Um zu prüfen, ob die loglineare Modellierung zu einer Biasreduktion führt, wird das Verhältnis von Bias zu Biaskorrektur berechnet. Tabelle 5.7 zeigt das Verhältnis von Bias zu Biaskorrektur im SOEP und im Mikrozensus.

Tabelle 5.7: Verhältnis von Biaskorrektur zu Bias bei Benutzung eines loglinearen Modells. Datenbasis: SOEP und MZ-Panel

t	Bias	Biaskorrektur/Bias	
		SOEP	MZ
Übergang Arbeitslos→Erwerbstätig			
1997	1.98	0.20	0.76
1998	3.13	0.15	0.80
1999	3.91	0.07	1.10
Übergang Nichterwerbsperson→Erwerbstätig			
1997	1.10	0.52	1.35
1998	3.59	0.33	0.94
1999	4.56	0.35	1.36

Es ist schwierig zu prüfen, ob die Anwendung dieses NMAR-Modells zu einer Biasreduktion führt, da im SOEP keine Randtabelle für die Zugezogenen beobachtet wird. Aus diesem Grund wurde nur ein vereinfachtes Modell mit einem einzigen Ausfallindikator für die Verluste durch Mobilität geschätzt. Zwar erfolgt die Korrektur der Schätzwerte in die Richtung des wahren Wertes unter Einschluss aller Personen, der Bias wird jedoch lediglich halbiert, vgl. Spalte SOEP in Tabelle 5.7. Diese schwache Reaktion der Modellschätzung gegenüber der Schätzung allein auf Basis der Immobilen ändert sich sofort, wenn man im Mikrozensus die Gruppe der Zugezogenen mit berücksichtigt, vgl. Spalte MZ in Tabelle 5.7. Im Mittel liegt die Korrektur bei 1.05, also gar nicht so weit entfernt vom Idealwert 1.

Tabelle 5.8 zeigt die Ergebnisse für das 4-Zeitpunkte loglineare Modell. Für die Schätzung des Modells wurden die Restriktionen in den Gleichungen (5.8), (5.9) und (5.10) gesetzt. Zusätzlich wurden für die Umzugswahrscheinlichkeiten in Gleichung (5.9) wieder dieselben Restriktionen wie im Zwei-Wellen-Modell gesetzt.

Tabelle 5.8: Verhältnis von Biaskorrektur zu Bias bei Benutzung eines loglinearen Modells. Datenbasis: SOEP und MZ-Panel

t	Bias	Biaskorrektur/Bias	
		SOEP	MZ
Übergang Arbeitslos→Erwerbstätig			
96→97	4.93	0.35	1.21
97→98	3.24	-0.36	1.28
98→99	3.24	0.15	2.12
Übergang Nichterwerbsperson→Erwerbstätig			
96→97	2.84	0.20	1.26
97→98	2.55	-0.26	1.33
98→99	2.26	0.04	2.69

Im Falle des Mikrozensus wird für die Übergänge 96→97 und 97→98 eine fast perfekte Biaskorrektur erreicht. Lediglich für den Wechsel 98→99 erhält man eine Überkorrektur des Wechsels von U→E und N→E. Im Falle des SOEP ist die Biaskorrektur sehr gering. Dies ist auf das Fehlen der zusätzlichen Informationen über die Zugezogenen zurückzuführen.

5.3.3.2 IEM Programme

Alle loglinearen Modelle wurden mit dem Statistikprogramm ℓ_{EM} geschätzt. Im Folgenden werden einige ℓ_{EM} Programme beschrieben, mit denen die Ergebnisse in den Tabellen 5.7 und 5.8 berechnet wurden. Für die Entscheidung, ob die Mobilität von Merkmal A , B oder A und B abhängt, wurden folgende ℓ_{EM} Modelle spezifiziert:

Modell 1:

```

1 res 1 * Anzahl der Responsevariablen
2 man 2 * Anzahl der Variablen
3 dim 2 3 3 * Anzahl der Ausp. von res und man
4 lab R A B * Labels von res und man
5 sub AB AB * Teiltabellen
6 mod A B|A {AB} R|AB {RA} * Modellspezifikation: R haengt von A ab
7 dat [6262 334 273 273 441 171 260 111 1863 * Tabelle der Immobilien
8 745 45 39 54 41 16 49 22 120] * Tabelle der Mobilien

```

Die Zahlen zu Beginn der einzelnen Programmzeilen gehören nicht zum ℓ_{EM} -Programm. Sie dienen allein der Numerierung der einzelnen Programmzeilen, um sie besser kommentieren zu können. ℓ_{EM} -Befehle bestehen aus einem Befehlswort

(z.B. `res` bzw. `man` in Zeilen 1 und 2), die klein und genau so geschrieben werden müssen, wie es in der Programmbeschreibung steht. Alle Programmzeilen, die mit einem Stern beginnen, werden als Kommentar betrachtet und beeinflussen die Programmverarbeitung nicht. Im Modell 1 wird in der ersten Zeile mit dem Befehl `res` eine Responsvariable spezifiziert, die anzeigt, ob B beobachtet wird oder nicht. In diesem Beispiel wird A immer beobachtet. Die zweite Zeile `man` spezifiziert die Anzahl der Untersuchungsvariablen. Mit dem Befehl `dim` in der dritten Zeile wird angegeben, wieviele Ausprägungen die Response- bzw. die Untersuchungsvariablen haben, zuerst die Ausprägungen der Responsevariablen dann die der Untersuchungsvariablen. Mit dem Befehl `lab` in der vierten Zeile können Labels bestehend aus jeweils einem Buchstaben für die drei Variablen spezifiziert werden. Der Befehl `sub` (steht für `subgroups`) in der fünften Zeile spezifiziert die zu verwendenden Daten je nach verschiedenen Ausprägungen der Responsevariable. Da sich das Modell 1 auf die SOEP Daten bezieht, d.h. man beobachtet sowohl für die immobilen als auch für die mobilen Personen die volle Tabelle, werden hier zwei Tabellen AB benutzt. Das eigentliche Modell spezifiziert man mit dem Befehl `mod`. Zunächst ist die Tabelle zu benennen, die analysiert werden soll; hier AB und RAB. Dann wird in geschweiften Klammern eine loglineare Modellformel angegeben. In diesem Fall ein saturiertes Modell für AB Beziehung und ein restringiertes Modell für RAB Beziehung (R hängt nur von A ab). Dies ist ein MAR-Modell, da die Mobilitätswahrscheinlichkeit nur von A abhängt und A wird immer beobachtet. Mit dem Befehl `dat` kann man entweder die zu analysierenden Häufigkeiten einzeln aufzählen oder auf eine Datei verweisen, die genau diese Häufigkeiten enthält. In dem Modell 1 wird von der ersten Möglichkeit gebrauch gemacht. Es müssen so viele Häufigkeiten angegeben werden, wie die Tabelle Dimensionen hat. In diesem Fall also $2 \times 3 \times 3 = 18$ Häufigkeiten, die in der oben beschriebenen Reihenfolge genannt werden müssen. Im ℓ_{EM} Output werden folgende Informationen ausgewiesen:

- Wiedergabe der Programmzeilen
- Statistiken über den Schätzalgorithmus (Anzahl Iterationen, Likelihood-Statistiken usw.)
- Beobachtete und geschätzte Häufigkeiten
- Schätzwerte für die (effektkodierten) log-linearen Parameter und die (un)bedingten Wahrscheinlichkeiten

Der ℓ_{EM} -Output des Modells 1 sieht dann wie folgt aus:

```

1 LEM: log-linear and event history analysis with missing data.
2 Developed by Jeroen Vermunt (c), Tilburg University, The Netherlands.
3 Version 1.0 (September 18, 1997).
4
5
6 *** INPUT ***
7
```

```

8   res 1                * Anzahl der Responsevariablen
9   man 2                * Anzahl der Variablen
10  dim 2 3 3           * Anzahl der Ausp. von res und man
11  lab R A B           * Labels von res und man
12  sub AB AB           * Teiltabellen
13  mod A B|A {AB} R|AB {RA} * Modellspezifikation: R haengt von A ab
14  dat [6262 334 273 273 441 171 260 111 1863 * Tabelle der Immobilien
15        745 45 39 54 41 16 49 22 120] * Tabelle der Mobilien
16
17
18 *** STATISTICS ***
19
20   Number of iterations = 2
21   Converge criterion   = 0.0000000000
22   Seed random values   = 5435
23
24   X-squared            = 65.7546 (0.0000)
25   L-squared           = 56.9971 (0.0000)
26   Cressie-Read        = 62.4106 (0.0000)
27   Dissimilarity index = 0.0114
28   Degrees of freedom  = 6
29   Log-likelihood       = -14165.90536
30   Number of parameters = 11 (+1)
31   Sample size          = 11119.0
32   BIC(L-squared)       = 1.0986
33   AIC(L-squared)       = 44.9971
34   BIC(log-likelihood)  = 28434.2912
35   AIC(log-likelihood)  = 28353.8107
36
37   Eigenvalues information matrix
38     9227.7780  6134.6675  4975.9420  3539.4301  2111.6534  1703.6417
39     1630.8989  800.6946  570.5893  546.9570  480.8274
40
41
42
43 *** FREQUENCIES ***
44
45 * SUBGROUP R 1 *
46
47   A B      observed  estimated  std. res.
48   1 1      6262.000  6252.414   0.121
49   1 2      334.000   338.185  -0.228
50   1 3      273.000   278.401  -0.324
51   2 1      273.000   290.557  -1.030
52   2 2      441.000   428.283   0.614
53   2 3      171.000   166.160   0.376
54   3 1      260.000   284.662  -1.462
55   3 2      111.000   122.525  -1.041
56   3 3      1863.000  1826.813   0.847
57
58   X-squared = 5.6882, L-squared = 5.8017
59
60 * SUBGROUP R 2 *
61
62   A B      observed  estimated  std. res.
63   1 1      745.000   754.586  -0.349
64   1 2      45.000    40.815   0.655
65   1 3      39.000    33.599   0.932
66   2 1      54.000    36.443   2.908
67   2 2      41.000    53.717  -1.735
68   2 3      16.000    20.840  -1.060

```

```

69 3 1      49.000      24.338      4.999
70 3 2      22.000      10.475      3.561
71 3 3     120.000     156.187     -2.896
72
73 X-squared = 60.0664, L-squared = 51.1954
74
75
76
77 *** PSEUDO R-SQUARED MEASURES ***
78
79 * P(B|A) *
80
81          baseline      fitted      R-squared
82 entropy              0.8083      0.4721      0.4159
83 qualitative variance  0.2348      0.1199      0.4896
84 classification error  0.3126      0.1481      0.5262
85 -2/N*log-likelihood  1.6166      0.9442      0.4159/0.4021
86 likelihood^(-2/N)    5.0357      2.5706      0.4895/0.6108
87
88 * P(R|A) *
89
90          baseline      fitted      R-squared
91 entropy              0.3288      0.3280      0.0026
92 qualitative variance  0.0914      0.0912      0.0016
93 classification error  0.1017      0.1017      0.0000
94 -2/N*log-likelihood  0.6577      0.6560      0.0026/0.0017
95 likelihood^(-2/N)    1.9303      1.9270      0.0017/0.0035
96
97 *** LOG-LINEAR PARAMETERS ***
98
99 * TABLE A [or P(A)] *
100
101 effect      beta      std err      z-value      exp(beta)      Wald      df      prob
102 A
103 1          1.0667      0.0147      72.730      2.9058
104 2          -0.9783      0.0225     -43.471      0.3760
105 3          -0.0884      0.0154      5295.83      0.9154      2 0.000
106
107 * TABLE AB [or P(B|A)] *
108
109 effect      beta      std err      z-value      exp(beta)      Wald      df      prob
110 B
111 1          0.5760      0.0241      23.901      1.7788
112 2          -0.5481      0.0282     -19.435      0.5781
113 3          -0.0279      0.0275      640.85      0.9725      2 0.000
114 AB
115 1 1        1.4336      0.0286      50.115      4.1939
116 1 2        -0.3594      0.0362     -9.930      0.6981
117 1 3        -1.0742      0.0362     -29.388      0.3416
118 2 1        -0.5190      0.0362     -14.346      0.5951
119 2 2         0.9930      0.0376      26.407      2.6994
120 2 3        -0.4740      0.0376     -12.608      0.6225
121 3 1        -0.9146      0.0376     -24.324      0.4007
122 3 2        -0.6336      0.0376     -16.851      0.5307
123 3 3         1.5482      0.0376      41.174      4.7031      4 0.000
124
125 * TABLE RA [or P(R|A)] *
126
127 effect      beta      std err      z-value      exp(beta)      Wald      df      prob
128 R
129 1          1.1083      0.0218      50.737      3.0292

```

```

130      2      -1.1083                                0.3301  2574.25   1  0.000
131      RA
132      1  1      -0.0510   0.0243   -2.101      0.9502
133      1  2      -0.0703   0.0364   -1.933      0.9321
134      1  3       0.1213                                1.1290
135      2  1       0.0510                                1.0524
136      2  2       0.0703                                1.0728
137      2  3      -0.1213                                0.8857   17.87   2  0.000
138
139
140
141 *** (CONDITIONAL) PROBABILITIES ***
142
143 * P(A) *
144
145      1              0.6923 (0.0044)
146      2              0.0896 (0.0027)
147      3              0.2181 (0.0039)
148
149 * P(B|A) *
150
151      1 | 1          0.9102 (0.0033)
152      2 | 1          0.0492 (0.0025)
153      3 | 1          0.0405 (0.0022)
154      1 | 2          0.3283 (0.0149)
155      2 | 2          0.4839 (0.0158)
156      3 | 2          0.1878 (0.0124)
157      1 | 3          0.1274 (0.0068)
158      2 | 3          0.0548 (0.0046)
159      3 | 3          0.8177 (0.0078)
160
161 * P(R|A) *
162
163      1 | 1          0.8923 (0.0035)
164      1 | 2          0.8886 (0.0100)
165      1 | 3          0.9212 (0.0055)
166      2 | 1          0.1077 (0.0035)
167      2 | 2          0.1114 (0.0100)
168      2 | 3          0.0788 (0.0055)

```

Wie aus dem ℓ_{EM} -Output zu ersehen ist, erreicht das Modell 1 keine ausreichende Modellanpassung ($L^2 = 56.99$, $df = 6$, $p = 0.00$). Diese Fehlanpassung beruht auf einer Fehlspezifikation der Mobilitätsgründe. Insofern ist hier anzunehmen, dass die zum zweiten Zeitpunkt fehlende Werte von diesen Werten bzw. von diesen und den Werten zum ersten Zeitpunkt abhängen.

Modell 2:

```

1 res 1
2 man 2
3 dim 2 3 3
4 lab R A B
5 sub AB AB
6 mod A B|A {AB} R|AB {RB}
7 dat mob.dat

```

Im Unterschied zum Modell 1 handelt es sich beim Modell 2 um ein NMAR-Modell, da die Mobilitätswahrscheinlichkeit von B abhängt und nicht immer beobachtet wird. Im Gegensatz zum Modell 1 wird hier zudem die Datentabelle aus der externen Datei `mob.dat` eingelesen. Allerdings erreicht auch dieses Modell keine ausreichende Modellanpassung ($L^2 = 39.21$, $df = 6$, $p = 0.00$). Somit ist anzunehmen, dass die Mobilität von den Merkmalsübergängen abhängt, d.h. sowohl von A als auch von B .

Modell 3:

```

1 res 1
2 man 2
3 dim 2 3 3
4 lab R A B
5 sub AB AB
6 mod A B|A {AB} R|AB
7 {fac(ABR,3)} des [0 0 0 0 0 0 0 0 0
8                   1 1 1 2 1 1 2 2 3]
9 dat mob.dat

```

Beim Modell 3 werden zusätzliche Restriktionen bezüglich der Tabelle RAB gesetzt. Hierbei bedeutet der Befehl `fac(ABR,3)`, dass drei Restriktionen an die Parameter des Modells $P(R|AB)$ gesetzt worden sind. In der Zeile 7 wird mit `des` eine Designmatrix mit entsprechenden Restriktionen spezifiziert. Hierbei bedeutet 0, dass keine Restriktion gesetzt wird. Alle Parameter mit derselben Zahl, z.B. 1, werden gleichgesetzt. Im obigen Programm werden folgende Parameter gleichgesetzt:

Restriktion 1:

$$\begin{aligned}
 P(R = 0|A = 1, B = 1) &= P(R = 0|A = 1, B = 2) = P(R = 0|A = 1, B = 3) \\
 &= P(R = 0|A = 2, B = 2) = P(R = 0|A = 2, B = 3)
 \end{aligned} \tag{5.11}$$

Restriktion 2:

$$P(R = 0|A = 2, B = 1) = P(R = 0|A = 3, B = 1) = P(R = 0|A = 3, B = 2) \tag{5.12}$$

Bei der ersten Restriktion handelt es sich um die Gruppe von Personen, die eine hohe Mobilität aufweisen, während sich bei der zweiten Restriktion um die Gruppe von Personen mit niedriger Mobilität handelt. Im Unterschied zu den beiden obigen Modellen (Modell 1 und Modell 2) erreicht das Modell 3 eine ausreichende Modellanpassung ($L^2 = 5.07$, $df = 6$, $p = 0.53$).

Die drei vorstehenden Modellen beziehen sich auf SOEP Daten und dienen zum einen zur Entscheidung, ob die fehlende Daten zufällig oder nicht zufällig zustande kommen und zum anderen zur Herleitung von Restriktionen. Diese Restriktionen können dann für den Mikrozensus übernommen werden. Da im Mikrozensus sowohl A als auch B fehlende Werte aufweisen können, hat die Response Variable R nun drei Ausprägungen.

Modell 4:

```

1 res 1
2 man 2
3 dim 3 3 3
4 lab R A B
5 sub AB A B
6 mod A B|A {AB} R|AB
7 {fac(ABR,3)} des [0 0 0 0 0 0 0 0 0
8                   1 1 1 2 1 1 2 2 3
9                   1 1 1 2 1 1 2 2 3]
10
11 dat [40020 1862 2296 1323 2358 836 2412 693 15291
12       6161 824 2181
13       5401 900 1969]

```

* Tabelle der Immobilien
* Tabelle der Fortgezogenen
* Tabelle der Zugezogenen

Es wurden die selben Restriktionen wie im Modell 3 gesetzt. Zusätzlich wurde die Restriktion Fortzug=Zuzug gesetzt (siehe design matrix des):

```

1 1 1 2 1 1 2 2 3
1 1 1 2 1 1 2 2 3.

```

Das Modell hat allerdings keine ausreichende Modellanpassung wie der Likelihood-Statistik entnommen werden kann ($L^2 = 150.01$, $df = 3$, $p = 0.00$).

5.4 Weiterführende Literatur

- Baker, S.G.; Laird, N.M. (1988): *Regression analysis for categorial variables with outcome subject to nonignorable nonresponse*, in: Journal of the American Statistical Association, **83**: 62-69.
- Basic, E. (2005): *Stabilität von Ergebnissen bei unterschiedlichen Arbeitsmarkt- tabbegrenzungen*. Arbeitspapier #9, Methodenverbund "Aufbereitung und Bereitstellung des Mikrozensus als Panelstichprobe". URL: <http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Content/Wissenschaftsforum/MethodenVerfahren/Mikrozensus/Arbeitspapiere/Arbeitspapier9,property=file.pdf>
- Basic, E.; Marek, I.; Rendtel, U. (2005): *The German Microcensus as a tool for longitudinal data analysis: An evaluation using SOEP data*. Arbeitspapier #3, Methodenverbund "Aufbereitung und Bereitstellung des Mikrozensus als Panelstichprobe". URL: <http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Content/Wissenschaftsforum/MethodenVerfahren/Mikrozensus/Arbeitspapiere/Arbeitspapier3,property=file.pdf>
- Basic, E.; Rendtel, U. (2006): *Estimation strategies in the presence of non-coverage in the German Microcensus-Panel: An evaluation using SOEP*

data. Arbeitspapier #8, Methodenverbund "Aufbereitung und Bereitstellung des Mikrozensus als Panelstichprobe". URL: <http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Content/Wissenschaftsforum/MethodenVerfahren/Mikrozensus/Arbeitspapiere/Arbeitspapier8,property=file.pdf>

Clarke, P.S., Tate, P.F. (2002): *An Application of Non-Ignorable Non-Response Models for Gross Flows Estimation in the British Labour Force Survey*, in: Australian & New Zealand Journal of Statistics, **44**: 413-425.

Greene, W.H. (2000): *Econometric Analysis*, 4th. ed. New York: Macmillan.

Haisken-DeNew, J.P.; Frick, J.R. (2005): *DTC. Desktop Companion to the German Socio-Economic Panel (SOEP)*. Version 8.0 – Dec. 2005. Berlin: Deutsches Institut für Wirtschaftsforschung (DIW). URL: <http://www.diw.de/deutsch/sop/service/dtc/dtc.pdf>

Hausman, J. (1978): *Specification tests in econometrics*, in: *Econometrica*, **46**: 1251-1271

Heidenreich, H.-J. (2002): *Längsschnittdaten aus dem Mikrozensus. Basis für neue Analysemöglichkeiten*, in: Allgemeines Statistisches Archiv, **86**: 213-231.

Neukirch, T. (2002): *Nonignorable attrition and selectivity biases in the Finnish subsample of the ECHP*. CHINTEX Working Paper #5. URL: <http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Content/Wissenschaftsforum/Chintex/ResearchResults/Downloads/WorkingPaper5,property=file.pdf>

Rendtel, U. (1995): *Panelausfälle und Panelrepräsentativität*. Frankfurt/M., New York: Campus Verlag.

Robins, J., Rotnitzky, A., Zhao, L. (1995): *Analysis of semi-parametric regression models for repeated outcomes in the presence of missing data*, in: *Journal of the American Statistical Association*, **90**: 106-121.

Rubin, D. (1976): *Inference and missing data*, in: *Biometrika*, **63**: 581-592.

Vermunt, J.K. (1997a): *ℓEM: A general program for the analysis of categorical data*. URL: <http://spitswww.uvt.nl/web/fsw/mto/lem/manual.pdf>.

Vermunt, J.K. (1997b): *Log-linear models for event histories*. Thousand Oaks: Sage Publications.