

Handbuch Mikrozensus-Panel 1996-1999

Methodenverbund

Aufbereitung und Bereitstellung des Mikrozensus als Panelstichprobe

Version 0.2 – Juli 2006

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung

Deutsche
Forschungsgemeinschaft

DFG

Vorwort

Der Mikrozensus mit seinen seit 1957 jährlich vorgelegten Querschnittsergebnissen über die Bevölkerung, ihre Bildungs- und Erwerbsbeteiligung und ihre soziale Lage – auch im Familien- und Haushaltszusammenhang – hat sich als unverzichtbare Informationsquelle bewährt. Jedoch sind die von ihm aufgezeigten Nettoveränderungen für die Sozial- und Wirtschaftswissenschaften nicht hinreichend. So empfiehlt das Gutachten der Kommission zur Verbesserung der informationellen Infrastruktur die Erschließung des Mikrozensus für Längsschnittanalysen¹. Ähnliches ist auch dem Armuts- und Reichtumsbericht der Bundesregierung aus dem Jahre 2001 zu entnehmen².

Bereits vor diesen Forderungen wurden von der Wissenschaft und der amtlichen Statistik die Notwendigkeit einer Längsschnittaufbereitung des Mikrozensus gesehen und erste Vorarbeiten geleistet, die in das Verbundprojekt "Aufbereitung und Bereitstellung des Mikrozensus als Panelstichprobe" mündeten.

Die mit der Erschließung und Bereitstellung des Mikrozensus als Panel verbundenen Arbeiten sind aufgrund der Tatsache, dass der Mikrozensus auf einer Flächenstichprobe basiert, komplex und sehr umfangreich. Diese Entwicklungsarbeiten können die statistischen Ämter oder einzelne Forscher nicht aus eigenen Mitteln bestreiten. Ein von amtlicher Statistik und Wissenschaft gemeinsam durchgeführtes Projekt ist daher unerlässlich und hat die Vorteile, dass in den Arbeiten nicht nur spezifische Erfahrungen der Forschung, sondern auch die Interessen der künftigen Nutzer/innen eines Scientific Use Files des Mikrozensus-Panels von Beginn an einfließen können. Es wurde daher ein Forschungsprojekt beim Bundesministerium für Bildung und Forschung sowie der Deutschen Forschungsgemeinschaft beantragt. Ziel des Verbundprojektes ist die Aufbereitung des Mikrozensus als Mikrozensus-Panel und die Bereitstellung eines Mikrozensus-Panels als Scientific Use File.

Als Kooperationspartner von Seiten der amtlichen Statistik waren an dem Projekt das Statistische Bundesamt sowie das Landesamt für Datenverarbeitung und Statistik Nordrhein-Westfalen, von Seiten der Wissenschaft der Fach-

¹Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik (Hrsg.) (2001): *Wege zu einer besseren informationellen Infrastruktur*. Gutachten der vom Bundesministerium für Bildung und Forschung eingesetzten Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik. Baden-Baden-Nomos; S. 26, 242, 248.

²Bundesministerium für Arbeit und Sozialordnung (BMA) (2001): *Lebenslagen in Deutschland. Erster Armuts- und Reichtumsbericht der Bundesregierung*. Bonn: BMA; S. 354.

bereich Wirtschaftswissenschaften der Freien Universität Berlin und das Zentrum für Umfragen, Methoden und Analysen in Mannheim beteiligt.

Mit dem vorliegenden Handbuch soll den Nutzern/Nutzerinnen des Scientific Use Files des Mikrozensus-Panels 1996-1999 der Umgang mit dem Datenmaterial erleichtert werden. Hierzu erfolgt 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 1996-1999 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 1996-1999 gebildet wurde,
- in welchem Umfang Datensätze längsschnittanalytisch analysiert werden können – angesprochen ist hier die Stichprobenentwicklung des Mikrozensus-Panels 1996-1999 –,
- welche Auswertungsstrategien in einem Mikrozensus-Panel herangezogen werden können,
- welche Längsschnittinformationen das Mikrozensus-Panel 1996-1999 enthält,
- wie der Datensatzaufbau des Mikrozensus-Panels 1996-1999 gestaltet ist,
- welche Erhebungs- und Analyseeinheiten in einem Mikrozensus-Panel vorliegen und in welchem Umfang sie genutzt werden können und
- welche intertemporalen Veränderungen es in Bezug der Erhebungsmerkmale innerhalb des Beobachtungszeitraum 1996 bis 1999 gegeben hat, sofern die Merkmale in das faktisch anonymisierte Mikrozensus-Panel 1996-1999 aufgenommen wurden.

Welche allgemeinen und spezifischen Schutzvorkehrungen zur Erstellung des faktisch anonymisierten Mikrozensus-Panels 1996-1999 getroffen wurden (einschließlich des Auswahlplans für die Substichprobenziehung), welche Stichprobenentwicklung im faktisch anonymisierten Mikrozensus-Panel 1996-1999 vor-

liegt und wie die Hochrechnungsfaktoren für die Analyse von Querschnittsanalysen im faktisch anonymisierten Mikrozensus-Panel 1996-1999 gebildet wurden, sind Gegenstand des dritten Kapitels. Da das vorliegende faktisch anonymisierte Mikrozensus-Panel 1996-1999 nur Personen aus Privathaushalten beinhaltet, können die im Querschnitt hochgerechneten Randverteilungen des faktisch anonymisierten Mikrozensus-Panels 1996-1999 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 1996 bis 1999 (unter Berücksichtigung der jeweiligen Hochrechnung) der Mikrozensus-Originalfiles sowie der Querschnitte des faktisch anonymisierten Mikrozensus-Panels 1996-1999 ausgewiesen. 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 Verbundprojektes wurden drei Ansätze verfolgt. 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.

Das Handbuch endet derzeit mit Methoden zur Aufdeckung und Reduktion von Verzerrungen infolge räumlicher Mobilität (Kapitel 5). Erläutert werden die Methoden beispielhaft an Erwerbsübergängen, wobei als Referenzstatistik das Sozio-ökonomische Panel herangezogen wird. Bezüglich der Biasreduktion werden drei Verfahren vorgestellt:

1. der Einbezug von Kontrollvariablen,
2. die Gewichtung mit dem Kehrwert der Immobilitätswahrscheinlichkeit und
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 Kontingenztafeln mit fehlenden Werten wurden mit dem Statistikprogramm ℓ EM durchgeführt³.

Abschließender Hinweis: Bitte beachten Sie, dass die Beispiele im vorliegenden Handbuch nicht eins zu eins reproduzierbar sind, da das der Wissenschaft zur Verfügung gestellte Scientific Use File des Mikrozensus-Panels 1996-1999 sich in einigen Punkten von dem im Verbundprojekt genutzten Arbeitsfile des Mikrozensus-Panels 1996-1999 unterscheidet. Gegenüber dem Arbeitsfile wurde im Scientific Use File

- a) ein erhöhter Unterstichprobenumfang realisiert (70% gegenüber 63,5% im Arbeitsfile),
- b) eine echte geschichtete Stichprobe anstelle einer systematischen Auswahl aus vorsortierter Datenanordnung mit Zufallsstart gezogen und
- c) die Auswahlbezirksgröße (gemessen an der durchschnittlichen Anzahl, der in einem Auswahlbezirk wohnenden Personen) und die Unterstichprobeneigenschaft explizit berücksichtigt.

³Das Statistikprogramm ℓ EM 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>.

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 1996-1999	35
2.3	Die Stichprobenentwicklung des Mikrozensus-Panels 1996-1999	37
2.4	Auswertungskonzepte in einem Mikrozensus-Paneldatensatz	43
2.5	Längsschnittinformationen im Mikrozensus-Paneldatensatz 1996-1999	48
2.6	Der Datensatzaufbau des Mikrozensus-Panels 1996-1999	53
2.7	Erhebungs- und Analyseeinheiten im Mikrozensus-Panel	54
2.8	Intertemporale Veränderungen der Erhebungsmerkmale im Erhebungszeitraum 1996 bis 1999	54
2.9	Weiterführende Literatur	59
3	Das faktisch anonymisierte Mikrozensus-Panel 1996-1999	61
3.1	Maßnahmen zur Erstellung des faktisch anonymisierten Mikrozensus-Panels 1996-1999	62
3.1.1	Allgemeine Schutzvorkehrungen	62
3.1.2	Spezifische Anonymisierungsmaßnahmen	63
3.2	Die Stichprobenentwicklung des faktisch anonymisierten Mikrozensus-Panels 1996-1999	66
3.3	Das Variablenspektrum des faktisch anonymisierten Mikrozensus-Panels 1996-1999	69
3.4	Die Bildung angepasster Hochrechnungsfaktoren für Querschnittsanalysen im Mikrozensus-Panel 1996-1999	70

3.5	Randverteilungen ausgewählter Merkmale in den Mikrozensus-Originalfiles und den Querschnitten des faktisch anonymisierten Mikrozensus-Panels 1996-1999	75
3.6	Metainformationen zum Mikrozensus-Panel 1996-1999	82
3.6.1	Das Schlüsselverzeichnis des faktisch anonymisierten Mikrozensus-Panels 1996-1999	83
3.6.2	Das systematische Verzeichnis der Wirtschaftszweige des Statistischen Bundesamtes in der Fassung für den Mikrozensus – Ausgabe 1993	86
3.6.3	Klassifikationen beruflicher Tätigkeiten	87
3.6.4	Gliederung der Hauptfachrichtung des Abschlusses an einer (Fach-)Hochschule – Ausgabe 1985	89
3.7	Weiterführende Literatur	90
4	Das Hochrechnungsverfahren für Längsschnittauswertungen im Mikrozensus-Panel 1996-1999	91
4.1	Die Grundidee der Hochrechnung	91
4.2	Logistic Regression Weighting	93
4.3	Grundgesamtheit und Stichprobe des Mikrozensus-Panels 1996-1999	96
4.4	Die Kalibrierung des Mikrozensus-Panels 1996-1999	97
4.5	Ein Beispiel	102
4.6	Die im Mikrozensus-Panel 1996-1999 bereitgestellten Hochrechnungs- und Mobilitätsfaktoren	102
4.7	Weiterführende Literatur	104
5	Möglichkeiten zur Korrektur von Verzerrungen aufgrund räumlicher Immobilität	105
5.1	Das Auftreten von räumlicher Mobilität	105
5.2	Methoden zur Biasaufdeckung	106
5.2.1	Beispiel	108
5.2.2	Die Größenordnung des Mobilitätsbias	109
5.2.3	Makro hausman.sas	110
5.3	Methoden zur Biasreduktion	110
5.3.1	Verwendung von Kontrollvariablen	111
5.3.1.1	Makro logit.sas	114
5.3.2	Gewichtung mit dem Kehrwert der Immobilitätswahrscheinlichkeit	115
5.3.2.1	Beispiel	116
5.3.2.2	Makro	119
5.3.3	Schätzung eines Modells für Kontingenztafeln mit fehlenden Werten	119
5.3.3.1	Beispiel	121

5.3.3.2 IEM Programme	123
5.4 Weiterführende Literatur	130

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 zweiten Rotationsviertel der vierten Stichprobe in den einzelnen Erhebungsjahren	29
Tabelle 2.1: Umfang plausibler und unplausibler Angaben sowie temporärer Ausfälle auf der Personenebene für den Mikrozensus-Paneldatensatz 1996-1999 nach der rein maschinellen Zusammenführung – ausgewiesen nach den einzelnen Erhebungsjahren .	39
Tabelle 2.2: Umfang plausibler und unplausibler Angaben sowie temporärer Ausfälle auf der Haushaltsebene für den Mikrozensus-Paneldatensatz 1996-1999 nach der rein maschinellen Zusammenführung – ausgewiesen nach den einzelnen Erhebungsjahren .	40
Tabelle 2.3: Umfang plausibler und unplausibler Angaben sowie temporärer Ausfälle auf der Personenebene für den Mikrozensus-Paneldatensatz 1996-1999 nach Datenkorrektur – ausgewiesen nach den einzelnen Erhebungsjahren	41
Tabelle 2.4: Umfang plausibler Angaben auf der Personenebene für den Mikrozensus-Paneldatensatz 1996-1999 vor und nach Datenkorrektur – ausgewiesen nach den einzelnen Erhebungsjahren . .	41
Tabelle 2.5: Umfang plausibler und unplausibler Angaben sowie temporärer Ausfälle auf der Haushaltsebene für den Mikrozensus-Paneldatensatz 1996-1999 nach Datenkorrektur – ausgewiesen nach den einzelnen Erhebungsjahren	42
Tabelle 2.6 : Umfang plausibler Angaben auf der Haushaltsebene für den Mikrozensus-Paneldatensatz 1996-1999 vor und nach Datenkorrekturen – ausgewiesen nach den einzelnen Erhebungsjahren .	43
Tabelle 2.7: Zusammenhang von Auswertungskonzept und Werteselektion erklärter Personensätze	48
Tabelle 2.8: Längsschnittinformationen im Mikrozensus-Panel 1996-1999	49

Tabelle 2.9: Intertemporale Veränderungen des Erhebungsmerkmals: Wichtigster Grund für weniger geleistete Arbeitsstunden als normalerweise in der Berichtswoche	55
Tabelle 2.10: Intertemporale Veränderungen des Erhebungsmerkmals: Wichtigster Grund für Arbeitsuche von erwerbstätigen Personen	56
Tabelle 2.11: Intertemporale Veränderungen des Erhebungsmerkmals: Art der aktiven Arbeitsuche nach einer Arbeitnehmertätigkeit innerhalb der letzten vier Wochen	57
Tabelle 2.12: Intertemporale Veränderungen in Bezug auf Variablen- namen	57
Tabelle 2.13: Intertemporale Veränderungen des Erhebungsmerkmals: Art des höchsten beruflichen Ausbildungs- oder (Fach-)Hochschulabschlusses	58
Tabelle 3.1: Auswahlplan der Substichprobenziehung zur Generierung eines faktisch anonymisierten Mikrozensus-Panels 1996-1999	64
Tabelle 3.2: Die Stichprobenentwicklung des faktisch anonymisierten Mikrozensus-Panels 1996-1999 auf der Personenebene	67
Tabelle 3.2: Die Stichprobenentwicklung des faktisch anonymisierten Mikrozensus-Panels 1996-1999 auf der Haushaltsebene	68
Tabelle 3.4: Auswahlwahrscheinlichkeit für eine Person aus der Un- terstichprobe des Mikrozensus-Originalfiles in die Unterstichpro- be des faktisch anonymisierten Mikrozensus-Panels 1996-1999 zu gelangen	71
Tabelle 3.5: Variablennamen und -labels der angepassten Hochrech- nungsfaktoren in den einzelnen Erhebungsjahren des Mikrozensus- Panels 1996-1999	73
Tabelle 3.6: Bevölkerung in Privathaushalten nach ausgewählten Merkmalen im MZ-Originalfile und dem Querschnitt des MZ- Panels im Erhebungsjahr 1996 mit Personengewichtung	77
Tabelle 3.7: Metainformationen zum Mikrozensus-Panel 1996-1999 . . .	82
Tabelle 3.6: Ausfallart sowie zugehöriger Variablenwert und Wertela- bel im Mikrozensus-Paneldatensatz 1996-1999	86
Tabelle 3.9: Die Gliederungsebenen der Klassifikation der Wirtschafts- zweige des Statistischen Bundesamtes in der Ausgabe von 1993 . . .	87
Tabelle 3.10: Die Gliederungsebenen der Klassifikation der Berufe des Statistischen Bundesamtes in der Ausgabe von 1992	88
Tabelle 3.11: Die Gliederungsebenen der Internationalen Standard- klassifikation der Berufe zur Verwendung innerhalb der Europäi- schen Gemeinschaft in der Ausgabe von 1988	88
Table 4.1: Erklärende Variablen für die Schätzung der Mobilitätsge- wichte im Mikrozensus-Panel 1996-1999	95

Tabelle 4.2: Variablennamen der einzelnen Längsschnitt- und Mobilitätsgewichte im Mikrozensus-Panel 1996-1999	103
Tabelle 5.1: Übergangswahrscheinlichkeiten zwischen Arbeitslosigkeit bzw. Nichterwerbsperson (1996) und Erwerbstätigkeit (t=1997, 1998, 1999). Datenbasis: SOEP	109
Tabelle 5.2: Übergangswahrscheinlichkeiten zwischen Arbeitslosigkeit bzw. Nichterwerbsperson (1996) und Erwerbstätigkeit (t=1997, 1998, 1999) bei Kontrolle nach Altersgruppe. Datenbasis: SOEP	112
Tabelle 5.3: Logistische Regression: Übergang von Arbeitslosigkeit (1996) nach Erwerbstätigkeit (1998). Datenbasis: SOEP	113
Tabelle 5.4: Logistische Regression: Übergang von Nichterwerbsperson (1996) nach Erwerbstätigkeit (1998). Datenbasis: SOEP	114
Tabelle 5.5: (Im-)mobilitätsmodell: Zeitintervall 1996-1998	117
Tabelle 5.6: Verhältnis von Biaskorrektur zu Bias für die Benutzung der GewichtungsvARIABLEN über die Immobilitätswahrscheinlichkeit. Datenbasis: SOEP und MZ-Panel	118
Tabelle 5.7: Verhältnis von Biaskorrektur zu Bias bei Benutzung eines loglinearen Modells. Datenbasis: SOEP und MZ-Panel	122
Tabelle 5.8: Verhältnis von Biaskorrektur zu Bias bei Benutzung eines loglinearen Modells. Datenbasis: SOEP und MZ-Panel	123

Abbildungsverzeichnis

Abbildung 2.1: Zusammenführbare Rotationsviertel der Mikrozensus- erhebungen 1996 bis 2004	35
Abbildung 2.2: Umfang der Analyseeinheiten in einer Paneluntersu- chung mit vier Befragungszeitpunkten ohne temporäre Ausfälle nach dem Analysekonzept mit Berücksichtigung aller verfügba- ren Fälle – schematische Darstellung	44
Abbildung 2.3: Umfang der Analyseeinheiten in einer Paneluntersu- chung mit vier Befragungszeitpunkten ohne temporäre Ausfälle nach dem Weiterverfolgungskonzept – schematische Darstellung .	45
Abbildung 2.4: Umfang der Analyseeinheiten in einer Paneluntersu- chung mit vier Befragungszeitpunkten ohne temporäre Ausfälle nach dem Rückverfolgungskonzept – schematische Darstellung . .	46
Abbildung 2.5: Umfang der Analyseeinheiten in einer Paneluntersu- chung mit vier Befragungszeitpunkten ohne temporäre Ausfälle nach dem Analysekonzept mit Berücksichtigung der vollständi- gen Fälle– schematische Darstellung	47
Abbildung 4.1: Definition der beiden betrachteten Mengen für die Randanpassung des Längsschnitts 1996-1999	98

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) 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).

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?	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. Darüber hinaus dient er der Rationalisierung anderer amtlicher Statistiken, wie z.B. der Einkommens- und Verbrauchsstichprobe sowie als Hochrechnungsrahmen nicht amtlicher Stichprobenerhebungen.
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 Jahreshauptgutachten 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 Berichtswche 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 vier Jahre in der Stichprobe.
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.

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 $4 \cdot (k+4)$ betrug. ($k \geq 1$: Zahl der Wohnungen des Gebäudes) oder wenn ein bewohntes Gebäude keine Wohnung aufwies”</i> (Meyer 1994: 107). 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. 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 kreisscharf 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 Haushaltsbezugs-person (Deutsche/r, Ausländer/in),
 - (c) für Deutsche: Wohnrechtsverhältnis der Haushaltsbezugs-person (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 Mikrozensusserhebungen 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	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	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	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	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 Arbeitserhebung 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 Auswahlatz für die Unterstichprobe an den Genauigkeitsanforderungen für die EU-Arbeitskräfteerhebung¹. Aufgrund dieser Genauigkeitsanforderungen kann der Auswahlatz 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 Auswahlatz von etwa 0,45 Prozent.

Da die Unterstichprobe so angelegt ist, dass einerseits die Belastung der Befragten gleichmäßig auf die Auswahlseinheiten 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 Auswahlatz 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 zweiten Rotationsviertel der vierten Stichprobe (RV 04/2) ausgewiesen sind – entnommen werden kann, sind im ersten Erhebungsjahr eines neu in die Auswahl gelangten Rotationsviertels rund 45 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 sind nur noch jene Personen in das Unterstichprobenprogramm einbezogen, die in Regierungsbezirken wohnen, in denen der Auswahlatz für das Unterstichprobenprogramm 1 Prozent beträgt. Aus diesem Rotationsschema des Unterstichprobenprogramms innerhalb eines Rotationsviertels und der unterschiedlichen Größe der Regierungsbezirke in den einzelnen Bundesländern folgt des Weiteren, dass die Auswahlätze 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 zweiten Rotationsviertel der vierten Stichprobe in den einzelnen Erhebungsjahren

Bundesland	Befragte Personen im zweiten Rotationsviertel der vierten Stichprobe im Erhebungsjahr ...							
	1996		1997		1998		1999	
	Insgesamt	darunter %-Anteil in UStp.	Insgesamt	darunter %-Anteil in UStp.	Insgesamt	darunter %-Anteil in UStp.	Insgesamt	darunter %-Anteil in UStp.
SH	6 180	40,9	6 207	80,4	6 201	80,5	6 259	–
HH	3 915	39,7	4 005	79,6	4 079	79,1	3 993	–
NI	15 723	40,2	16 369	84,3	16 583	84,0	16 427	4,2
HB	1 437	100,0	1 414	100,0	1 428	100,0	1 383	100,0
NW	37 825	39,6	38 491	79,3	38 775	80,0	38 580	–
HE	13 042	51,1	13 577	88,0	13 509	88,2	13 508	15,2
RP	9 041	54,6	8 955	89,5	9 074	90,3	9 059	19,7
BW	24 005	38,6	23 693	78,9	23 268	81,3	23 335	–
BY	27 584	48,7	27 678	87,3	27 652	88,2	27 679	11,3
SL	2 408	59,3	2 376	100,0	2 374	100,0	2 344	19,7
BE	7 887	48,7	7 809	88,1	7 723	87,0	7 624	7,0
BB	5 957	40,2	5 967	79,9	5 951	78,6	5 922	–
MV	3 890	39,8	3 844	79,2	3 765	76,5	3 628	–
SN	10 238	44,2	10 172	84,5	10 007	86,2	9 908	5,1
ST	6 115	67,8	6 152	100,0	6 112	100,0	5 926	49,5
TH	5 827	40,5	5 731	79,6	5 639	80,6	5 531	–
D	181 064	44,9	182 440	83,9	182 140	84,5	181 106	7,4

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 neu bearbeitete 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
- 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/download/d/mv/arbeitspapier12.pdf>

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 1996-1999 und nicht auf das faktisch anonymisierte Mikrozensus-Panelfile 1996-1999.

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.

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

Infolge der partiellen Rotation der Auswahlbezirke können zum einen Zweijahrespanels auf der Basis von jeweils drei Rotationsvierteln – also einer Auswahl von 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. In einem Zweijahreslängsschnitt liegen für etwa 450 000 Personen Angaben zu zwei Erhebungsjahren vor und in einem Vierjahreslängsschnitt kann immerhin noch auf rund 124 000 Personen zurückgegriffen werden.

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	
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 1996-1999

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.

Von den pro Erhebungsjahr insgesamt etwa 180 000 zur Verfügung stehenden Personen aus Privathaushalten konnten damit im Zeitraum 1996 bis 1999, unter Berücksichtigung der räumlichen Mobilität und den demographischen Ereignissen Geburt und Tod, etwa 86 bis 88 Prozent als plausibel im Sinne einer längsschnittanalytischen Auswertungsstrategie eingestuft werden. Der Anteil unplausibler Personensätze entsprach demnach 12 bis 14 Prozent. Dieser Anteil wurde als relativ hoch angesehen, weshalb nach der rein maschinellen Zusammenführung die unplausiblen Personensätze genauer untersucht wurden. Dazu wurden alle Personensätze eines Haushalts, in dem mindestens ein unplausibler Personensatz vorkam, mit einem reduzierten Variablensatz in eine Datei geschrieben. Unter Berücksichtigung der Haushaltszusammensetzung, des Familienstandes und des Eheschließungsjahres wurden in einem weiteren Schritt Korrekturen an der Haushaltsnummer, dem Geburtsjahr und/oder der Geschlechtsangabe durchgeführt. Des Weiteren wurden Veränderungen infolge natürlicher und räumlicher Bevölkerungsbewegungen gekennzeichnet. Bei der Inspektion der Daten zeigte sich, dass die Unstimmigkeiten vor allem dadurch entstanden, dass

- der Tod eines Ehepartners/einer Ehepartnerin oftmals nicht angegeben wurde,
- das Geburtsjahr (insbesondere von Kindern) um ± 2 Jahre variierte,
- der Fortzug einer Person nicht angegeben wurde oder
- der Fort-/Zuzug eines Haushalts nicht registriert wurde (und somit implizit zu fehlerhaften Haushaltsnummern führt).
- In wenigen Fällen zeigte sich zudem, dass falsche Zuordnungen von Personen zu einem Haushalt über einzelne Erhebungszeitpunkte vorlagen.

Es war daher naheliegend, diese – im Zeitablauf – Inkompabilitäten nachträglich zu bereinigen. Bis auf wenige Ausnahmen wurden diese Korrektoren maschinell

vorgenommen. Zusätzlich wurden drei neue Variablen gebildet, die eine Kennzeichnung darüber enthalten, ob eine Korrektur bei der Haushaltsnummer, dem Geburtsjahr bzw. dem Geschlecht vorgenommen wurde oder nicht (siehe hierzu [Kapitel 2.5](#)). In Abhängigkeit von der Anzahl divergierender Angaben über die einzelnen Erhebungszeitpunkte, wurden die Korrekturen wie folgt vorgenommen:

1. Es liegt eine gleiche Anzahl divergierender Angaben vor: Dies ist nur dann möglich, wenn Befragungsergebnisse zu zwei oder vier Erhebungszeitpunkten vorhanden sind. In diesen Fällen wurde jeweils die zeitlich erste Angabe als "wahrer" Wert aufgefasst und die nachfolgenden Werte entsprechend korrigiert.
2. Es liegt eine ungleiche Anzahl divergierender Angaben vor: In diesen Fällen wurde der jeweils häufigst vorkommende Wert als "wahr" angesehen und die anderen Werte entsprechend angepasst.

Die Wortwahl "manuelle Korrektur" ergibt sich demnach aus der "Inaugenscheinnahme" der Daten und der darauf aufbauenden Entscheidung über vorzunehmende Datenkorrekturen.

Die Nachbearbeitung des Mikrozensus-Paneldatensatzes 1996-1999 führt zu einer deutlichen Zunahme plausibler Sätze sowohl auf der Personenebene als auch auf der Ebene der Privathaushalte. So steigt der Anteil plausibler Angaben auf der Ebene der Personen, der bei der maschinellen Zusammenführung zwischen 86 und 88 Prozent liegt, auf 95,0 bis 96,3 Prozent.

Ausgeschlossen wurden Personen aus Gemeinschaftsunterkünften, da das Geburtsjahr und das Geschlecht in diesen Einrichtungen aufgrund starker Klumpungen umfangreichere Nachbearbeitungen erfordert, die im Mikrozensus-Panel 1996-1999 noch nicht umgesetzt werden konnten.

Hinsichtlich der Datenkorrekturen auf der Haushaltsebene ist anzumerken, dass eine Kennzeichnung temporärer Ausfälle nicht vorgenommen wurde und somit auch keine Unterscheidung zwischen Haushalten mit mindestens einem unplausiblen Personensatz und Haushalten mit einem temporären Ausfall vorgenommen werden kann.

2.3 Die Stichprobenentwicklung des Mikrozensus-Panels 1996-1999

Im vorhergehenden Unterkapitel wurde die Bildung des Mikrozensus-Panels 1996-1999 beschrieben. Es wurde zunächst eine rein maschinelle Zusammenführung vorgenommen, die im Anschluss nachbearbeitet wurde. In welchem

Umfang sich in den jeweiligen Phasen des Zusammenführungsprozesses Personensätze für den Mikrozensus-Paneldatensatz 1996-1999 zusammenführen lassen, ist Gegenstand nachfolgender Ausführungen.

Zunächst wird die Stichprobenentwicklung mit einer rein maschinellen Zusammenführung aufgezeigt und im Anschluss daran die Stichprobenentwicklung nach der Nachbearbeitung. In der Tabelle 2.1, in der der Umfang plausibler und unplausibler Angaben sowie temporärer Ausfälle auf der Personenebene für den Mikrozensus-Paneldatensatz 1996-1999 nach der rein maschinellen Zusammenführung für die einzelnen Erhebungsjahre ausgewiesen ist, sind sowohl personelle Veränderungen (Zuzüge, Geburten, Fortzüge und Sterbefälle) als auch Veränderungen, die den gesamten Haushalt (Zu-/Fortzug bzw. "Tod" eines Haushalts) betreffen, berücksichtigt.

Bei vier Erhebungszeitpunkten mit der Kennung, ob für den jeweiligen Erhebungszeitpunkt personelle Angaben vorliegen oder nicht, ergeben sich 16 Kombinationsmöglichkeiten. Von diesen werden für die weiteren Auswertungen zehn in die Analysen einbezogen und als "plausibel" eingestuft. So scheidet die Kombinationsmöglichkeit aus, nach der zu keinem Erhebungszeitpunkt eine personelle Angabe vorliegt. Kombinationsmöglichkeiten, bei denen Angaben zu zwei oder drei, von temporären Ausfällen unterbrochenen Erhebungszeitpunkten vorliegen, werden ebenfalls ausgeschlossen. Bei diesen temporären Ausfällen ist allerdings 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 hingegen 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 den vorgenannten Setzungen bezüglich plausibler und unplausibler Angaben sowie über temporäre Ausfälle liegt der Anteil plausibler Angaben auf der Ebene der Personen zwischen 86 und 88 Prozent, wobei der Anteil von Personen, für die über alle vier Erhebungszeitpunkte Angaben vorliegen, etwa 62,5 Prozent beträgt (siehe nachstehende Tabelle 2.1).

Dies bedeutet, dass zu den einzelnen Erhebungszeitpunkten für 22 bis 23 Prozent der Personen aufgrund von Wanderungen, Geburten oder Sterbefällen keine Informationen über alle Erhebungszeitpunkte vorliegen. Zwischen 1,4 und 1,7 Prozent der Angaben zum Erhebungszeitpunkt t_i sind im Erhebungszeitpunkt t_j mit einer Ausfallkennung gekennzeichnet ($i, j = 1996, 1997, 1998, 1999$ und $i \neq j$). Schließlich sind für die unplausiblen Angaben Anteile festzustellen, die zwischen 11,9 und 13,9 Prozent variieren.

Tabelle 2.1: Umfang plausibler und unplausibler Angaben sowie temporärer Ausfälle auf der Personenebene für den Mikrozensus-Panel Datensatz 1996-1999 nach der rein maschinellen Zusammenführung – ausgewiesen nach den einzelnen Erhebungsjahren

Personen in Privathaushalten	Angaben im Erhebungsjahr ...							
	1996		1997		1998		1999	
	Anzahl	%	Anzahl	%	Anzahl	%	Anzahl	%
Insgesamt	179 260	100,0	180 488	100,0	180 248	100,0	179 278	100,0
davon mit:								
- unplausiblen Angaben	22 724	12,7	25 040	13,9	23 668	13,1	21 332	11,9
- plausiblen Angaben ..	156 535	87,3	155 488	86,1	156 580	86,9	157 946	88,1
- Angaben im Erhebungsjahr...								
1996 1997 1998 1999								
X - - -	17 820	9,9	-	-	-	-	-	-
X X - -	13 255	7,4	13 255	7,3	-	-	-	-
X X X -	10 803	6,0	10 803	6,0	10 803	6,0	-	-
X X X X	112 222	62,6	112 222	62,2	112 222	62,3	112 222	62,6
- X - -	-	-	3 449	1,9	-	-	-	-
- X X -	-	-	2 300	1,3	2 300	1,3	-	-
- X X X	-	-	10 350	5,7	10 350	5,8	10 350	5,8
- - X -	-	-	-	-	3 549	2,0	-	-
- - X X	-	-	-	-	14 343	8,0	14 343	8,0
- - - X	-	-	-	-	-	-	18 113	10,1
Summe	154 100	86,0	152 379	84,4	153 568	85,2	155 028	85,5
- einem Ausfall in mindestens einem der anderen Erhebungsjahre ^{a)}	2 435	1,4	3 069	1,7	3 013	1,7	2 918	1,6

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.

In einem weiteren Schritt werden die zusammengeführten Datensätze auf der Haushaltsebene untersucht. Dabei werden die Angaben eines Haushalts dann als plausibel angenommen, wenn die Anzahl der zusammengeführten plausiblen Datensätze auf der Personenebene zum Erhebungszeitpunkt t_i der Anzahl der befragten Personen in einem Haushalt zum Zeitpunkt t_i entspricht. Betrachtet man unter dieser Annahme die plausiblen Angaben von Privathaushalten, so variiert der Anteil plausibler Haushalte an den befragten Privathaushalten insgesamt in den einzelnen Erhebungsjahren zwischen 78,8 und 80,2 Prozent und spiegelbildlich dazu der Anteil unplausibler Angaben von Privathaushalten zwischen 19,8 und 21,2 Prozent (siehe Tabelle 2.2). Sind nur die Privathaushalte von Interesse, für die plausible Angaben über alle vier Erhebungszeitpunkte vorliegen, so können etwa 58 Prozent der Privathaushalte in die Analyse einfließen.

Tabelle 2.2: Umfang plausibler und unplausibler Angaben sowie temporärer Ausfälle auf der Haushaltsebene für den Mikrozensus-Paneldatensatz 1996-1999 nach der rein maschinellen Zusammenführung – ausgewiesen nach den einzelnen Erhebungsjahren

Privathaushalte	Angaben im Erhebungsjahr ...							
	1996		1997		1998		1999	
	Anzahl	%	Anzahl	%	Anzahl	%	Anzahl	%
Insgesamt	84 179		84 553		84 723		84 827	
davon								
befragt	81 445	100,0	82 334	100,0	82 339	100,0	82,535	100,0
davon mit:								
- unplausiblen Angaben	14 685	18,0	15 516	18,8	15 698	19,1	15 303	18,5
- plausiblen Angaben ..	66 760	82,0	66 818	81,2	66 641	80,9	67 232	81,5
- Angaben im Erhebungsjahr...								
1996 1997 1998 1999								
X - - -	7 542	9,3	-	-	-	-	-	-
X X - -	5 793	7,1	5 747	7,0	-	-	-	-
X X X -	4 346	5,3	4 355	5,3	4 274	5,2	-	-
X X X X	47 626	58,5	47 626	57,8	47 626	57,8	47 626	57,7
- X - -	-	-	1 629	2,0	-	-	-	-
- X X -	-	-	1 102	1,3	1 083	1,3	-	-
- X X X	-	-	4 459	5,4	4 482	5,4	4 481	5,4
- - X -	-	-	-	-	1 628	2,0	-	-
- - X X	-	-	-	-	5 914	7,2	5 930	7,2
- - - X	-	-	-	-	-	-	7 872	9,5
Summe	65 307	80,2	64 217	78,8	65 431	78,9	65 649	79,8
- einem Ausfall in mindestens einem der anderen Erhebungsjahre	1 453	1,8	1 900	2,3	1 634	2,0	1 423	1,7
ausgefallen^{a)}	2 734	3,2	2 219	2,6	2 384	2,8	2 292	2,7

a) Bezugsgröße für Prozentanteil ist "Insgesamt"

Die Nachbearbeitung des Mikrozensus-Paneldatensatzes 1996-1999 führt zu einer deutlichen Ausweitung der Plausibilität sowohl auf der Personenebene als auch auf der Ebene der Privathaushalte. So steigt der Anteil plausibler Angaben auf der Ebene der Personen, der bei der maschinellen Zusammenführung zwischen 86 und 88 Prozent liegt, auf 95,0 bis 96,3 Prozent (siehe hierzu die Tabelle 2.1 sowie die nachfolgenden Tabellen 2.3 und 2.4). Werden nur die plausiblen Personensätze ohne temporäre Ausfälle betrachtet, ist ein Anstieg von 84,5 bis 86 Prozent um rund 9,5 Prozentpunkte auf 94 bis 95,5 Prozent festzustellen. Der Anteil von Personen, für die über alle vier Erhebungsjahre Angaben vorliegen, vergrößert sich um durchschnittlich 6,9 Prozentpunkte von 62,4 Prozent auf rund 69 Prozent.

Tabelle 2.3: Umfang plausibler und unplausibler Angaben sowie temporärer Ausfälle auf der Personenebene für den Mikrozensus-Paneldatensatz 1996-1999 nach Datenkorrektur – ausgewiesen nach den einzelnen Erhebungsjahren

Personen in Privathaushalten	Angaben im Erhebungsjahr ...							
	1996		1997		1998		1999	
	Anzahl	%	Anzahl	%	Anzahl	%	Anzahl	%
Insgesamt	179 260	100,0	180 488	100,0	180 248	100,0	179 278	100,0
davon mit:								
- unplausiblen Angaben	6 670	3,7	9 836	5,4	8 984	5,0	7 896	4,4
- plausiblen Angaben ..	172 590	96,3	170 652	94,6	171 264	95,0	171 382	95,6
- Angaben im Erhebungsjahr...								
1996 1997 1998 1999								
X - - -	20 045	11,2	-	-	-	-	-	-
X X - -	14 834	8,3	14 834	8,2	-	-	-	-
X X X -	11 879	6,6	11 879	6,6	11 879	6,6	-	-
X X X X	124 561	69,5	124 561	69,0	124 561	69,1	124 561	69,5
- X - -	-	-	3 756	2,1	-	-	-	-
- X X -	-	-	2 423	1,3	2 423	1,3	-	-
- X X X	-	-	11 988	6,6	11 988	6,6	11 988	6,7
- - X -	-	-	-	-	3 762	2,1	-	-
- - X X	-	-	-	-	15 065	8,4	15 065	8,4
- - - X	-	-	-	-	-	-	18 879	10,5
Summe	171 319	95,6	169 441	93,9	169 678	94,1	170 493	95,1
- einem Ausfall in mindestens einem der anderen Erhebungsjahre^{a)}	1 271	0,7	1 211	0,7	1 586	0,9	889	0,5

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.

Tabelle 2.4: Umfang plausibler Angaben auf der Personenebene für den Mikrozensus-Paneldatensatz 1996-1999 vor und nach Datenkorrektur – ausgewiesen nach den einzelnen Erhebungsjahren

EJa)	Plausible Personensätze							
	Insgesamt	darunter plausible Personensätze ...						
		vor den Datenkorrekturen		nach den Datenkorrekturen		Differenz: (5) - (3)	% von (7) an (3)	Differenz: (6) - (4)
		Anzahl	% von (2)	Anzahl	% von (2)			
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1996	179 260	154 100	86,0	171 319	95,6	17 219	11,2	9,6
1997	180 488	152 379	84,4	169 441	93,9	17 069	11,2	9,5
1998	180 248	153 568	85,2	169 678	94,1	16 110	10,5	8,9
1999	179 278	155 028	85,5	170 493	95,1	15 465	10,0	9,6
1996/99	179 819 ^{b)}	112 222	62,4	124 561	69,3	12 339	11,0	6,9

a) EJ = Erhebungsjahr

b) ungewichtetes arithmetisches Mittel der einzelnen Erhebungsjahre

Auf der Ebene der Privathaushalte konnten nach den Datenkorrekturen ohne Berücksichtigung temporärer Ausfälle zwischen 90 und 92,5 Prozent als

plausibel eingestuft werden. Gegenüber der maschinellen Zusammenführung bedeutet dies eine Verbesserung um rund 12 Prozentpunkte (siehe hierzu Tabelle 2.2 sowie die nachfolgenden Tabellen 2.5 und 2.6).

Tabelle 2.5: Umfang plausibler und unplausibler Angaben sowie temporärer Ausfälle auf der Haushaltsebene für den Mikrozensus-Paneldatensatz 1996-1999 nach Datenkorrektur – ausgewiesen nach den einzelnen Erhebungsjahren

Privathaushalte	Angaben im Erhebungsjahr . . .							
	1996		1997		1998		1999	
	Anzahl	%	Anzahl	%	Anzahl	%	Anzahl	%
Insgesamt	84 179		84 553		84 723		84 827	
davon befragt	81 445	100,0	82 334	100,0	82 339	100,0	82 535	100,0
darunter mit:								
- unplausiblen Angaben	o ^{a)}	o ^{a)}	o ^{a)}	o ^{a)}	o ^{a)}	o ^{a)}	o ^{a)}	o ^{a)}
- plausiblen Angaben . .	o ^{a)}	o ^{a)}	o ^{a)}	o ^{a)}	o ^{a)}	o ^{a)}	o ^{a)}	o ^{a)}
- Angaben im Erhebungsjahr . . .								
1996 1997 1998 1999								
X - - -	8 162	10,0	-	-	-	-	-	-
X X - -	5 763	7,1	5 763	7,0	-	-	-	-
X X X -	4 140	5,1	4 140	5,0	4 140	5,0	-	-
X X X X	57 268	70,3	57 268	69,6	57 268	69,6	57 268	69,4
- X - -	-	-	1 649	2,0	-	-	-	-
- X X -	-	-	1 016	1,2	1 016	1,2	-	-
- X X X	-	-	4 528	5,5	4 528	5,5	4 528	5,5
- - X -	-	-	-	-	1 657	2,0	-	-
- - X X	-	-	-	-	6 200	7,5	6 200	7,5
- - - X	-	-	-	-	-	-	8 435	10,2
Summe	75 333	92,5	74 364	90,3	74 809	90,9	76 431	92,6
- einem Ausfall in mindestens einem der anderen Erhebungsjahre	o ^{a)}	o ^{a)}	o ^{a)}	o ^{a)}	o ^{a)}	o ^{a)}	o ^{a)}	o ^{a)}
ausgefallen^{b)}	2 734	3,2	2 219	2,6	2 384	2,8	2 292	2,7

a) Nicht auswertbar, da nach der Datenkorrektur keine Unterscheidung zwischen Haushalten mit mindestens einem unplausiblen Personensatz und temporärem Ausfall eines Haushalts vorgenommen wurde.

b) Bezugsgröße für Prozentanteil ist "Insgesamt".

Der Anteil der Privathaushalte, die über alle vier Erhebungsjahre als plausibel gelten können, beziffert sich auf etwa 70 Prozent, was einer Steigerung von knapp 12 Prozentpunkten gegenüber der rein maschinellen Zusammenführung entspricht. Dementsprechend beträgt der Anteil plausibler mobiler Privathaushalte an den insgesamt befragten Privathaushalten zwischen 21 und 23 Prozent.

Tabelle 2.6: Umfang plausibler Angaben auf der Haushaltsebene für den Mikrozensus-Paneldatensatz 1996-1999 vor und nach Datenkorrektur – ausgewiesen nach den einzelnen Erhebungsjahren

EJa)	Plausible Angaben auf der Haushaltsebene									
	Insgesamt	darunter plausible Haushaltssätze . . .						Differenz: (5) - (3)	% von (7) an (3)	Differenz: (6) - (4)
		vor den Daten- korrekturen		nach den Daten korrekturen						
		Anzahl	% von (2)	Anzahl	% von (2)					
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)		
1996	81 445	65 307	80,2	75 333	92,5	10 026	15,4	12,3		
1997	82 334	64 918	78,8	74 364	90,3	9 446	14,6	11,5		
1998	80 339	65 007	79,0	74 809	90,9	9 802	15,1	11,9		
1999	82 535	65 809	79,7	76 431	92,6	10 622	16,1	12,9		
1996/99	82 163 ^{b)}	47 626	58,0	57 268	69,7	9 642	20,2	11,7		

a) EJ = Erhebungsjahr

b) Ungewichtetes arithmetisches Mittel der einzelnen Erhebungsjahre.

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**

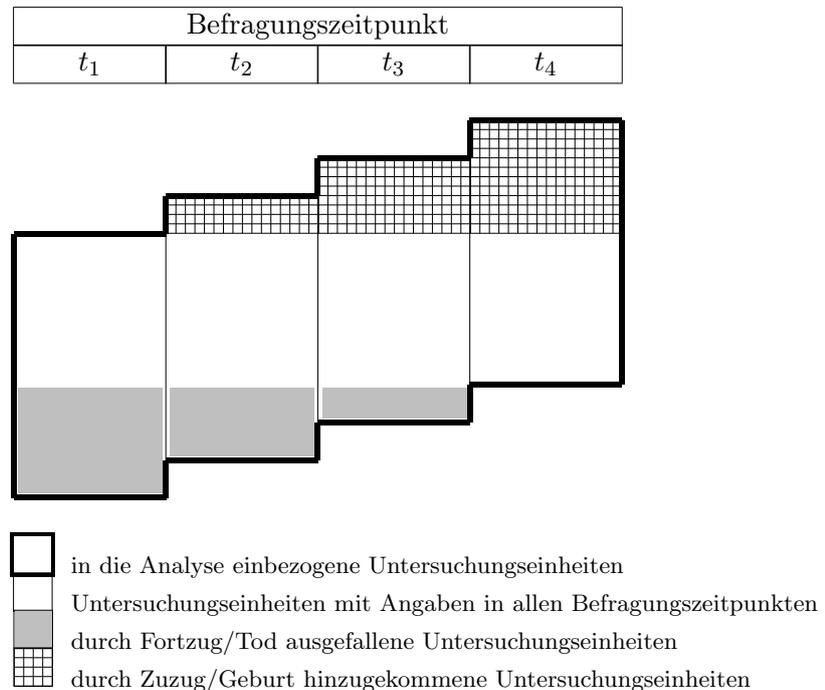
²In der Literatur wird dieses Auswertungskonzept als "complete information analysis" oder "available case analysis" bezeichnet. Hier wird jedoch eine deutsche Terminologie gewählt, die von Toutenburg et al. (2004) verwendet wird.

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.

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.

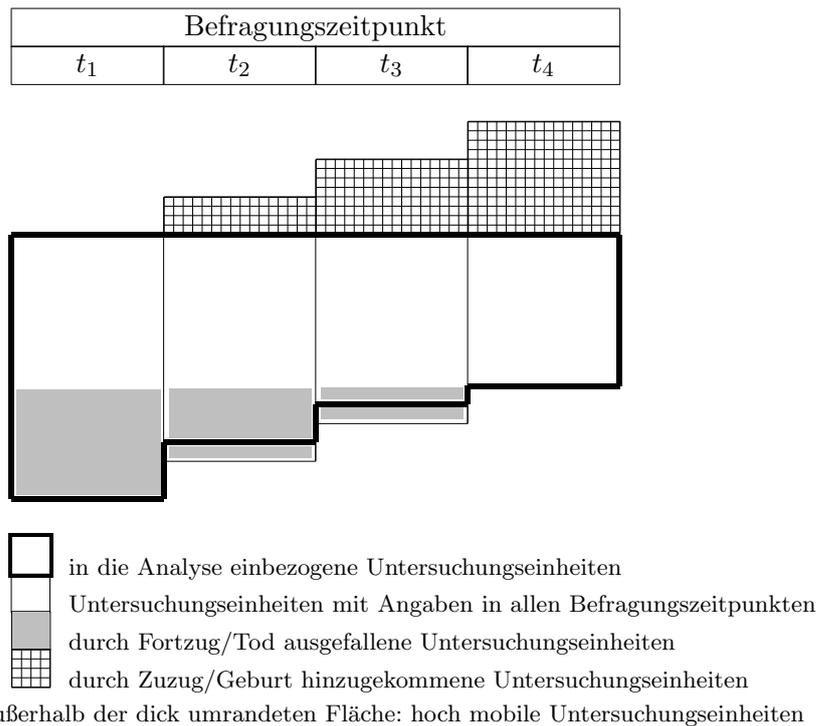
Abbildung 2.2: Umfang der Analyseeinheiten in einer Paneluntersuchung mit vier Befragungszeitpunkten ohne temporäre Ausfälle nach dem Analysekonzept mit Berücksichtigung aller verfügbaren Fälle – schematische Darstellung



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

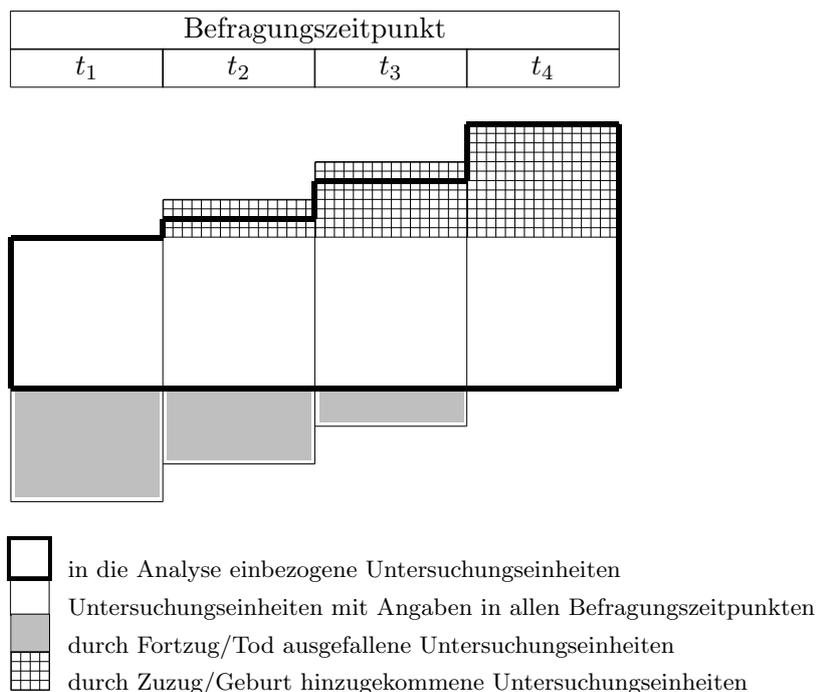
Fläche), werden nicht berücksichtigt. Dieses Konzept konzentriert sich damit auf die "Verluste" in einem Mikrozensus-Panel.

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-Paneldatensatz.

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

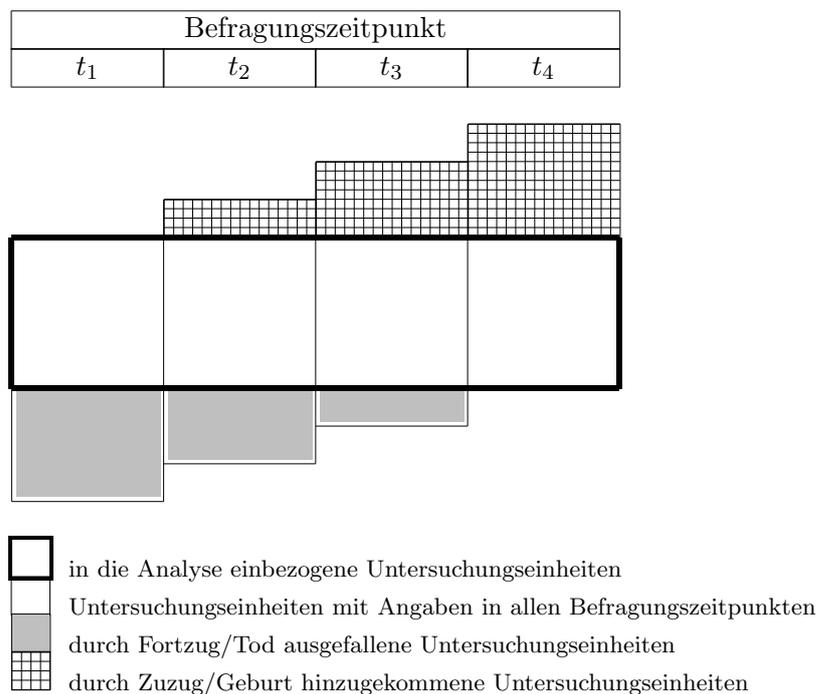
- (a) im Weiterverfolgungskonzept die Personen, die nach dem Erhebungszeitpunkt t_1 zugezogen oder geboren wurden, aus der Analyse ausgeschlossen werden und
- (b) im Rückverfolgungskonzept die Personen aus der Analyse ausgeschlossen bleiben, die vor dem Befragungszeitpunkt t_4 fortgezogen oder verstorben 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 `perkl` (wenn die Auswertungen auf der Personenebene stattfinden sollen) bzw. `hherkl` (wenn die Auswertungen nur für Per-

sonen 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-Panel datensatz 1996-1999

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 1996-1999 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-Panel datensatz 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. Gemäß den Datenschutzvorschriften wurden die Nummern für die Auswahlbezirke und Haushalte systemfrei vergeben.

Tabelle 2.8: Längsschnittinformationen im Mikrozensus-Panel 1996-1999

Längsschnittinformationen im Mikrozensus-Panel 1996-1999	
Variablenname	Inhalt
perkl	<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 1996, die Zeichen an dritter und vierter Stelle über den Personenstatus im Erhebungsjahr 1997, die Zeichen an fünfter und sechster Position über den Personenstatus im Erhebungsjahr 1998 und schließlich die Zeichen an siebter und achter Position über den Personenstatus im Erhebungsjahr 1999.</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; (c) ein kleines o, gefolgt von der Endziffer eines Erhebungsjahres weist darauf hin, dass in dem entsprechenden Erhebungsjahr der zu befragende Haushalt ausgefallen ist; (d) die Zeichenfolge "- " weist darauf hin, dass in dem entsprechenden Erhebungsjahr keine Angaben vorliegen und es keine Informationen über eventuelle Veränderungen gibt.

Fortsetzung der Tabelle auf nächster Seite

Längsschnittinformationen im Mikrozensus-Panel 1996-1999	
Variablenname	Inhalt
	Nachfolgend sind einige Beispiele zur Interpretation der Variablenlabels gegeben:
96xxxxxx	Es liegen nur für das Erhebungsjahr 1996 Angaben vor und die Person ist im Übergang 1996 → 1997 fortgezogen oder verstorben.
xx97xxxx	Es liegen nur für das Erhebungsjahr 1997 Angaben vor und die Person ist im Übergang 1996 → 1997 zugezogen oder geboren und im Übergang 1997 → 1998 fortgezogen oder verstorben.
96o7xxxx	Es liegen nur für das Erhebungsjahr 1996 Angaben vor. Im Erhebungsjahr 1997 ist der Haushalt ausgefallen und im Übergang 1997 → 1998 fortgezogen.
96o7- - -	Es liegen nur für das Erhebungsjahr 1996 Angaben vor. Im Erhebungsjahr 1997 ist der Haushalt ausgefallen und für die Erhebungsjahre 1998 und 1999 liegen keine Angaben vor, auch nicht über den Verbleib des Haushalts (unplausible Personensätze).

zusamm

Kennung über die Art der Zusammenführung

Die Variable **zusamm** kennzeichnet den Zusammenführungsprozess und gibt darüber Auskunft, ob die Zusammenführung rein maschinell mittels der Ordnungsnummern der Auswahlbezirke und Haushalte einerseits und den Personenmerkmalen Geburtsjahr und Geschlecht andererseits erfolgte (Kode = 1), oder ob eine nachträgliche Korrektur durchgeführt wurde (Kode = 2).

verluste

Kennzeichnung der Art eines Verlustes

Mit der Variablen **verluste** wird die Art eines Verlustes gekennzeichnet, wobei unterschieden wird zwischen:

- (a) dem Tod einer Person,
- (b) dem Tod/Fortzug einer Person,
- (c) dem Fortzug einer Person,
- (d) dem Fortzug eines Haushalts,
- (e) dem Tod/Fortzug eines Haushalts sowie
- (f) einem nicht zuordenbaren Verlust eines Haushalts.

Fortsetzung der Tabelle auf nächster Seite

Längsschnittinformationen im Mikrozensus-Panel 1996-1999	
Variablenname	Inhalt
gewinne	<p>Kennzeichnung der Art eines Zugangs</p> <p>Mit der Variablen gewinne wird die Art eines Zugangs gekennzeichnet, wobei unterschieden wird zwischen:</p> <ul style="list-style-type: none"> (a) der Geburt einer Person, (b) dem Zuzug einer Person, (c) dem Zuzug eines Haushalts sowie (d) einem nicht zuordenbaren Zugang.
flag_hh	<p>Kennung, ob Haushalt Nummer längsschnittkonsistent</p> <p>Die Variable flag_hh kennzeichnet, ob über alle Erhebungsjahre hinweg in denen ein Haushalt befragt wurde, eine identische Ordnungsnummer vergeben wurde (Kode = 1). In allen anderen Fällen weist die Variable den Wert 0 auf.</p>
flag_gj	<p>Kennung, ob Geburtsjahr längsschnittkonsistent sind</p> <p>Die Variable flag_gj kennzeichnet, ob über alle Erhebungsjahre hinweg in denen eine Person befragt wurde, das Geburtsjahr identische Werte aufweist (Kode = 1). In allen anderen Fällen weist die Variable den Wert 0 auf.</p>
flag_sex	<p>Kennung, ob Angaben zum Geschlecht längsschnittkonsistent sind</p> <p>Die Variable flag_sex kennzeichnet, ob über alle Erhebungsjahre hinweg in denen eine Person befragt wurde, die Angaben zum Geschlecht identische Werte aufweist (Kode = 1). In allen anderen Fällen weist die Variable den Wert 0 auf.</p>

Fortsetzung der Tabelle auf nächster Seite

Längsschnittinformationen im Mikrozensus-Panel 1996-1999	
Variablenname	Inhalt
hherkl	<p>Kennung plausibler Privathaushalte</p> <p>Die Variable hherkl kennzeichnet analog zu perkl den Zusammenführungserfolg auf der Ebene der Privathaushalte. Als plausibel wird ein Haushalt dann gekennzeichnet, wenn – solange ein Haushalt befragt wird – die Anzahl der Personensätze im Erhebungsjahr t_i in einem Haushalt gleich der Anzahl der plausiblen Personensätze ist.</p> <p>Im Gegensatz zur Kennung plausibler Personensätze, werden temporäre Ausfälle nicht berücksichtigt und es erfolgt auch keine differenzierte Unterscheidung der unplausiblen Fälle. Dementsprechend werden auf der Haushaltsebene nur Zusammenführungen gekennzeichnet, bei denen eine lückenlose Erklärung aller Personen in einem Haushalt für den Erhebungszeitraum von 1996 bis 1999 vorliegt.</p> <p>Bitte beachten Sie, dass perkl nicht mit hherkl bei Personen übereinstimmt, die in einen, bereits im Mikrozensus einbezogenen Haushalt zuziehen bzw. aus einem, weiterhin im Mikrozensus einbezogenen Haushalt ausziehen.</p> <p>Des Weiteren stimmt perkl nicht mit hherkl überein, wenn eine Person geboren wird bzw. eine Person verstirbt und der Haushalt, in dem dieses demographische Ereignis stattfindet, bereits bzw. weiterhin in den Mikrozensus einbezogen ist.</p>
hhk	<p>Kennung eines Datensatzes für Auswertungen auf der Haushaltsebene</p> <p>Die Variable hhk 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 Haushaltsbezugsperson vorgenommen wurde, und das für diese Stellvertreterposition auch keine weiteren Typisierungen gebildet wurden.</p>

2.6 Der Datensatzaufbau des Mikrozensus-Panels 1996-1999

Im Mikrozensus-Paneldatensatz 1996-1999 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 Variablenamen nicht mehr möglich⁴. Um die Umbenennungen der Variablenamen gegenüber dem jeweiligen Mikrozensus-Originalfile so gering wie möglich zu halten und unter der Vorgabe, dass die Maximallänge eines Variablenamens acht Zeichen nicht überschreiten sollte, wurde von den ursprünglichen Variablenamen in jedem Erhebungsjahr das erste Zeichen wie folgt ersetzt:

Erhebungsjahr	alter Variablenname ^{a)}	neuer Variablenname ^{a)}
1996	ef...	af...
1997	ef...	bf...
1998	ef...	cf...
1999	ef...	df...

a) Die Auslassungspunkte symbolisieren die jeweilige Feldnummer

Daraus ergibt sich folgender Datensatzaufbau im Mikrozensus-Panel 1996-1999 für die Variablen aus den Mikrozensus-Originalfiles:

Daten- satz von ...	Variablenname im MZ-Panel 1996-1999 für das Erhebungsjahr ...											
	1996			1997			1998			1999		
	afl	...	afp	bfl	...	bfq	cfl	...	cfr	dfl	...	dfs
Person 1												
Person 2												
⋮												
Person n												

p = letzte Feldnummer der Merkmale im Erhebungsjahr 1996

q = letzte Feldnummer der Merkmale im Erhebungsjahr 1997

r = letzte Feldnummer der Merkmale im Erhebungsjahr 1998

s = letzte Feldnummer der Merkmale im Erhebungsjahr 1999

Für Analysen mit wiederholten Messungen wird teilweise ein anderer Datensatzaufbau erwartet, nämlich, dass die Werte der wiederholten Messungen einer

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 Variablenamen setzen sich in der Regel daher aus zwei Bestandteilen zusammen. Aus dem Präfix **ef**, mit dem alle Variablenamen 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.

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 1996-1999 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.

2.8 Intertemporale Veränderungen der Erhebungsmerkmale im Erhebungszeitraum 1996 bis 1999

Alle Aspekte, so auch die zu erhebende Merkmale des Mikrozensus selbst als auch die in ihr einbezogenen Arbeitskräfteerhebung der Europäischen Union sind gesetzlich geregelt bzw. Verordnungen der Europäischen Union unterworfen. Änderungen hinsichtlich der Erhebungsmerkmale finden also nicht ad-hoc statt, sondern nur im Rahmen politischer und administrativer Entscheidungsprozesse.

Innerhalb des Erhebungszeitraums 1996 bis 1999 sind daher nur wenige Merkmale bzw. Merkmalsausprägungen von intertemporalen Veränderungen betroffen, die nachfolgend aufgezeigt werden.

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.

Bitte beachten Sie des Weiteren, dass intertemporale Veränderungen in Bezug auf Frageformulierung, Erläuterungen zu den einzelnen Fragen etc. hier nicht erörtert werden.

1. Wichtigster Grund, weshalb in der Berichtswoche weniger Stunden als normalerweise gearbeitet wurde.

⇒ Variablenamen im Mikrozensus-Panel: **xf145** (x=a,b,c,d)

Im Erhebungsjahr 1999 wurde als neue Merkmalsausprägung der Erziehungsurlaub aufgenommen, was nicht nur zu einer Ausweitung der Merkmalskategorien führte, sondern auch zu einer teilweisen veränderten Kodierung der Merkmalskategorien.

Erstmals erwähnt wurde der Aspekt des Erziehungsurlaubs im Erhebungsjahr 1998, allerdings wurde er in einem Klammersausdruck der Kategorie "Persönliche, familiäre Verpflichtungen oder sonstige persönliche Gründe" nachgestellt.

Tabelle 2.9: Intertemporale Veränderungen des Erhebungsmerkmals: Wichtigster Grund für weniger geleistete Arbeitsstunden als normalerweise in der Berichtswoche

Merkmalsausprägung	Kodierung	
	in den EJ 1996-1998	im EJ 1999
– Krankheit, Unfall	1	1
– Kur, Heilstättenbehandlung	2	2
– Arbeitsschutzbestimmungen, auch Mutterschaft	3	3
– Erziehungsurlaub	–	4
– Urlaub, Dienstbefreiung	4	5
– Streik, Aussperrung	5	6
– Schlechtwetterlage	6	7
– Kurzarbeit	7	8
– Aufnahme einer Tätigkeit in der Berichtswoche	8	9
– Beeindigung einer Tätigkeit in der Berichtswoche ..	9	10
– Arbeitsstunden zu anderen Terminen geleistet (auch gleitende Arbeitszeit und andere flexible Arbeitszeiten)	10	11
– Teilnahme an Schulausbildung, Aus- oder Fortbildung außerhalb des Betriebes	11	12
<i>In den Erhebungen 1996, 1997 und 1999:</i>		
– Persönliche, familiäre Verpflichtungen oder sonstige persönliche Gründe		
<i>In der Erhebung 1998:</i>		
– Persönliche, familiäre Verpflichtungen oder sonstige persönliche Gründe (auch Erziehungsurlaub)	12	13
– Sonstige Gründe	13	14

EJ = Erhebungsjahr.

2. Grund, warum eine erwerbstätige Person eine andere odere weitere Tätigkeit sucht.

⇒ Variablenamen im Mikrozensus-Panel: **xf207** (x=a,b,c,d)

Im Erhebungsjahr 1999 wurden zwei neue Merkmalsausprägungen aufgenommen: "Suche nach Tätigkeit mit längerer Arbeitszeit" und "Suche nach Tätigkeit mit kürzerer Arbeitszeit". Die Erweiterung der Merkmalskategorien führte bei zwei Merkmalsausprägungen zu einer veränderten Kodierung, wie der nachfolgenden Aufstellung entnommen werden kann.

Tabelle 2.10: Intertemporale Veränderungen des Erhebungsmerkmals: Wichtigster Grund für Arbeitsuche von erwerbstätigen Personen

Merkmalsausprägung	Kodierung	
	in den EJ 1996-1998	im EJ 1999
- Bevorstehende Beendigung der gegenwärtigen Tätigkeit	1	1
- Jetzige Tätigkeit nur Übergangstätigkeit	2	2
- Suche nach 2. Tätigkeit	3	3
- Suche nach Tätigkeit mit längerer Arbeitszeit	-	4
- Suche nach Tätigkeit mit kürzerer Arbeitszeit	-	5
- Suche nach besseren Arbeitsbedingungen	4	6
- Aus anderen Gründen	5	7

EJ = Erhebungsjahr.

3. Art der aktiven Arbeitsuche nach einer Arbeitnehmertätigkeit innerhalb der letzten vier Wochen

⇒ Variablenamen im Mikrozensus-Panel: **df227** und **xf248** (x=a,b,c,d)

Im Erhebungsjahr 1999 wurde bei dieser Frage als neue Merkmalskategorie "Tests, Vorstellungsgespräche, Prüfungen" aufgenommen.

Da bei dieser Frage Mehrfachnennungen zugelassen sind, hat die Erweiterung zwei Auswirkungen. Zum einen liegt für das Erhebungsjahr 1999 eine neue Variable vor (**df227**), in der dokumentiert ist, ob die Antwortkategorie "Tests, Vorstellungsgespräche, Prüfungen" von einer Person genannt wurde oder nicht und zum anderen führte dies zu einer Erweiterung und in Teilbereichen zu einer veränderten Kodierung der Merkmalskategorien innerhalb des typisierten Merkmals **xf248** (x=a,b,c,d). In dem typisierten Merkmal **xf248** wird festgehalten, welche aktive Methode der Arbeitsuche in der Liste der Nennungen als erste aktive Methode genannt wurde. Die nachfolgende Aufstellung der Antwortkategorien entspricht dabei der Reihenfolge der Antwortmöglichkeiten zur Art der aktiven Arbeitsuche.

Tabelle 2.11: Intertemporale Veränderungen des Erhebungsmerkmals: Art der aktiven Arbeitsuche nach einer Arbeitnehmertätigkeit innerhalb der letzten vier Wochen

Merkmalsausprägung	Kodierung	
	in den EJ 1996-1998	im EJ 1999
– Suche über das Arbeitsamt	1	1
– Suche über private Vermittlung	2	2
– Aufgabe von Inseraten	3	3
– Bewerbung auf Inserate	4	4
– Bewerbung auf eine nicht ausgeschriebene Stelle ..	5	5
– Suche über persönliche Verbindung	6	6
– Durchsehen von Inseraten	7	7
– Tests, Vorstellungsgespräche, Prüfungen	–	8
– Sonstige Bemühungen	8	9

EJ = Erhebungsjahr.

4. Infolge der Ablösung der Verordnung (EWG) Nr. 3711/91 des Rates vom 16. Dezember 1991 (Amtsblatt Nr. L 351 vom 20.12.1991, S. 1-3) durch die Verordnung (EG) Nr. 577/98 des Rates vom 9. März 1998 zur Durchführung einer Stichprobenerhebung in der Europäischen Gemeinschaft (Amtsblatt Nr. L 77 vom 14.3.1998, S. 3-7) wurden neue Merkmale im Bereich der Aus- und Weiterbildung eingeführt. Damit verbunden war, dass die bis dahin vergebenen Variablennamen, die sich am Datensatzaufbau orientierten, nicht mehr beibehalten werden konnten. Demzufolge war fast der gesamte Bereich der Aus- und Weiterbildung von diesen Änderungen betroffen. Da aber dieser Bereich auch fast ausschließlich im Rahmen des Ergänzungsprogramms erhoben wurde und das Ergänzungsprogramm für das Erhebungsjahr 1999 nicht ausgewiesen ist – zu den Gründen siehe Abschnitt 1.3 –, sind im Mikrozensus-Panel 1996-1999 nur zwei Variablen betroffen:

Tabelle 2.12: Intertemporale Veränderungen in Bezug auf Variablennamen

Erhebungsmerkmal	Variablenname	
	in den EJ 1996-1998	im EJ 1999
Beruflicher Ausbildungs- oder (Fach-)Hochschulabschluss vorhanden?	xf289	df288
Art des höchsten beruflichen Ausbildungs- oder (Fach-)Hochschulabschlusses	xf290	df289

EJ = Erhebungsjahr.

Darüber hinaus wurden die Antwortmöglichkeiten des Merkmals "Art des höchsten beruflichen Ausbildungs- oder (Fach-)Hochschulabschlusses" erweitert. Sowohl hinsichtlich der Anzahl der Antwortmöglichkeiten als auch hinsichtlich der Aufzählung erreichter Abschlüsse auf einem bestimmten Ausbildungsniveau. Die diesbezüglichen intertemporalen Veränderungen können Sie nachfolgender Aufstellung entnehmen.

Tabelle 2.13: Intertemporale Veränderungen des Erhebungsmerkmals: Art des höchsten beruflichen Ausbildungs- oder (Fach-)Hochschulabschlusses

Erfragter höchster beruflicher Ausbildungs- oder (Fach-)Hochschulabschluss ... in den Erhebungsjahren 1996-1998		im Erhebungsjahr 1999	
Merkmalskategorie	Kode	Merkmalskategorie	Kode
- Anlernausbildung oder berufliches Praktikum	1	- Anlernausbildung oder berufliches Praktikum	1
		- Berufsvorbereitungsjahr ..	2
- Abschluss einer Lehrausbildung oder gleichwertiger Berufsfachschulabschluss .	2	- Abschluss einer Lehrausbildung, Vorbereitungsdienst für den mittleren Dienst in der öffentlichen Verwaltung	3
		- Berufsqualifizierender Abschluss an einer Berufsfach- oder Kollegscheule, Abschluss einer 1-jährigen Schule des Gesundheitswesens	4
- Meister-/Techniker- oder gleichwertiger Fachschulabschluss	3	- Meister-/Technikerausbildung oder gleichwertiger Fachschulabschluss, Abschluss einer 2- oder 3-jährigen Schule des Gesundheitswesens, Abschluss einer Fach- oder Berufsakademie, Abschluss einer Verwaltungsfachhochschule	5
- Abschluss der Fachschule der ehemaligen DDR	4	- Abschluss der Fachschule der ehemaligen DDR	6
- Fachhochschulabschluss (auch Ingenieurschulabschluss)	5	- Fachhochschulabschluss (auch Ingenieurschulabschluss, ohne Verwaltungsfachhochschule)	7
- Hochschulabschluss	6	- Hochschulabschluss	8
		- Promotion	9
- keine Angabe	9	- keine Angabe	99

2.9 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
- Schimpl-Neimanns, B. (2006): *Filekonzept zum Mikrozensuspanel*. Arbeitspapier#12, Methodenverbund "Aufbereitung und Bereitstellung des Mikrozensus als Panelstichprobe".
URL: <http://www.destatis.de/download/d/mv/arbeitspapier12.pdf>
- 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>
- Zühlke, S. (2001): *Längsschnittdaten auf der Basis des Mikrozensus. Methodische Probleme und Lösungsansätze*, in: Statistische Analysen und Studien Nordrhein-Westfalen; **53**: 3-13.

Kapitel 3

Das faktisch anonymisierte Mikrozensus-Panel 1996-1999

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 1996-1999 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 1996-1999 sind im nachfolgenden Kapitel 3.1 näher erläutert.

Wie bereits erwähnt enthält das faktisch anonymisierte Mikrozensus-Panel 1996-1999 nicht nur die plausibel zusammengeführten Datensätze, sondern umfasst alle Personensätze aus Privathaushalten der ausgewählten Auswahlbezirke. Damit besteht die Möglichkeit mit dem faktisch anonymisierten Mikrozensus-Panel 1996-1999 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

¹Auswahlplan der Substichprobenziehung des faktisch anonymisierten Mikrozensus-Panels 1996-1999.

²Anonymisierungskonzept.

Auswertungen, als auch für Inferenzstatistiken, mit denen Parameter der Grundgesamtheit geschätzt 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 1996-1999 nur Personen aus Privathaushalten beinhaltet, können die im Querschnitt hochgerechneten Randverteilungen des faktisch anonymisierten Mikrozensus-Panels 1996-1999 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 1996 bis 1999 (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 1996-1999 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 1996-1999

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 Datenauszügen 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 1996-1999 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 Auswahlsatz 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 das Erhebungsjahr 1999 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 Auswahlsätze in den einzelnen Regierungsbezirken (siehe hierzu die Erläuterungen auf Seite 28). Damit wären Regierungsbezirke mit einem hohen Auswahlsatz identifizierbar. Nach den Datenschutzregeln zur Umsetzung der faktischen Anonymisierung dürfen aber keine Regierungsbezirke als regionale Einheiten erkennbar sein, weshalb die Merkmale der Unterstichprobe für das Erhebungsjahr 1999 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 1996-1999

Auswahlplan der Substichprobenziehung des faktisch anonymisierten Mikrozensus-Panels 1996-199	
Auswahlgesamtheit	Personen in Privathaushalten des zweiten Rotationsviertels der vierten Stichprobe.
Auswahleinheiten	Auswahlbezirke
Auswahltechnik	Sortierung der Auswahlbezirke nach: <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>Die Sortierung folgt damit weitgehend der Schichtung des Originalmaterials. Neu aufgenommen in die Sortierung wurde die Größenklasse der Auswahlbezirke sowie die Kennung eines Auswahlbezirks über dessen Einbezug in die Unterstichprobe. 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 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.

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 1996-1999

Die Substichprobenziehung des faktisch anonymisierten Mikrozensus-Panels 1996-1999 umfasst 70 Prozent der Auswahlseinheiten, in denen im Erhebungszeitraum 1996 bis 1999 mindestens eine Person aus einem Privathaushalt befragt wurde. Durch die bei der Substichprobenziehung kontrollierte Größe der Auswahlseinheiten – gemessen an der durchschnittlichen Wohnbevölkerung in einer Auswahlseinheit – müssten damit annähernd auch 70 Prozent einer jeden, durch `perkl` bzw. `hherkl` definierten Teileinheit, in der Substichprobe enthalten sein.

Für die Personenebene wird der angestrebte Auswahlssatz von 70 Prozent weitgehend eingehalten (siehe Tabelle 3.2). Allerdings ist hinsichtlich der einzelnen Teileinheiten eine systematische Verschiebung zu beobachten. Teileinheiten, die im Verlauf des Erhebungszeitraums von 1996 bis 1999 aus einer Auswahlseinheit fortziehen, weisen mit einem Auswahlssatz von 69,5 Prozent eine geringfügige Untererfassung (gemessen am erwarteten Auswahlssatz von 70 Prozent) auf. Die Unterrepräsentation fällt noch etwas stärker bei den Immobilien aus, bei denen ein Auswahlssatz von 69,1 Prozent festgestellt werden kann. Demgegenüber ist bei den Teileinheiten, die im Verlauf des Erhebungszeitraums von 1996 bis 1999 in eine Auswahlseinheit zugezogen sind, mit einem Auswahlssatz von etwa 71 Prozent eine leichte Überrepräsentation gegeben.

Ein ähnliches Bild zeichnet sich auf der Ebene der Haushalte ab. Allerdings ist die Unterrepräsentation der Teileinheiten, die im Verlauf des Erhebungszeitraums 1996 bis 1999 aus einer Auswahlseinheit fortziehen sowie die Unterrepräsentation bei den Immobilien weniger stark und die Überrepräsentation der Teileinheiten, die im Verlauf des Erhebungszeitraums 1996 bis 1999 in eine Auswahlseinheit zugezogen sind, stärker ausgeprägt (siehe Tabelle 3.3).

Tabelle 3.2: Die Stichprobenentwicklung des faktisch anonymisierten Mikrozensus-Panels 1996-1999 auf der Personenebene

Personen in Privat- haushalten	Anzahl der Personen im MZ-Panel Originalfile (MZPO) und im faktisch anonymisierten MZ-Panel (fMZP) sowie der Auswahlsatz im Erhebungsjahr											
	1996			1997			1998			1999		
	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	179 260	124 383	69.4	180 488	125 248	69.5	180 248	125 685	69.7	179 278	124 831	69.6
davon mit:												
- unplausiblen Angaben	6 670	4 628	69.4	9 386	6 917	73.7	8 984	6 423	71.5	7 896	5 527	70.0
- plausiblen Angaben . .	172 590	119 755	69.4	170 652	118 469	69.4	171 264	125 685	69.6	171 382	119 304	69.6
davon mit:												
- Angaben zum Erhe- bungsjahr												
1996 1997 1998 1999												
X - - - -	20 045	14 136	70.5	-	-	-	-	-	-	-	-	-
X X - - -	14 834	10 356	69.8	14 834	10 356	69.8	-	-	-	-	-	-
X X X - -	11 879	8 242	69.4	11 879	8 242	69.4	11 879	8 242	69.4	-	-	-
X X X X -	124 561	86 112	69.1	124 561	86 112	69.1	124 561	86 112	69.1	124 561	86 112	69.1
- X - - -	-	-	-	3 756	2 647	70.5	-	-	-	-	-	-
- X X - -	-	-	-	2 423	1 718	70.9	2 423	1 718	70.9	-	-	-
- X X X -	-	-	-	11 988	8 530	71.2	11 988	8 530	71.2	11 988	8 530	71.2
- - X - -	-	-	-	-	-	-	3 762	2 670	71.0	-	-	-
- - - X X	-	-	-	-	-	-	15 065	10 864	72.1	15 065	10 864	72.1
- - - - X	-	-	-	-	-	-	-	-	-	18 879	13 168	69.7
Summe	171 319	118 846	69.4	169 441	117 605	69.4	169 678	118 136	69.6	170 493	118 674	69.6
- einem Ausfall in min- destens einem der anderen Erhebungs- jahrea)	1 271	909	71.5	1 211	864	71.3	1 586	1 126	71.0	889	630	70.9

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.

Tabelle 3.3: Die Stichprobenentwicklung des faktisch anonymisierten Mikrozensus-Panels 1996-1999 auf der Haushaltsebene

Personen in Privat- haushalten	Anzahl der Haushalte im MZ-Panel Originalfile (MZPO) und im faktisch anonymisierten MZ-Panel (fMZP) sowie der Auswahlatz im Erhebungsjahr											
	1996			1997			1998			1999		
	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	81 445	56 680	69,6	82 334	57 422	69,7	82 339	57 571	69,9	82 535	57 600	69,8
davon mit:												
- Angaben zum Erhe- bungsjahr												
1996 1997 1998 1999												
X - - -	8 162	5 775	70,8	-	-	-	-	-	-	-	-	-
X X - -	5 763	4 020	69,8	5 763	4 020	69,8	-	-	-	-	-	-
X X X -	4 140	2 890	69,8	4 140	2 890	69,8	4 140	2 890	69,8	-	-	-
X X X X	57 268	39 785	69,5	57 268	39 785	69,5	57 268	39 785	69,5	57 268	39 785	69,5
- X - -	-	-	-	1 649	1 228	74,5	-	-	-	-	-	-
- X X -	-	-	-	1 016	752	74,0	1 016	752	74,0	-	-	-
- X X X	-	-	-	4 528	3 388	74,8	4 528	3 388	74,8	4 528	3 388	74,8
- - X -	-	-	-	-	-	-	1 657	1 257	75,9	-	-	-
- - X X	-	-	-	-	-	-	6 200	4 676	75,4	6 200	4 676	75,4
- - - X	-	-	-	-	-	-	-	-	-	8 435	6 175	73,2
Summe	75 333	52 470	69,7	74 364	52 063	70,0	74 809	52 748	70,5	76 431	54 024	70,7

3.3 Das Variablenspektrum des faktisch anonymisierten Mikrozensus-Panels 1996-1999

Das Variablenspektrum des faktisch anonymisierten Mikrozensus-Panels 1996-1999 orientiert sich an der Variablenauswahl der faktisch anonymisierten Mikrozensus-Grundfiles der entsprechenden Erhebungsjahre. Der Datensatz enthält mit Ausnahme des Erhebungsjahres 1999 also auch die Merkmale des Ergänzungsprogramms, die mit einem reduzierten Auswahlssatz erhoben werden, als auch die Merkmale des Zusatzprogramms, die nicht jährlich, sondern in einem Abstand von vier Jahren und teilweise ebenfalls mit einem reduzierten Auswahlssatz erhoben werden.

Die Merkmale, die mit einem reduzierten Auswahlssatz erhoben werden, wurden für das Erhebungsjahr 1999 nicht in den Datensatz aufgenommen, da der Auswahlssatz, der am Ergänzungsprogramm beteiligten Haushalte/Personen in der Regel innerhalb eines Rotationsviertels von Erhebungsjahr zu Erhebungsjahr variiert (siehe [Tabelle 1.2](#)). Für das Erhebungsjahr 1999 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-Grundfiles mit einem deutlich geringeren Auswahlssatz³. Da als Auswahlseinheiten der Substichprobenziehung des faktisch anonymisierten Mikrozensus-Panels 1996-1999 die so genannten Auswahlbezirke herangezogen wurden, ergibt sich für die einzelnen Erhebungsjahre kein einheitlicher Auswahlssatz 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-Grundfiles 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 Auswahlssatz 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 1996-1999 zu gelangen beträgt etwa 0.175

3.4 Die Bildung angepasster Hochrechnungsfaktoren für Querschnittsanalysen im Mikrozensus-Panel 1996-1999

Um eine möglichst gute Anpassung der Mikrozensus-Panel-Querschnitte an die jeweiligen Querschnitte der Mikrozensus der Erhebungsjahre 1996 bis 1999 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 1996-1999 des t -ten Erhebungsjahres gelangt ($t = 1996, 1997, 1998, 1999$).

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 dabei nach der Kreuzkombination folgender Variablen:

- Bundesland (16 Bundesländer)
- × Geschlecht (männlich, weiblich)
- × Staatsangehörigkeit (deutsch, nicht deutsch)
- × Altersgruppe (< 15 Jahre, $15 \leq 45$ Jahre, $45 \leq 65$ Jahre, ≥ 65 Jahre)
- × Erwerbsstatus (erwerbstätig, erwerbs-/arbeitslos, Nichterwerbsperson).

Für einen angepassten Hochrechnungsfaktor einer Person i in der j -ten Subpopulation des t -ten Erhebungsjahres mit ($t = 1996, 1997, 1998, 1999$) ($HF_{j,t}^{Anp}$) 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.

$n_{j,t}^{MZQ_g}$ = Anzahl der hochgerechneten Personen im Mikrozensus-Originalfile in der j -ten Subpopulation des t -ten Erhebungsjahres.

$n_{j,t}^{MZP_g}$ = 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 $n_{j,t}^{MZQ_g} / n_{j,t}^{MZP_g}$ bezeichnet dabei den Korrekturfaktor.

Unter Berücksichtigung, dass Personen im Alter von unter 15 Jahren nach Definition nicht erwerbstätig oder erwerbs- bzw. arbeitslos sein können, ergeben sich insgesamt 620 Kombinationsmöglichkeiten, die zwar in den Mikrozensus-Originalfiles alle besetzt sind, auf Grund des kleineren Stichprobenumfangs nicht aber in der Substichprobe. In diesen Fällen wurde als Korrekturfaktor der Kehrwert der erwarteten Auswahlwahrscheinlichkeit, dass eine Person im k -ten Bundesland aus dem Mikrozensus-Originalfile in die Substichprobe des faktisch anonymisierten Mikrozensus-Panels aufgenommen wurde, eingesetzt. In der Folge dieser Mittelwertersetzung der Korrekturfaktoren bei leeren Tabellenfeldern kommt es allerdings zu geringen Abweichungen hinsichtlich der hochgerechneten Fallzahlen.

Da das faktisch anonymisierte Mikrozensus-Panel 1996-1999 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 1996-1999 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 1996-1999 gelangt:

Tabelle 3.4: Auswahlwahrscheinlichkeit für eine Person aus der Unterstichprobe des Mikrozensus-Originalfiles in die Unterstichprobe des faktisch anonymisierten Mikrozensus-Panels 1996-1999 zu gelangen

Erhebungsjahr	Auswahlwahrscheinlichkeit
1996	$0,25 * 0,449 * 0,70 = 0,0786$
1997	$0,25 * 0,839 * 0,70 = 0,1468$
1998	$0,25 * 0,845 * 0,70 = 0,1478$

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".

Die Variablenamen 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 1996-1999

Angepasster Hochrechnungsfaktor . . .		
für das Erhebungsjahr	Variablenname	Variablenlabel
1996	af750d	(EF750) MZ96: Personen-Hochrechnungsfaktor - (korrigiert für deskriptive Statistiken)
	af750b	(EF750) MZ96: Personen-Hochrechnungsfaktor - (korrigiert für Inferenzstatistiken)
	af751d	(EF751) MZ96: HH- und Familien-Hochrechnungsfaktor - (korrigiert für deskriptive Statistiken)
	af751b	(EF751) MZ96: HH- und Familien-Hochrechnungsfaktor - (korrigiert für Inferenzstatistiken)
	af755d	(EF755) MZ96: Personen-Hochrechnungsfaktor - UStp (korrigiert für deskriptive Statistiken)
	af755b	(EF755) MZ96: Personen-Hochrechnungsfaktor - UStp (korrigiert für Inferenzstatistiken)
	af756d	(EF756) MZ96: HH- und Familien-Hochrechnungsfaktor - UStp (korrigiert für deskriptive Statistiken)
	af756b	(EF756) MZ96: HH- und Familien-Hochrechnungsfaktor - UStp (korrigiert für Inferenzstatistiken)
1997	bf750d	(EF750) MZ97: Personen-Hochrechnungsfaktor - (korrigiert für deskriptive Statistiken)
	bf750b	(EF750) MZ97: Personen-Hochrechnungsfaktor - (korrigiert für Inferenzstatistiken)
	bf751d	(EF751) MZ97: HH- und Familien-Hochrechnungsfaktor - (korrigiert für deskriptive Statistiken)
	bf751b	(EF751) MZ97: HH- und Familien-Hochrechnungsfaktor - (korrigiert für Inferenzstatistiken)
	bf755d	(EF755) MZ97: Personen-Hochrechnungsfaktor - UStp (korrigiert für deskriptive Statistiken)
	bf755b	(EF755) MZ97: Personen-Hochrechnungsfaktor - UStp (korrigiert für Inferenzstatistiken)
	bf756d	(EF756) MZ97: HH- und Familien-Hochrechnungsfaktor - UStp (korrigiert für deskriptive Statistiken)
	bf756b	(EF756) MZ97: HH- und Familien-Hochrechnungsfaktor - UStp (korrigiert für Inferenzstatistiken)

Fortsetzung der Tabelle auf nächster Seite

Angepasster Hochrechnungsfaktor . . .		
für das Erhebungsjahr	Variable-name	Variablenlabel
1998	cf750d	(EF750) MZ98: Personen-Hochrechnungsfaktor - (korrigiert für deskriptive Statistiken)
	cf750b	(EF750) MZ98: Personen-Hochrechnungsfaktor - (korrigiert für Inferenzstatistiken)
	cf751d	(EF751) MZ98: HH- und Familien-Hochrechnungsfaktor - (korrigiert für deskriptive Statistiken)
	cf751b	(EF751) MZ98: HH- und Familien-Hochrechnungsfaktor - (korrigiert für Inferenzstatistiken)
	cf755d	(EF755) MZ98: Personen-Hochrechnungsfaktor - UStp (korrigiert für deskriptive Statistiken)
	cf755b	(EF755) MZ98: Personen-Hochrechnungsfaktor - UStp (korrigiert für Inferenzstatistiken)
	cf756d	(EF756) MZ98: HH- und Familien-Hochrechnungsfaktor - UStp (korrigiert für deskriptive Statistiken)
	cf756b	(EF756) MZ98: HH- und Familien-Hochrechnungsfaktor - UStp (korrigiert für Inferenzstatistiken)
	cf757d	(EF757) MZ98: Wohnungsfaktor - (korrigiert für deskriptive Statistiken)
	cf757d	(EF757) MZ98: Wohnungsfaktor - (korrigiert für Inferenzstatistiken)
1999	df750d	(EF750) MZ99: Personen-Hochrechnungsfaktor - (korrigiert für deskriptive Statistiken)
	df750b	(EF750) MZ99: Personen-Hochrechnungsfaktor - (korrigiert für Inferenzstatistiken)
	df751d	(EF751) MZ99: HH- und Familien-Hochrechnungsfaktor - (korrigiert für deskriptive Statistiken)
	df751b	(EF751) MZ99: HH- und Familien-Hochrechnungsfaktor - (korrigiert für Inferenzstatistiken)

3.5 Randverteilungen ausgewählter Merkmale in den Mikrozensus-Originalfiles und den Querschnitten des faktisch anonymisierten Mikrozensus-Panels 1996-1999

Wie bereits erwähnt beinhaltet das faktisch anonymisierte Mikrozensus-Panel 1996-1999 nur Personen aus Privathaushalten. Die im Querschnitt hochgerechneten Randverteilungen können daher nicht generell mit veröffentlichten Tabellen aus den Fachserien des Mikrozensus verglichen werden, da in diesen zumeist die Personen in Gemeinschaftsunterkünften mit enthalten sind. Für ausgewählte Merkmale (zur Liste der ausgewählten Merkmale siehe unten stehende Auflistung) und unterschiedliche Bevölkerungskonzepte

- Bevölkerung in Privathaushalten,
- Bevölkerung in Privathaushalten am Ort ihrer Hauptwohnung,
- erwerbstätige Bevölkerung in Privathaushalten am Ort ihrer Hauptwohnung und
- Bevölkerung in Privathaushalten am Familienwohnsitz

wurden daher die Randverteilungen der ausgewählter Merkmale für die Erhebungsjahre 1996 bis 1999 (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 1996-1999 ausgewiesen. Die entsprechenden Tabellen finden Sie in einem gesonderten Skript.

Zur Illustration finden Sie nachfolgend die Tabelle der Randverteilungen der Bevölkerung in Privathaushalten nach ausgewählten Merkmalen im Mikrozensus-Originalfile und dem Querschnitt des Mikrozensus-Panels mit Personengewichtung für das Erhebungsjahr 1996 (im Anhangband ist dies die Tabelle 1a).

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 in Privathaushalten nach ausgewählten Merkmalen im MZ-Originalfile und dem Querschnitt des MZ-Panels im Erhebungsjahr 1996 mit Personengewichtung

Bevölkerung in Privathaushalten im Erhebungsjahr 1996 mit Personengewichtung						
Gegenstand der Nachweisung	MZ-Originalfile		Querschnitt des MZ-Panels		Differenz	
	Anzahl	Prozent	Anzahl	Prozent	absolut	relativ
Insgesamt	82 069	100.0	82 069	100.0	0	0.00
davon						
im Bundesland						
Schleswig-Holstein	2 770	3.4	2 771	3.4	0	0.02
Hamburg	1 710	2.1	1 711	2.1	1	0.04
Niedersachsen	7 779	9.5	7 775	9.5	-4	-0.05
Bremen	678	0.8	671	0.8	-7	-1.01
Nordrhein-Westfalen	17 842	21.7	17 848	21.7	7	0.04
Hessen	6 055	7.4	6 056	7.4	1	0.02
Rheinland-Pfalz	4 012	4.9	4 020	4.9	7	0.19
Baden-Württemberg	10 413	12.7	10 401	12.7	-12	-0.11
Bayern	12 137	14.8	12 156	14.8	18	0.15
Saarland	1 076	1.3	1 078	1.3	3	0.26
Berlin	3 511	4.3	3 495	4.3	-16	-0.46
Brandenburg	2 533	3.1	2 526	3.1	-7	-0.27
Mecklenburg-Vorpommern	1 813	2.2	1 809	2.2	-4	-0.23
Sachsen	4 541	5.5	4 549	5.5	8	0.18
Sachsen-Anhalt	2 709	3.3	2 719	3.3	10	0.35
Thüringen	2 490	3.0	2 484	3.0	-6	-0.23
nach dem Geschlecht						
männlich	40 133	48.9	40 038	48.8	-95	-0.24
weiblich	41 937	51.1	42 031	51.2	95	0.23
nach dem Familienstand						
ledig	31 365	38.2	31 199	38.0	-165	-0.53
verheiratet	40 745	49.6	40 816	49.7	71	0.17
verwitwet	6 278	7.6	6 319	7.7	41	0.66
geschieden	3 681	4.5	3 734	4.6	53	1.45

Fortsetzung der Tabelle auf nächster Seite

Bevölkerung in Privathaushalten im Erhebungsjahr 1996 mit Personengewichtung						
Gegenstand der Nachweisung	MZ-Originalfile		Querschnitt des MZ-Panels		Differenz	
	Anzahl	Prozent	Anzahl	Prozent	absolut	relativ
Insgesamt	82 069	100.0	82 069	100.0	0	0.00
davon						
nach Altersgruppen						
0 < 5 Jahre	3 843	4.7	3 860	4.7	18	0.47
5 < 10 Jahre	4 620	5.6	4 499	5.5	-122	-2.63
10 < 15 Jahre	4 537	5.5	4 550	5.5	12	0.27
15 < 20 Jahre	4 479	5.5	4 579	5.6	100	2.23
20 < 25 Jahre	4 739	5.8	4 665	5.7	-74	-1.56
25 < 30 Jahre	6 306	7.7	6 321	7.7	15	0.23
30 < 35 Jahre	6 918	8.4	6 858	8.4	-60	-0.86
35 < 40 Jahre	6 305	7.7	6 301	7.7	-4	-0.06
40 < 45 Jahre	5 803	7.1	5 708	7.0	-95	-1.64
45 < 50 Jahre	5 368	6.5	5 413	6.6	45	0.84
50 < 55 Jahre	5 127	6.2	5 024	6.1	-103	-2.01
55 < 60 Jahre	6 308	7.7	6 305	7.7	-3	-0.05
60 < 65 Jahre	4 740	5.8	4 689	5.7	-51	-1.07
65 < 70 Jahre	4 309	5.3	4 460	5.4	151	3.51
70 < 75 Jahre	3 583	4.4	3 661	4.5	77	2.16
75 < 80 Jahre	2 178	2.7	2 252	2.7	74	3.39
≥ 80 Jahre	2 905	3.5	2 924	3.6	19	0.65
nach der Staatsangehörigkeit						
Deutsche(r)	74 914	91.3	74 920	91.3	6	0.01
Ausländer(in) aus einem EU-Staat	1 828	2.2	1 768	2.2	-60	-3.28
Ausländer(in) aus einem Nicht-EU-Staat ..	5 327	6.5	5 381	6.6	54	1.01
nach Art der Erwerbsbeteiligung						
erwerbstätig	36 459	44.4	36 190	44.1	-269	-0.74
erwerbslos nach ILO	3 476	4.2	3 459	4.2	-17	-0.50
sonstig erwerbslos	533	0.6	515	0.6	-18	-3.36
Nichterwerbsperson	41 601	50.7	41 905	51.1	304	0.73

Fortsetzung der Tabelle auf nächster Seite

Bevölkerung in Privathaushalten im Erhebungsjahr 1996 mit Personengewichtung						
Gegenstand der Nachweisung	MZ-Originalfile		Querschnitt des MZ-Panels		Differenz	
	Anzahl	Prozent	Anzahl	Prozent	absolut	relativ
Insgesamt	82 069	100.0	82 069	100.0	0	0.00
davon						
nach der Stellung im Beruf in der gegenwärtigen Tätigkeit						
Selbstständig ohne Beschäftigte	1 685	2.1	1 668	2.0	-16	-0.96
Selbstständig mit Beschäftigten	1 818	2.2	1 793	2.2	-25	-1.35
Mithelfende(r) Familienangehörige(r)	391	0.5	379	0.5	-12	-3.04
Beamtin/Beamter, Richter(in)	2 110	2.6	2 137	2.6	28	1.32
Angestellte(r)	16 260	19.8	16 098	19.6	-163	-1.00
(Heim-)Arbeiter(in)	12 211	14.9	12 122	14.8	-89	-0.73
Kaufm./techn. Auszubildende(r)	707	0.9	716	0.9	9	1.21
Gewerbl. Auszubildende(r)	767	0.9	777	0.9	11	1.40
Zeit-/Berufssoldat(in) (einschl. BGS und Bereitschaftspolizei) ..	250	0.3	245	0.3	-4	-1.74
Grundwehrdienstleistender	166	0.2	159	0.2	-7	-4.05
Zivildienstleistender	96	0.1	95	0.1	-1	-0.96
trifft nicht zu (Person < 15 Jahre, nicht erwerbstätig)	45 610	55.6	45 879	55.9	269	0.59
nach Voll-/Teilzeittätigkeit						
Vollzeittätigkeit	30 417	37.1	30 184	36.8	-233	-0.76
Teilzeittätigkeit	6 042	7.4	6 006	7.3	-37	-0.61
trifft nicht zu (Person < 15 Jahre, nicht erwerbstätig)	45 610	55.6	45 879	55.9	269	0.59

Fortsetzung der Tabelle auf nächster Seite

Bevölkerung in Privathaushalten im Erhebungsjahr 1996 mit Personengewichtung						
Gegenstand der Nachweisung	MZ-Originalfile		Querschnitt des MZ-Panels		Differenz	
	Anzahl	Prozent	Anzahl	Prozent	absolut	relativ
Insgesamt	82 069	100.0	82 069	100.0	0	0.00
davon						
nach dem überwiegenden Lebensunterhalt						
Erwerbs-/Berufstätigkeit	34 301	41.8	34 079	41.5	-222	-0.65
Arbeitslosengeld bzw.						
Arbeitslosenhilfe	2 878	3.5	2 855	3.5	-22	-0.78
Rente, Pension	16 408	20.0	16 761	20.4	353	2.15
Unterhalt durch Eltern, Ehemann/-frau oder andere Angehörige ...	25 243	30.8	25 182	30.7	-61	-0.24
eigenes Vermögen, Ver- mietung, Zinsen, Alten- teil	383	0.5	406	0.5	23	6.05
Sozialhilfe	1 535	1.9	1 519	1.9	-16	-1.01
Leistungen aus einer Pflegeversicherung	86	0.1	99	0.1	13	14.71
sonstige Unterstützungen	1 236	1.5	1 168	1.4	-68	-5.48
nach den Individualeinkommensklassen						
< 300 DM	2 4250	3.0	2 457	3.0	31	1.30
300 < 600 DM	4 5750	5.6	4 669	5.7	94	2.06
600 < 1 000 DM	6 1230	7.5	6 209	7.6	86	1.41
1 000 < 1 400 DM	6 9839	8.5	6 937	8.5	-47	-0.68
1 400 < 1 800 DM	6 9246	8.4	6 925	8.4	1	0.01
1 800 < 2 200 DM	8 0199	9.8	8 002	9.8	-18	-0.22
2 200 < 2 500 DM	6 0122	7.3	5 953	7.3	-59	-0.98
2 500 < 3 000 DM	6 0674	7.4	5 974	7.3	-93	-1.54
3 000 < 3 500 DM	3 9387	4.8	3 914	4.8	-25	-0.63
3 500 < 4 000 DM	2 2762	2.8	2 236	2.7	-40	-1.77
4 000 < 4 500 DM	1 5265	1.9	1 487	1.8	-40	-2.62
4 500 < 5 000 DM	1 0702	1.3	1 048	1.3	-22	-2.03
5 000 < 5 500 DM	7424	0.9	726	0.9	-16	-2.22
5 500 < 6 000 DM	4958	0.6	489	0.6	-6	-1.30
6 000 < 6 500 DM	3724	0.5	356	0.4	-16	-4.39
6 500 < 7 000 DM	2482	0.3	256	0.3	8	3.06
7 000 < 7 500 DM	1821	0.2	186	0.2	4	2.30
≥ 7 500 DM	7936	1.0	784	1.0	-10	-1.21
selbstständige(r) Land- wirt(in) in der Haupt- tätigkeit	3379	0.4	320	0.4	-18	-5.31
kein Einkommen	20 0685	24.5	20 093	24.5	24	0.12
keine Angabe	2 8856	3.5	3 048	3.7	162	5.63

Fortsetzung der Tabelle auf nächster Seite

Bevölkerung in Privathaushalten im Erhebungsjahr 1996 mit Personengewichtung						
Gegenstand der Nachweisung	MZ-Originalfile		Querschnitt des MZ-Panels		Differenz	
	Anzahl	Prozent	Anzahl	Prozent	absolut	relativ
Insgesamt	82 069	100.0	82 069	100.0	0	0.00
davon						
nach dem höchsten allgemeinbildenden Schulabschluss						
Volks-/Hauptschulabschluss	32 689	39.8	32 087	39.1	-603	-1.84
Abschluss der allgemeinbildenden POS der ehemaligen DDR	5 837	7.1	5 665	6.9	-173	-2.96
Realschul- oder gleichwertiger Abschluss ...	10 743	13.1	10 766	13.1	23	0.22
Fachhochschulreife	2 288	2.8	2 227	2.7	-60	-2.64
Allgemeine oder fachgebundene Hochschulreife	9 369	11.4	9 352	11.4	-16	-0.17
keine Angabe	375	0.5	452	0.6	77	20.51
trifft nicht zu (Person < 15 Jahre, kein allgemeinbildender Abschluss bzw. keine Angabe hierzu)	20 768	25.3	21 520	26.2	752	3.62
nach dem höchsten beruflichen Ausbildungsabschluss						
Anlernausbildung, berufliches Praktikum	1 825	2.2	1 837	2.2	12	0.66
Abschluss einer Lehrausbildung oder gleichwertigem Berufsfachschulabschluss	31 382	38.2	30 531	37.2	-851	-2.71
Meister-/Technikerabschluss oder gleichwertiger Fachschulabschluss	3 659	4.5	3 647	4.4	-12	-0.32
Abschluss der Fachschule der ehemaligen DDR .	1 101	1.3	1 071	1.3	-30	-2.74
Fachhochschulabschluss (auch Ingenieurschulabschluss)	2 506	3.1	2 468	3.0	-38	-1.53
Hochschulabschluss	4 327	5.3	4 347	5.3	20	0.46
keine Angabe	524	0.6	501	0.6	-23	-4.44
trifft nicht zu (Person < 15 Jahre, kein beruflicher Abschluss bzw. keine Angabe hierzu) .	36 746	44.8	37 669	45.9	923	2.51

3.6 Metainformationen zum Mikrozensus-Panel 1996-1999

Neben dem vorliegenden Handbuch und der Randverteilung ausgewählter Merkmale zum Mikrozensus-Panel 1996-1999 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 1996-1999 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 1996-1999

Dateiname	Inhaltsbeschreibung
Randverteilungen.pdf	Randverteilungen ausgewählter Merkmale aus den Mikrozensus-Originalfiles und den Querschnitten des Mikrozensus-Panels 1996-1999 für die Erhebungsjahre 1996 bis 1999
FB96_oEP.pdf	Fragebogen für Selbstausfüller ohne Ergänzungsprogramm für das Erhebungsjahr 1996
FB96_mEP.pdf	Fragebogen für Selbstausfüller mit Ergänzungsprogramm für das Erhebungsjahr 1996
FB97_oEP.pdf	Fragebogen für Selbstausfüller ohne Ergänzungsprogramm für das Erhebungsjahr 1997
FB97_mEP.pdf	Fragebogen für Selbstausfüller mit Ergänzungsprogramm für das Erhebungsjahr 1997
FB98_oEP.pdf	Fragebogen für Selbstausfüller ohne Ergänzungsprogramm für das Erhebungsjahr 1998
FB98_mEP.pdf	Fragebogen für Selbstausfüller mit Ergänzungsprogramm für das Erhebungsjahr 1998
FB99_oEP.pdf	Fragebogen für Selbstausfüller ohne Ergänzungsprogramm für das Erhebungsjahr 1999
FB99_mEP.pdf	Fragebogen für Selbstausfüller mit Ergänzungsprogramm für das Erhebungsjahr 1999
FMZP9699_SV.pdf	Schlüsselverzeichnis des faktisch anonymisierten Mikrozensus-Panels 1996-1999 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)
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)
HFR85_MZ.pdf	Gliederung der Hauptfachrichtung des Abschlusses an einer (Fach-)Hochschule – Ausgabe 1985 – (HFR 85)

Bitte beachten Sie, dass die WZ 93 und die HFR 85 nicht den aktuellen Stand der jeweiligen Systematik wiedergeben, hier aber auf sie Bezug genommen wird, weil sie für den Erhebungszeitraum des Mikrozensus-Panels 1996-1999 Gültigkeit hatten.

Für die Systematik der Wirtschaftszweige liegt seit dem Jahr 2003 eine neue Klassifikation vor und die Gliederung der Abschlüsse an einer (Fach-)Hochschule wurde in den Jahren 2000 und 2003 revidiert und gilt nun auch für die beruflichen Abschlüsse insgesamt.

3.6.1 Das Schlüsselverzeichnis des faktisch anonymisierten Mikrozensus-Panels 1996-1999

Das Schlüsselverzeichnis beinhaltet die Beschreibung des faktisch anonymisierten Mikrozensus-Panels 1996-1999 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 FMZP9699.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 1996 beginnt der Variablenname mit einem **a**, bei Variablen des Erhebungsjahres 1997 beginnt der Variablenname mit einem **b** usw. Die erste Stelle eines Variablennamens kennzeichnet also den Erhebungszeitpunkt. Nur bei Variablen, die längsschnittspezifische bzw. längsschnittkonsistente Informationen beinhalten, wird von dieser Regel abgewichen.

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 Werteausprägungen können Sie der zweiten Spalte des Schlüsselverzeichnisses "Variablelabel und Wertelabel in den einzelnen Querschnitterhebungen 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 1998, ist somit wie folgt gegeben:

[EF1] MZ98: 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 Fragenummer 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 1996-1999 wird der Wissenschaft in vier verschiedenen Datenformaten bereitgestellt. Als SPSS-Datenfile, als SAS-Datenfile, als Stata-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: FMZP9699.dat" angegeben, wobei FMZP9699.dat der Name der ASCII-Datei ist.

Weiterhin sind noch zwei Aspekte zu beachten:

1. Variablen, die auf Fragen basieren, deren Beantwortung freiwillig sind und die aus diesen Fragen abgeleiteten Variablen sind im Schlüsselverzeichnis hellgrau hinterlegt.
2. 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 1996-1999 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 3.6: Ausfallart sowie zugehöriger Variablenwert und Wertelabel im Mikrozensus-Paneldatensatz 1996-1999

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

Bitte beachten Sie, dass die Vergabe dieser generellen Codes im Schlüsselverzeichnis bei den einzelnen Variablen nicht aufgeführt wird und die Werte in den Datendateien nicht als `missing`-Werte deklariert wurden.

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 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 3.10: 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 3.11: 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 Klassifikationen ist also gemeinsam, dass sie eine hierarchische Gliederung beruflicher Tätigkeiten vornehmen, 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 mit 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 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.

3.6.4 Gliederung der Hauptfachrichtung des Abschlusses an einer (Fach-)Hochschule – Ausgabe 1985

Für eine systematische Erfassung und Darstellung der Hauptfachrichtung des höchsten Abschlusses an einer Fachhochschule oder Hochschule im Mikrozensus, liegt in der Ausgabe von 1985 ein hierarchisch, zweigliedriges Ordnungssystem vor, das auf der untersten, der zweiten Gliederungsebene 78 Fachrichtungen ausweist und auf der ersten Gliederungsebene neun breite Felder.

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.)

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.

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.
- 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/download/d/mv/arbeitspapier12.pdf>
- 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 (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/download/d/mv/arbeitspapier11.pdf>

Kapitel 4

Das Hochrechnungsverfahren für Längsschnittauswertungen im Mikrozensus-Panel 1996-1999

Ivo Marek (Freie Universität Berlin¹)

Sandra Rohloff (Statistisches Bundesamt, Gruppe II A²)

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 1996 und 1999 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, würde

¹Mobilitätsgewichtung.

²Hochrechnung.

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 die, in der Grundgesamtheit bekannten Verteilungen, auch andere, durch die räumliche Mobilität bedingte Abweichungen von Verteilungen in der Stichprobe an die entsprechenden Verteilungen in der Grundgesamt angepasst 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 Regressions-schätzung 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 innerhalb der durch die Hilfsvariablen definierten Teilgruppen der Ausfallmechanismus unabhängig von den Untersuchungsvariablen ist.

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 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 ein Logit-Modell ("Logistic Regression Weighting"). 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 Logistic Regression Weighting

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 Immobilien 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 unbekannt 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 ein Logit-Modell – für Immobilität berechnen.

Um das Konzept formal darzustellen, wird zunächst die folgende Notation eingeführt. Es sei U die Längsschnittgrundgesamtheit für einen vorgegebenen Zeitraum $[t_i, t_j]$, wobei t_i der Anfangszeitpunkt und t_j der Endzeit-

punkt ist. Eine beliebige Person der Grundgesamtheit wird mit $k \in U$ bezeichnet. Für jede Person $k \in U$ existiert ein Vektor von $J + 1$ Merkmalen $\mathbf{x}_k = (1, x_{k1}, x_{k2}, \dots, x_{kJ})'$. Das Ziel der Schätzung sei der Gesamtwert $t_y = \sum_U y_k$, wobei y_k der Wert der Untersuchungsvariablen y für die k -te Person ist.

Alle Personen $k \in U$ haben ein Design-Gewicht $d_k > 0$, dass durch das Ziehungsdesign des Mikrozensus festgelegt ist. Die realisierte Stichprobe sei $s \subset U$. Die immobilen Personen werden in der Menge $r \subset s$ zusammengefasst. Für diese sei das Längsschnittmerkmal y bekannt. Zusätzlich zeigt der Indikator m_k an, ob eine Person $k \in U$ im Zeitraum $[t_i, t_j]$ mobil ist ($m_k = 1$) oder nicht ($m_k = 0$). Für alle $k \in r$ gilt $m_k = 0$. Sowohl für die immobilen Personen als auch mobilen Personen ist der Vektor x_k bekannt.

Zur Berechnung eines unverzerrten Schätzwertes für t_y muss die Wahrscheinlichkeit am Wohnort im Zeitraum $[t_i, t_j]$ zu verbleiben $\theta_k = P(m_k = 0)$ für alle immobilen Personen durch ein Modell vorhergesagt werden. Die unbekanntenen θ_k werden dann durch die Schätzwerte $\hat{\theta}_k$ ersetzt und die modifizierten Gewichte ergeben sich aus $w_k = \hat{\theta}_k^{-1} d_k$. Die Berechnung eines Schätzwertes für t_y erfolgt durch

$$\hat{t}_y = \sum_r w_k y_k$$

Zur Berechnung von Korrekturgewichten $\hat{\theta}_k^{-1}$ wird im Mikrozensus der Ansatz des „Logistic Regression Weighting“ verfolgt. Bei „Logistic Regression Weighting“ ist das Mobilitätsmodell von der Form: $\theta_k = 1 + e^{-\mathbf{x}_k' \lambda}$ für $k \in U$. Als Schätzwert für λ kann der ML-Schätzer der Logistischen Regression gewählt werden. Hierbei gilt für das Merkmal y , vgl. Harms (2005):

$$E(\hat{t}_y | s) \approx \sum_s d_k y_k.$$

Da die statistische i.i.d.-Annahme in Bezug auf räumliche Mobilität zwischen den einzelnen Haushaltsmitgliedern nicht zutrifft³, aber dennoch die Mobilität in einem Haushalt berücksichtigt werden soll, wird das Mobilitätsmodell auf Basis von synthetischen Haushalten geschätzt. Ein synthetischer Haushalt ist eine Gruppe von in einem Haushalt zusammenlebenden Individuen, die das gleiche Mobilitätsverhalten aufweisen. Liegt beispielsweise ein Drei-Personenhaushalt vor und eine Person zieht um, so wird dieser Haushalt in zwei synthetische Haushalte untergliedert. Der erste synthetische Haushalt umfasst die räumlich mobile Person und der zweite die beiden räumlich immobilen Personen.

³Die i.i.d.-Annahme würde im Falle der räumlichen Mobilität besagen, dass die Zufallsvariable der Mobilitätswahrscheinlichkeiten der einzelnen Haushaltsmitglieder voneinander unabhängig sind und dieselbe Verteilung haben. Die Abkürzung i.i.d. steht für: independent, identically distributed.

Für die Beschreibung der Korrekturgewichte $\hat{\theta}_k^{-1}$ gehen mehrere haushalts-spezifische und personengebundene Merkmale als erklärende Variablen ein. Insgesamt werden folgende Merkmale als erklärende Variablen genutzt um die Korrekturgewichte $\hat{\theta}_k^{-1}$ zu bestimmen.

Tabelle 4.1: Erklärende Variablen für die Schätzung der Mobilitätsgewichte im Mikrozensus-Panel 1996-1999

Merkmal	Ausprägung	Anzahl Personen im Haushalt mit Ausprägung ...
Altersgruppen:	unter 15 Jahre	1 / mehr als 1
	15 bis 24 Jahre	alle (Referenz)
	25 bis 34 Jahre	1 / mehr als 1
	35 bis 44 Jahre	1 / mehr als 1
	45 bis 54 Jahre	1 / mehr als 1
	55 bis 64 Jahre	1 / mehr als 1
	über 65 Jahre	1 / mehr als 1
Bildungsabschluss	niedrig (k.A., kein Abschluss)	1 / mehr als 1
	mittel (Haupt-, Realschule u.ä.)	alle (Referenz)
	hoch (Abitur, Fachhochschulreife)	1 / mehr als 1
	Person in allgemeinbildender Schule	1 / mehr als 1
	Person in weiterführender Schule	1 / mehr als 1
Ausbildungsabschluss	niedrig (keine Ausbildung, k.A.)	1 / mehr als 1
	mittel (Lehre, Fachschulabschluss u.ä.)	alle (Referenz)
	hoch (Fach-, Hochschulabschluss)	1 / mehr als 1
Erwerbsstatus	erwerbstätig	alle (Referenz)
	erwerbslos	1 / mehr als 1
	Nichterwerbsperson	1 / mehr als 1
Familienstand	verheiratet	1 / mehr als 1
	nicht verheiratet	alle (Referenz)
Haushaltsgröße:	1 Person	alle
	2 Personen	alle (Referenz)
	3 Personen	alle
	4 Personen	alle
	mehr als 4 Personen	alle
Nettoeinkommen	bis unter 1 800 DM	alle (Referenz)
	1 800 DM und mehr	1 / mehr als 1
	kein Einkommen	1 / mehr als 1
	Selbständige/r Landwirt/in	mind. 1
	keine Angabe	mind. 1
Region:	Früheres Bundesgebiet ohne Berlin	alle (Referenz)
	Neue Länder mit Berlin	alle
Staatsangehörigkeit:	deutsch	alle (Referenz)
	nicht deutsch	1 / mehr als 1

Hier werden Merkmale verwendet die für immobile und mobile Personen bekannt sind, das sind insbesondere die Querschnittsmerkmale zu einem Zeitpunkt.

Im Scientific Use File stehen die Kehrwerte der geschätzten Bleibewahrscheinlichkeiten $\hat{\theta}_k$ 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 4.6 entnehmen.

4.3 Grundgesamtheit und Stichprobe des Mikrozensus-Panels 1996-1999

Die Grundgesamtheit, auf die für das Mikrozensus-Panel 1996-1999 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 $i=1996$ und – je nach zu analysierendem Längsschnitt – $j=1997, 1998$ oder 1999) 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 1996-1999

Um Aussagen über die Gesamtbevölkerung Deutschlands in den Jahren 1996 bis 1999 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 möglichen, 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 angepasst⁴.

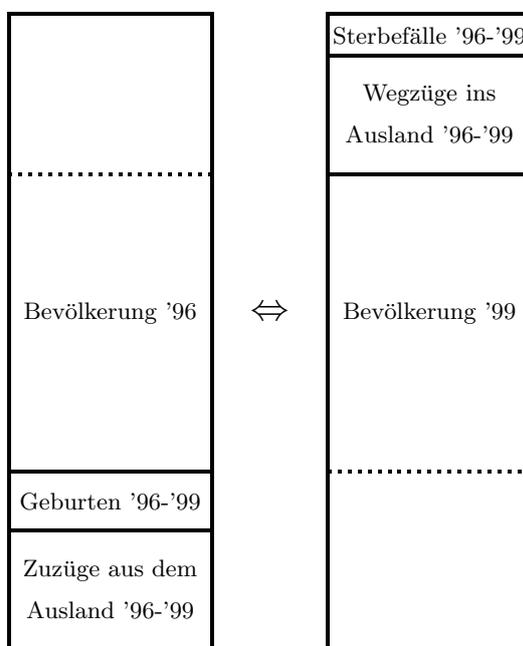
Die Anpassung des Längsschnitts erfolgt parallel an die Bevölkerung von 1996 (t_i) und an die Bevölkerung des im Längsschnitt zuletzt betrachteten Jahres (t_j). Bei Verwendung der Gewichte ist lediglich darauf zu achten, dass die Summe der betrachteten Eckwertvektoren gleich ist. Hierzu werden zu den Angaben der immobilen Personen die Zahlen der Geburten, Zuzüge aus dem Ausland, Sterbefälle und Fortzüge ins Ausland addiert.

Betrachtet man beispielsweise den Längsschnitt 1996-1999, so werden Personen, die nach der Erhebung 1996 sterben oder ins Ausland ziehen, zwar 1996

⁴Diese parallele Randanpassung von Stichprobenergebnisse an bekannte bzw. geschätzte Randverteilungen wird als Kalibrierung bezeichnet.

befragt, doch sie fehlen in der Menge der befragten Personen von 1999. Um dieses Defizit auszugleichen, wird diese Menge zu derjenigen der immobilen Personen von 1999 hinzugefügt. Ebenso wurden Geburten und Zuzüge aus dem Ausland zwischen 1996 und 1999 nicht in der Stichprobenmenge 1996 erfasst. Diese Personen werden daher zur Menge der Personen von 1996 dazugezählt. Jede Person der betrachteten Stichprobe muss in der Menge der Personen von 1996 und in der Menge der Personen von 1999 vertreten sein. Beide Mengen sind in Abbildung 4.1 abgebildet.

Abbildung 4.1: Definition der beiden betrachteten Mengen für die Randanpassung des Längsschnitts 1996-1999



Personen, die während des gesamten betrachteten Längsschnitts im Mikrozensus erhoben wurden, sind in den Mengen Bevölkerung 1996 und Bevölkerung 1999 vertreten. Geburten sind unter Geburten 1996-1999 und in der Menge Bevölkerung 1999 zu finden. Analog gilt dies für Zuzüge aus dem Ausland. Sterbefälle und Fortzüge ins Ausland sind in der Menge Bevölkerung 1996 und 1999 in der Menge Sterbefälle 1996-1999 bzw. Fortzüge ins Ausland 1996-1999 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 aus der Bevölkerung durch eine verallgemeinerte Regressionsschätzung, 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 2 multipliziert werden, da die Ereignisse fallbezogen vorliegen, die Gewichtungsfaktoren aber für der Personenebene ermittelt werden.

In die Berechnung von g_k geht dann die Differenz des Vektors \mathbf{t}_x der Gesamtwerte der Hilfsmerkmale und des Horwitz-Thompson-Schätzers⁵ der Gesamtwerte ($\hat{\mathbf{t}}_{x,HT}$) 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 diese stets für die Gesamtbevölkerung Deutschlands ermittelt werden. Da das Mikrozensus-Panel 1996-1999 nur aus Personen in Privathaushalten besteht, ist die Bevölkerungszahl entsprechend anzupassen. In den Mikrozensus-Originalfiles stehen für alle vier Jahre die Hochrechnungsfaktoren für die Gesamtbevölkerung Deutschlands zur Verfügung. Die Zahl der Personen in Gemeinschaftsunterkünften in der Grundgesamtheit kann somit anhand des Hochrechnungsfaktors geschätzt werden. Diese kann anschließend von der Bevölkerungszahl subtrahiert werden, um so die Anzahl der Personen in Privathaushalten in Deutschland zu erhalten. Da für Personen in Gemeinschaftsunterkünften die personengebundenen Merkmale wie Geschlecht, Alter und Staatsangehörigkeit zur Verfügung stehen, kann

⁵Der 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.

von jedem Eckwert die entsprechende Zahl der Personen in Gemeinschaftsunterkünften abgezogen werden. Bei den Eheschließungen, Ehescheidungen, Zuzügen aus dem Ausland und Geburten ist dies nicht möglich, hier werden die Personen in Gemeinschaftsunterkünften vernachlässigt. Man geht stattdessen davon aus, dass die Ereignisse wie Eheschließungen, Ehescheidungen, Geburten und Zuzüge aus dem Ausland in Gemeinschaftsunterkünften so selten auftreten, dass ihre Anzahl vernachlässigbar ist. Bei Sterbefällen ist diese Zahl nicht vernachlässigbar, weshalb sie als Differenz der anderen Eckwerte wie folgt ermittelt wird:

$$N_{Verl,t_i,t_j} = (N_{Quer,t_i} + N_{Geb,t_i,t_j} + N_{Zuzug,t_i,t_j}) - N_{Quer,t_j}$$

mit:

- N_{Quer,t_i} : Anzahl der Bevölkerung in Privathaushalten im Erhebungszeitpunkt t_i
- N_{Quer,t_j} : Anzahl der Bevölkerung in Privathaushalten im Erhebungszeitpunkt t_j
- N_{Geb,t_i,t_j} : Anzahl der Geburten zwischen $t_{i,Mai}$ und $t_{j,April}$ in Privathaushalten
- N_{Zuzug,t_i,t_j} : Anzahl der Zuzüge aus dem Ausland in Privathaushalte zwischen $t_{i,Mai}$ und $t_{j,April}$
- N_{Verl,t_i,t_j} : Anzahl der Sterbefälle in Privathaushalten und Fortzüge ins Ausland aus Privathaushalten zwischen $t_{i,Mai}$ und $t_{j,April}$

Dadurch erhält man auch hier näherungsweise die Anzahl der Personen in Privathaushalten.

Ein weiteres Problem bei den Eckwerten besteht darin, 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 Interpolation eines Teils der Eckwerte und die Hochrechnung der Gemeinschaftsunterkünfte im Querschnitt können zu Ungenauigkeiten in den Eckwerten führen. Subtrahiert man jedoch die Zahl der Personen in Gemeinschaftsunterkünften nicht, so führt dies zu einer Übergewichtung der Personen in Privathaushalten. Auch eine fehlende Interpolation kann die Ergebnisse verzerren. Es handelt sich hier in beiden Fällen zwar um keine optimalen Lösungen, aber um die momentan bestmöglichen. In Zukunft kann die Qualität der Hochrechnung verbessert werden, wenn möglichst zeitnahe Eckwerte aus der Bevölkerungsfortschreibung vorliegen und diese die gleiche Personengruppe umfasst wie die Stichprobe.

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 4.6 entnehmen.

4.5 Ein Beispiel

Eine wesentliche Auswertungsform des Mikrozensus-Panels ist die Schätzung von Aggregaten und Anteilswerten zur wirtschaftlichen und sozialen Lage von Personen in der Bundesrepublik Deutschland. Es geht beispielsweise um die Frage, wie hoch der Anteil der Personen ist, die in einem bestimmten Zeitintervall aus der Erwerbslosigkeit ins Berufsleben zurückgefunden haben. In diesem Abschnitt soll die Anwendung der verschiedenen Gewichtungsfaktoren für diese Fragestellung illustriert werden.

Ausgangspunkt ist die Schätzung von Erwerbsübergängen im Zeitraum 1996 bis 1999. Hierbei werden jeweils für die Jahre 1996 und 1999 drei verschiedene Erwerbszustände unterschieden: erwerbstätig (E), erwerbslos (U) und Nichterwerbsperson (N). Die Abgrenzung der Erwerbszustände erfolgt über die Variable `Erwerbstyp`, `af504` für 1996 und `df504` für 1999.⁶ Somit ergeben sich neun Erwerbsübergänge.

Für die obige Fragestellung ist der Übergang erwerbslos (U) - erwerbstätig (E) von Interesse. Der Anteil der Personen, die im Zeitraum 1996 bis 1999 von der Erwerbslosigkeit ins Berufsleben zurückgefunden haben, bestimmt sich aus dem Anteil der in 1999 erwerbstätigen Personen unter den in 1996 erwerbslosen Personen. Es werden nur Personen berücksichtigt, die 1996 älter als 15 Jahre und 1999 jünger als 66 Jahre sind.⁷ Es ergeben sich für die drei verschiedenen Gewichte: 38.49%, 37.96% bzw. 38.39%. Das heißt der Anteil der erwerbslosen Personen die in das Berufsleben zurückgefunden haben liegt bei ca 38%.

4.6 Die im Mikrozensus-Panel 1996-1999 bereitgestellten Hochrechnungs- und Mobilitätsfaktoren

Insgesamt stehen im Mikrozensus-Panel 1996-1999 zwei unterschiedliche Längsschnittgewichte und ein Mobilitätsgewicht für jeweils unterschiedliche Zeiträu-

⁶Die Variable `Erwerbstyp` hat die Ausprägungen Erwerbstätig (1), Erwerbslos (2), Sonstig Erwerbslos (3) und Nichterwerbsperson (4). Die beiden Ausprägungen Erwerbslos und Sonstig Erwerbslos werden zusammengefasst.

⁷Das Alter der Personen bestimmt sich aus der Variable `gebj` (Geburtsjahr).

me 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 1996-1999

Variablenname	Gewichtungsfaktor	Übergang
HG_9697	Längsschnittgewicht zur Hochrechnung ohne Berücksichtigung der Mobilitätsgewichte für den Übergang ...	1996 ⇒ 1997
HG_9698		1996 ⇒ 1998
HG_9699		1996 ⇒ 1999
HGM_9697	Längsschnittgewicht zur Hochrechnung mit Berücksichtigung der Mobilitätsgewichte für den Übergang ...	1996 ⇒ 1997
HGM_9698		1996 ⇒ 1998
HGM_9699		1996 ⇒ 1999
MG_9697	Mobilitätsgewicht (Kehrwert der Bleibewahrscheinlichkeit) für den Übergang ...	1996 ⇒ 1997
MG_9698		1996 ⇒ 1998
MG_9699		1996 ⇒ 1999
MG_9798		1997 ⇒ 1998
MG_9799		1997 ⇒ 1999
MG_9899		1998 ⇒ 1999

Wie Sie vorstehender Tabelle entnehmen können, stehen derzeit die Längsschnittgewichte zur Hochrechnung ohne und mit Berücksichtigung der Mobilitätsgewichte nur mit Ausgangsbasis des Erhebungsjahres 1996 zur Verfügung, wohingegen für die Mobilitätsgewichte auch die Erhebungsjahre 1997 und 1998 als Ausgangsbasis zur Verfügung stehen.

4.7 Weiterführende Literatur

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/download/d/mv/arbeitspapier3.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.

Harms, T. (2005): *Reweighting and Calibration Estimators for Complex Data Structures*. Dissertation.

Lundström, S.; Särndal, C.-E. (2002): *Estimation in the presence of Nonresponse and Frame Imperfections*. Stockholm: Statistics Sweden.

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/download/d/mv/arbeitspapier6.pdf>

Särndal, C.-E.; Swensson, B.; Wretman, J. (1992) *Model Assisted Survey Sampling*. New-York: Springer-Verlag.

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 be-

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.

ruhigend zu sehen, dass beim Vergleich der 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 Haushaltsabspaltungen, 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 Mobilen 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

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 .

Fall ist die Differenz unter der Null-Hypothese (kein Mobilitätsbias) approximativ multivariat normalverteilt mit Erwartungswertvektor 0 und einer Kovarianzmatrix 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

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

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.

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.

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

- 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.

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

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 Immobilien. 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)
	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)
	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 immobilien 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 Immoblie und

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

Gesamt (Immobilie 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.

Aufruf

Die Syntax des Makros ist wie folgt:

- dataset = Datensatz
- v1 = Zielgröße
- var1 = kategoriale Variable(n)
- var2 = metrische Variable(n)
- mob = Mobilitätsindikator mit Ausprägungen 0 für Immobilie 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 immobilien 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 MODELL 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 Datenanalyzesystems 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

Merkmale 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.

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

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

```

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 Responsevariable 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 immobilien 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                0.9154  5295.83  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                0.9725  640.85  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 -14.346 0.5951
118 2 1 -0.5190 0.0362 26.407 2.6994
119 2 2 0.9930 0.0376 26.407 2.6994
120 2 3 -0.4740 0.0376 26.407 2.6994
121 3 1 -0.9146 0.0376 26.407 2.6994
122 3 2 -0.6336 0.0376 26.407 2.6994
123 3 3 1.5482 0.0376 26.407 2.6994
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 0.0218 50.737 3.0292
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 0.0364 -1.933 0.9321
135 2 1 0.0510 0.0243 -2.101 0.9502
136 2 2 0.0703 0.0364 -1.933 0.9321
137 2 3 -0.1213 0.0364 -1.933 0.9321
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:

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

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 categorical 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-
tabgrenzungen*. Arbeitspapier #9, Methodenverbund "Aufbereitung und
Bereitstellung des Mikrozensus als Panelstichprobe".
URL: <http://www.destatis.de/download/d/mv/arbeitspapier9.pdf>
- Basic, E.; Marek, I.; Rendtel, U. (2005): *The German Microcensus as a tool
for longitudinal data analysis: An evaluation using SOEP data*. Arbeits-
papier #3, Methodenverbund "Aufbereitung und Bereitstellung des Mi-
krozensus als Panelstichprobe".
URL: <http://www.destatis.de/download/d/mv/arbeitspapier3.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 Bereitstel-
lung des Mikrozensus als Panelstichprobe".
URL: <http://www.destatis.de/download/d/mv/arbeitspapier8.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 Ger-
man 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 Fin-
nish subsample of the ECHP*. CHINTEX Working Paper #5.
URL: <http://www.destatis.de/chintex/download/paper5.pdf>
- Rendtel, U. (1995): *Panelausfälle und Panelrepräsentativität*. Frankfurt/M.,
New York: Campus Verlag.
- Robins, J., Rotnitzky, A., Zhao, L. (1995): *Analysis of semi-parametric regres-
sion 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.