

Dokumentation des Campus Files des Mikrozensus-Panels 1996-1999¹

Inhaltsverzeichnis

1. Einleitung.....	2
2. Das Campus File des Mikrozensus-Panels 1996-1999.....	3
2.1. Spezifische Anonymisierungsmaßnahmen	6
2.2 Die Stichprobenentwicklung des Campus Files des Mikrozensus- Panels 1996-1999.....	7
2.3 Die Bildung angepasster Hochrechnungsfaktoren für die Querschnittsanalysen im Campus-File des Mikrozensus-Panels 1996-1999 .	9
2.3.1 Angepasste Hochrechnungsfaktoren für deskriptive Auswertungen 10	
2.3.2 Angepasste Hochrechnungsfaktoren für inferenzstatistische Auswertungen	11
2.4 Die im Campus File des Mikrozensus-Panels 1996 – 1999 bereitgestellten Längsschnitt- und Mobilitätsgewichte	14
2.5 Randverteilungen ausgewählter Merkmale in den Mikrozensus- Originalfiles für ein Erhebungsjahr und den Querschnitten des Campus File des Mikrozensus-Panels 1996-1999.....	15
2.6 Berechnung der Stichprobenfehler und der Konfidenzintervalle für ausgewählte Merkmale	20
3. Beispiele für multivariate Längsschnittanalysen mit dem Campus File des MZ-Panels 1996 - 1999	24
3.1 Übergang aus abhängiger Beschäftigung in Erwerbslosigkeit	24
3.2 Übergang von Familienstand „verheiratet“ in den Familienstand „geschieden“	26
Literatur	29
4. Anhang.....	31

¹ Die Erstellung des Campus-Files des Mikrozensus-Panels 1996-1999 erfolgte beim Forschungsdatenzentrum NRW in enger Kooperation mit dem German Microdata Lab (GML). Aufgrund seiner umfangreichen Unterstützungs- und Beratungsleistungen gebührt Herrn Schimpl-Neimanns besonderen Dank. Zudem danken wir Herrn Rendtel (Freie Universität Berlin) für seine hilfreichen Kommentare bei der Erstellung der Dokumentation.

1. Einleitung

Längsschnittdaten spielen im Rahmen von sozio-ökonomischen Analysen in der empirischen Forschung seit geraumer Zeit eine zentrale Rolle. Zur Analyse dieser Daten wurden mittlerweile umfangreiche Methodiken und Analyseinstrumente entwickelt, die inzwischen auch in der Lehre an den Universitäten zum Standardrepertoire gehören. Leider stehen der positiven Entwicklung in der Methodenlehre bislang nur wenige „echte“ Längsschnitt-Einzeldatensätze für Auswertungen in dem praktischen Bereich der Statistikausbildung zur Verfügung. In der Regel behilft man sich dann mit selbst erstellten, kleineren Datensätzen. Mit dem jetzt vorliegenden Campus File des Mikrozensus-Panels 1996-1999 ist es nun erstmals möglich, neben den Daten des Sozio-oekonomischen Panels (SOEP) auch amtliche Längsschnitt-Einzeldaten in der Lehre zu verwenden.

Das Campus File des Mikrozensus-Panels ist ein so genanntes Public Use File (PUF), welches speziell für Lehrende und Studierende erstellt wurde. Public Use Files sind mittlerweile unverzichtbare, effektive Hilfsmittel, um Auswertungstechniken und -methoden in der praktischen Methodenausbildung auf Basis von amtlichen Einzeldaten kennen zu lernen und zu üben. Zurzeit gibt es von Seiten der amtlichen Statistik Public Use-Files zum Querschnittsdatensatz Mikrozensus 1998 und 2002, zu der Lohn- und Einkommenssteuerstatistik 1998 und 2001, zu der Sozialhilfestatistik 1998, zu der Kostenstrukturerhebung in kleineren und mittleren Unternehmen 1999, zu der Gehalts- und Lohnstrukturerhebung 2001 sowie zu der Europäischen Erhebung zur beruflichen Weiterbildung 2000.²

Das Campus-MZ-Panel-File stellt eine Substichprobe aus dem Mikrozensus-Panel 1996-1999 dar und wurde durch verschiedene Anonymisierungsmaßnahmen so verändert, dass der Datenschutz voll gewährleistet ist. Dennoch lassen sich mit seiner Hilfe, neben methodischen Analysen, modellhaft Studien zu vielen, insbesondere in den Sozialwissenschaften relevanten Fragestellungen durchführen. So besteht die Möglichkeit, die vielfältig enthaltenen sozio-ökonomischen Merkmale zur Familienstruktur oder zum Erwerbsverhalten bzw. zur „Erwerbsbeteiligung in Deutschland“ im Längsschnitt zu nutzen. Natürlich ist dabei zu beachten, dass die kleine 10% Stichprobe aus dem Originalmaterial, die das Campus-File darstellt, einen hohen Stichprobenfehler impliziert, weshalb bestimmte Verteilungen von den in der Grundgesamtheit erwarteten Verteilungen abweichen können. Die einzelnen durchgeführten Anonymisierungsmaßnahmen als auch der Stichprobenfehler für ausgewählte Merkmale und Jahre sind in dieser Dokumentation in den Abschnitten 2.1 bzw. 2.6 im Detail erläutert.

Trotz der stellenweise größeren Abweichungen gegenüber den Originalverteilungen lässt sich der Campus File MZ-Panel ohne weiteres zur Erstellung einer empirisch orientierten Hausarbeit verwenden. Eine möglicherweise darauf aufbauende Diplomarbeit/Dissertation könnte dann mit dem von den Forschungsdatenzentren (FDZ) bereitgestellten Scientific Use File (SUF) des Mikrozensus-Panel 1996 – 1999 zur Off-Site-Nutzung angefertigt werden, wenn diese Arbeit in ein Forschungsvorhaben einer inländischen Hochschule/Forschungseinrichtung integriert ist. Außerdem bieten die FDZ die Möglichkeit an, weniger anonymisiertes Datenmaterial an einem Gastwissenschaftlerarbeitsplatz oder über das so genannte Fernrechnen zu nutzen.

² Die Public-Use-Files sowie deren Beschreibung können im Internet unter der folgenden Adresse bezogen werden: <http://www.forschungsdatenzentrum.de/campus-file.asp>.

Da das Campusfile des Mikrozensus-Panels analog dem Scientific-Use-File (SUF) eine Substichprobe des Original MZ-Panels darstellt³, orientiert sich die vorliegende Dokumentation weitestgehend an den Ausführungen zum SUF des MZ-Panels in Kapitel 3 des „Handbuch Mikrozensus Panel 1996 – 1999“ (Statistisches Bundesamt 2006). Weitergehende allgemeine Informationen zum Mikrozensus bzw. zu dessen Stichprobenentwicklung, zum Datensatzaufbau des Mikrozensus-Panels 1996-1999 und vorhandener intertemporale Veränderungen von Erhebungsmerkmalen im Erhebungszeitraum sowie zu den möglichen Längsschnittauswertungskonzepten können dem Handbuch in Kapitel 1 und 2 entnommen werden.

2. Das Campus File des Mikrozensus-Panels 1996-1999

Das dem Campus-File zugrunde liegende Mikrozensus-Panel 1996-1999 basiert auf den vom Statistischen Bundesamt zusammengeführten Querschnitterhebungen eines Rotationsviertels (jährlich wird ein Viertel der Auswahlbezirke ausgetauscht) des Mikrozensus und umfasst den Zeitraum von 1996 – 1999.⁴ Im Mikrozensus werden alle Personen eines ausgewählten Bezirks befragt. Ein Auswahlbezirk umfasst dabei durchschnittlich neun Wohneinheiten. Die ausgewählten Bezirke werden in vier aufeinander folgenden Jahren in die Erhebung einbezogen, so dass die Paneldaten einen Zeitraum von vier Jahren abdecken. Es liegen jedoch nicht für alle Haushalte und Personen Informationen für den gesamten Zeitraum vor. Dies ist insbesondere dann der Fall, wenn Haushalte und Personen fortziehen. Solche Haushalte werden nicht weiter befragt, sondern durch die nachziehenden Bewohner(innen) ersetzt. (Herter-Eschweiler 2006a, S.2).

Gegenüber einer Haushaltssubstichprobe besteht der Vorteil, dass innerhalb der Auswahlbezirke Populationsveränderungen erfasst werden können, so dass insbesondere Fragestellungen zur räumlichen Mobilität und der damit verbundenen Stichprobenselektivität im Längsschnitt analysiert werden können. Nachteilig ist allerdings, dass mit diesem Konzept auch ein größerer Klumpeneffekt und damit ein höherer Stichprobenfehler verbunden ist (siehe Schimpl-Neimanns 2006).

Das Mikrozensus-Panel wurde ausführlich vom Statistischen Bundesamt beschrieben (Statistisches Bundesamt 2006). Das Campus-File des Mikrozensus Panels stellt eine 10 Prozent Substichprobe dieses Originalmaterials dar. Der Auswahlplan des Campus-Files des MZ-Panels entspricht – bis auf den Auswahlatz - dem Vorgehen beim faktisch anonymisierten MZ-Panel (Scientific-Use File) und ist detailliert in Tabelle 1 beschrieben (vgl. auch Schimpl-Neimanns 2006, Herter-Eschweiler 2003).

³ Zur Erstellung des Campus-Files wurden Programme verwendet, die auch schon zur Konstruktion des faktisch anonymisierten MZ-Panel 1996-1999 genutzt wurden. Einen besonderen Dank gilt daher Herrn Herter-Eschweiler, der uns die Programme zur Verfügung gestellt hat (siehe hierzu Herter-Eschweiler 2006b). Die Programme wurden entsprechend angepasst bzw. modifiziert.

⁴ Als Auswahlgrundlage des Mikrozensus dienen für das frühere Bundesgebiet Wohneinheiten, die im Rahmen der Volkszählung 1987 ermittelt wurden und für die neuen Bundesländer das damalige Einwohnermelderegister der ehemaligen DDR. Insgesamt wurden bei der VZ87 20 so genannte Vorratsstichproben mit einem Auswahlatz von einem Prozent nach einem systematischen Zufallsprinzip gezogen. Zusätzlich werden Neubaugebiete durch eine jährliche Ergänzungsauswahl aus der Bautätigkeitsstatistik mit einbezogen. Die Vorratsstichproben werden dann in Rotationsviertel aufgeteilt. Jedes Jahr wird eines davon ausgewählt, welches vier Jahre im Mikrozensus verbleibt und dann wieder ausscheidet.

Rotation im Mikrozensus und Zusammenführung der Längsschnittstichprobe

Abbildung 1 ist zu entnehmen, wie sich die einzelnen Mikrozensusquerschnitte aus den Rotationsgruppen zusammensetzen. Im Mikrozensus wird jährlich ein Viertel aller in der Stichprobe enthaltenen Auswahlbezirke ausgetauscht (Verfahren der partiellen Rotation), so dass jede Person maximal vier Jahre in der Stichprobe verbleibt.

Zur Zusammenführung der einzelnen Querschnittserhebungen wurden die Merkmale Bundesland, Regierungsbezirk, Auswahlbezirk und Haushalt sowie zusätzlich Geburtsjahr und Geschlecht als Identifikatoren genutzt. Nach weiteren manuellen Korrekturen⁵ konnten auf Personenebene insgesamt ca. 96 Prozent und auf Haushaltsebene ca. 90 Prozent der Datensätze für das Original-Panel verknüpft werden. Personen in Gemeinschaftsunterkünften wie Alten- und Studentenwohnheime konnten nicht zusammengeführt werden, da die Zusatzmerkmale Geburtsjahr und Geschlecht in diesen Einrichtungen aufgrund der vorliegenden starken Klumpungen zur Identifizierung eher ungeeignet sind und umfangreiche Nachbearbeitungen erforderlich wären (vgl. Herter-Eschweiler 2006a). Es wurden daher Personen aus Gemeinschaftsunterkünften im MZ-Panel nicht mit berücksichtigt.⁶

Für das Mikrozensus-Panel wurde das zweite Rotationsviertel der vierten Vorratsstichprobe verwendet um die Erhebungsjahre 1996 bis 1999 für Längsschnittanalysen zu nutzen.

Abbildung 1: Rotation im Mikrozensus (vgl. Schimpl-Neimanns 2006, Herter-Eschweiler 2006)

Stichprobennummer	Rotationsviertel	Erhebungsjahr									
		1994	1995	1996	1997	1998	1999	2000	2001	2002	
3	3	■	■	■							
3	4	■	■	■	■						
4	1		■	■	■	■					
4	2			1. Befr.	2. Befr.	3. Befr.	4. Befr.				
4	3				■	■	■	■			
4	4					■	■	■	■		
5	1						■	■	■	■	

⁵ Eine manuelle Nachbearbeitung bzw. Korrekturen waren insbesondere aus folgenden Gründen notwendig (vgl. Herter-Eschweiler 2006, S.6):

1. der Tod eines Ehepartners/einer Ehepartnerin wurde nicht angegeben,
2. es lagen inkonsistente Angaben zum Geburtsjahr vor,
3. es lagen inkonsistente Angaben zum Geschlecht vor,
4. die Mobilität des Haushalts wurde nicht erkannt,
5. es lag eine inkonsistente Vergabe der Haushaltsnummer im Beobachtungszeitraum vor.

⁶ Im Mikrozensus Panel 2001 – 2004 sind allerdings die Gemeinschaftsunterkünfte mit enthalten.

Tabelle 1: Auswahlplan der Substichprobenziehung zur Generierung des Campus-Files zum Mikrozensus-Panel 1996 – 1999

Stichprobeneigenschaften	Campus-File des Mikrozensus Panels 1996-1999
Auswahlgesamtheit	Personen in Privathaushalten des zweiten Rotationsviertels der vierten Stichprobe
Auswahleinheiten	Auswahlbezirke
Auswahltechnik	<p>Sortierung der Auswahlbezirke nach:</p> <ol style="list-style-type: none"> 1. Bundesland 2. Größenklasse (basierend auf Dezilen) der Auswahlbezirke 3. Kennung über Grundauswahl/Aktualisierung der Auswahlbezirke 4. Kennung über den Einbezug eines Auswahlbezirks in der Unterstichprobe 5. Regierungsbezirk 6. Regionale Anpassungsschicht (regionale Einheiten mit durchschnittlich 500.000 Einw., die durch Zusammenlegungen von Kreisen und kreisfreien Städten bzw. einzelnen kreisfreien Städten gebildet wurden) 7. Regionale Schicht (regionale Einheiten mit durchschnittlich 250.000 Einw., die durch Zusammenlegungen von Kreisen und kreisfreien Städten gebildet wurden. Großstädte bilden eine eigene oder mehrerer regionale Schichten) 8. Regionale Untergruppe 9. Gemeindegrößenklasse 10. Nummer des Auswahlbezirks <p>Die Sortierung folgt damit weitgehend der Schichtung des Originalmaterials. Analog zum Scientific Use-Files des MZ-Panels wurde die Größenklasse der Auswahlbezirke neu aufgenommen. Die Messung der Größenklasse 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 den Privathaushalten vorliegen.</p> <p>Schichtung:</p> <ul style="list-style-type: none"> ▪ Jeweils zehn in der Reihenfolge der Sortierung aufeinander folgende Auswahlbezirke bilden eine Schicht ▪ 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 Auswahlbezirken zugelassen. Die letzte Schicht kann ebenfalls weniger als zehn Auswahlbezirke umfassen.
Stichprobenumfang	Es wird ein einheitlicher Auswahlsatz von 1/10 in allen Schichten angestrebt. Dazu wird in jeder Schicht mit Hilfe eines „Pseudozufallsgenerators“ eine Zahl a ($1 \leq a \leq 10, a \in \mathbb{N}$) gezogen. Jeder Auswahlbezirk innerhalb einer Schicht mit gezogener Zufallsauswahl gilt als ausgewählt. Bei den „unvollständigen Schichten“ gelten diejenigen Auswahlbezirke als gezogen, deren laufende Nummern innerhalb der Schicht sich unter der einen zufällig aus den Zahlen eins bis zehn gezogenen Zahlen befinden. Der Stichprobenumfang einer solchen Schicht hängt also vom Zufall ab, die Ziehungswahrscheinlichkeit von 1/10 bleibt aber erhalten.
Auswahlsatz	
Hochrechnung	Das Campus File enthält die angepassten MZ-Hochrechnungsfaktoren für Personen und Haushalte/Familien (siehe hierzu Abschnitt 2.3).

2.1. Spezifische Anonymisierungsmaßnahmen

Wie bereits erwähnt, wurde für das Campus-File des Mikrozensus-Panels eine 10 Prozent Substichprobe zum Zwecke der Anonymisierung gezogen. Zusätzlich wurden spezifische Anonymisierungsmaßnahmen ergriffen, die auch beim SUF -File des MZ-Panels eingesetzt worden sind (vgl. Bihler/Herter-Eschweiler/Wirth 2006):

1. Direkte Identifikatoren (wie Personen- und sonstige Ordnungsnummern), soweit sie im Mikrozensus vorhanden sind, wurden gelöscht und durch systemfreie Ordnungsnummern ersetzt.
2. Die Anordnung der Datensätze der anonymisierten Einzelangaben ist durch die Vergabe von systemfreien Ordnungsnummern selbst systemfrei.
3. Die anonymisierten Einzelangaben beschränken sich auf eine zufällig gezogene Substichprobe der Auswahlbezirke von 10 % des Mikrozensus Panels 1996 - 1999.
4. Die Ziehung der Substichprobe erfolgte nach der gleichen Methode wie bei dem Scientific Use File des Mikrozensus Panels 1996 – 1999. Die Beschreibung des Auswahlplans zur Generierung der Substichprobe enthält die Tabelle 1.
5. Als unmittelbare Regionalangabe wird nur das Bundesland weitergegeben.
6. Das Merkmal "Gemeindegrößenklassen" wurde auf fünf Ausprägungen beschränkt, damit keine einzelne Gemeinde mit weniger als 500 000 Einwohnern identifizierbar ist und in jeder Gemeindegrößenklasse mindestens 400 000 Einwohner vertreten sind. Darüber hinaus wurden weitere Zusammenfassungen in einzelnen Ländern vorgenommen.
7. Staatsangehörigkeiten oder Gruppen von Nationalitäten, denen weniger als ca. 50 000 Personen angehören, sind zu größeren Gruppen zusammengefasst und daher nicht einzeln identifizierbar.
8. Jede ausgewiesene Merkmalsausprägung umfasst in der univariaten Verteilung der Grundgesamtheit hochgerechnet mindestens 5 000 Fälle. Gegebenenfalls wurde dies durch sachgerechte Zusammenfassung von Merkmalsausprägungen erreicht.
9. Es werden keine Merkmale der Unterstichprobe für das Erhebungsjahr 1999 ausgewiesen.⁷

⁷ Dies galt schon für das faktisch anonymisierte MZ-Panel (SUF): Aufgrund der Konzeption der Unterstichprobe werden wechselnde Teile eines Rotationsviertels in das Unterstichprobenprogramm einbezogen. Sie ist prinzipiell so angelegt, dass eine Entlastung der zu befragenden Person in der ersten und vierten Erhebung erreicht werden soll und sich die Befragung im Wesentlichen auf die zweite und dritte Erhebung konzentriert. Eine Befragung in der vierten Erhebung ist eher die Ausnahme. Andererseits ergeben sich infolge der Genauigkeitsansprüche an die EU-Arbeitskräfteerhebung unterschiedliche Auswahlsätze in den einzelnen Regierungsbezirken. Dies kann dazu führen, dass Regierungsbezirke mit einem hohen Auswahlatz identifizierbar sind. 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.

2.2 Die Stichprobenentwicklung des Campus Files des Mikrozensus-Panels 1996-1999

Die Substichprobenziehung für das Campus-File des Mikrozensus-Panels 1996-1999 umfasst 10 Prozent der Auswahleinheiten, in denen im Erhebungszeitraum 1996 bis 1999 mindestens eine Person aus einem Privathaushalt befragt wurde. Durch die bei der Substichprobenziehung kontrollierte Größe der Auswahleinheiten – gemessen an der durchschnittlichen Wohnbevölkerung in einer Auswahleinheit – müssten damit annähernd auch 10 Prozent einer jeden, durch die Variablen „perkl“ bzw. „hherkl“ definierten Teileinheit in der Substichprobe enthalten sein (siehe hierzu Herter-Eschweiler, Robert (2002), S. 49 und 52).⁸

Für die Personen- als auch Haushaltsebene wird der angestrebte Auswahlatz von 10 Prozent weitgehend eingehalten (für die Personenebene siehe Tabelle 2). Allerdings sind hinsichtlich einzelner Teileinheiten in Bezug auf Verluste und Gewinne von Personen/Haushalten leichte Verschiebungen zu beobachten. Verluste entstehen vor allen Dingen durch den Fortzug von Haushalten. Insgesamt lassen sich im Campus-File 4.195 Haushalte eindeutig identifizieren, die im Verlauf des Erhebungszeitraums von 1996 bis 1999 aus einer Auswahleinheit fortziehen. Die Auswahlätze liegen für diese Teileinheiten durchgängig mit 9,0 Prozent (1997), 9,4 Prozent (1998) und 9,8 Prozent (1999) etwas unterhalb des erwarteten Auswahlatzes von 10 Prozent. Insgesamt liegt der Auswahlatz von fortgezogenen Haushalten bei 9,4 Prozent und von fortgezogenen Personen bei 10,5 Prozent. Verstorbene Personen weisen über den gesamten Erhebungszeitraum einen Auswahlatz von 9,3 Prozent auf. Der Auswahlatz von immobilen Personen liegt mit 9,9 Prozent sehr nah an dem erwarteten Wert.

Gewinne entstehen vor allem durch zugezogene Haushalte (von 1996 – 1999: 4.616 Haushalte). Der Auswahlatz liegt hier bei 10,1 Prozent. Bei zugezogenen Personen (von 1996 – 1999: 566 Personen) liegt er bei 9,2 Prozent und bei Geburten (von 1996 – 1999: 333 Geburten) bei 9,4 Prozent.

Personen, die einen Ausfall in mindestens einem der Erhebungsjahre aufweisen bzw. für die eine Lücke zwischen den Erhebungsjahren besteht, sind mit einem Auswahlatz von 11,2 Prozent leicht überrepräsentiert. Zu beachten ist hier, dass diese temporären Ausfälle nicht einzelne Personen im Haushalt, sondern aufgrund der bestehenden Auskunftspflicht immer alle Personen im Haushalt betreffen. Der gegenüber dem erwarteten Wert erhöhte Auswahlatz für temporäre Ausfälle lässt sich auch beim faktisch anonymisierten Mikrozensus-Panel nachweisen (siehe Bihler et al. 2006, S.67).

⁸ Die Variablen „perkl“ und „hherkl“ kennzeichnen den Zusammenführungserfolg auf der Personen- bzw. Haushaltsebene. Aus „perkl“ lässt sich erkennen, für welche Erhebungszeitpunkte Personensätze plausibel – d.h. unter Berücksichtigung räumlicher Mobilität, demografischen Ereignissen und temporären Ausfällen – zusammengeführt werden konnten. Die Variable „hherkl“ kennzeichnet einen Haushalt als plausibel, wenn die Anzahl der Personensätze im Erhebungsjahr t_i in einem Haushalt gleich der Anzahl der plausiblen Personensätze ist. Die Variablen „perkl“ und „hherkl“ stimmen nicht überein, wenn bei schon einbezogenen Haushalten Personen aus- oder zuziehen, sowie Personen geboren werden oder versterben.

Tabelle 2: Stichprobenentwicklung des Campus-Files des MZ-Panels 1996 – 1999 auf der Personenebene

Personen in Privathaushalten	Anzahl der Personen im MZ-Panel Originalfile (MZPO) und im faktisch anonymisierten Campus-Files des MZ-Panels (CF-MZP) sowie der Auswahlsatz im Erhebungsjahr ...											
	1996			1997			1998			1999		
	Anzahl MZPO	Anzahl CF-MZP	Auswahlsatz	Anzahl MZPO	Anzahl CF-MZP	Auswahlsatz	Anzahl MZPO	Anzahl CF-MZP	Auswahlsatz	Anzahl MZPO	Anzahl CF-MZP	Auswahlsatz
Insgesamt	179.260	17.565	9,8	180.488	17.896	9,9	180.248	17.962	10,0	179.278	17.819	9,9
davon mit												
unplausiblen Angaben	6.670	642	9,6	9.836	997	10,1	8.984	948	10,6	7.896	820	10,4
plausiblen Angaben	172.590	16.923	9,8	170.652	16.899	9,9	171.264	17.014	9,9	171.382	16.999	9,9
davon mit												
Angaben zum Erhebungsjahr												
1996 1997 1998 1999												
X - - -	20.045	1.860	9,3	-	-	-	-	-	-	-	-	-
X X - -	14.834	1.364	9,2	14.834	1.364	9,2	-	-	-	-	-	-
X X X -	11.879	1.199	10,1	11.879	1.199	10,1	11.879	1.199	10,1	-	-	-
X X X X	124.561	12.362	9,9	124.561	12.362	9,9	124.561	12.362	9,9	124.561	12.362	9,9
- X - -	-	-	-	3.756	380	10,1	-	-	-	-	-	-
- X X -	-	-	-	2.423	254	10,5	2.423	254	10,5	-	-	-
- X X X	-	-	-	11.988	1.199	10,0	11.988	1.199	10,0	11.988	1.199	10,0
- - X -	-	-	-	-	-	-	3.762	366	9,7	-	-	-
- - X X	-	-	-	-	-	-	15.065	1.451	9,6	15.065	1.451	9,6
- - - X	-	-	-	-	-	-	-	-	-	18.879	1.895	10,0
Summe	171.319	16.785	9,8	169.441	16.758	9,9	169.678	16.831	9,9	170.493	16.907	9,9
einem Ausfall in mindestens einem der anderen Erhebungsjahre ¹⁾	1.271	138	10,9	1.211	141	11,6	1.586	183	11,5	889	92	10,3

¹⁾ Es ist zu beachten, dass es auf der Personenebene keine temporären Ausfälle geben kann, da für alle Mitglieder eines Haushalts eine Auskunftspflicht besteht. Temporäre Ausfälle betreffen also stets alle Mitglieder eines Haushalts, weil der Haushalt im entsprechenden Erhebungsjahr nicht erreicht werden konnte.

2.3 Die Bildung angepasster Hochrechnungsfaktoren für die Querschnittsanalysen im Campus-File des Mikrozensus-Panels 1996-1999

Für Querschnittsauswertungen stehen in den einzelnen Erhebungsjahren für Analysen auf der Basis des Einbezugs aller Personen und auf Basis der Unterstichprobe sowie für die Personen- und Haushaltsebene neben den Original-Hochrechnungsfaktoren des Mikrozensus auch jeweils zwei für das Campus-File des MZ-Panels 1996-1999 angepasste Hochrechnungsfaktoren zur Verfügung:

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

Die Grundlage für die angepassten Hochrechnungsfaktoren bilden die Original-Hochrechnungsfaktoren des Mikrozensus, deren Berechnung hier nachfolgend kurz skizziert wird.

Die Bildung der Original-Hochrechnungsfaktoren im Mikrozensus ab 1990 bis 2004

Die Original-Hochrechnungsfaktoren des Mikrozensus werden über ein zweistufiges Verfahren gebildet (vgl. Statistisches Bundesamt 2006, S. 24):

1. Zunächst erfolgt eine Kompensation der bekannten Ausfälle auf der Haushaltsebene innerhalb der regionalen Untergruppen für jeweils 19 Kompensationsklassen, die sich aus der Kombination folgender Merkmale ergeben:
 - (a) Haushaltgröße (differenziert nach 1, 2, 3 oder mehr Personen)
 - (b) Staatsangehörigkeit der Haushaltsbezugsperson (Deutsche/r, Nichtdeutsche/r)
 - (c) Für Deutsche: Wohnrechtsverhältnis der Haushaltsbezugsperson (Haupt-/Nebenwohnsitz)
 - (d) Für 1-Personenhaushalte: Geschlecht und Altersgruppenzugehörigkeit (unter 60 Jahre, 60 Jahre und älter)
2. Die Stichprobenergebnisse werden dann an Eckzahlen der laufenden Bevölkerungsfortschreibung auf der Ebene der regionalen Anpassungsschichten für Deutsche und Nichtdeutsche in der Gliederung nach Geschlecht angepasst.⁹

⁹ Für Soldaten/Soldatinnen und Grundwehr- sowie Zivildienstleistende erfolgt eine gesonderte Anpassung auf der Ebene der Regierungsbezirke bzw. Bundesländer auf der Basis der Bestandsmeldungen des Verteidigungs- bzw. Innenministeriums.

Der endgültige Personenfaktor ergibt sich dann aus der Multiplikation des haushaltsbezogenen Kompensationsfaktors und des personenbezogenen Anpassungsfaktors. Der Personenfaktor variiert zwischen den Personen eines Haushalts. Bei haushaltsbezogenen Auswertungen sind daher die Ergebnisse abhängig von der Person, die als Haushaltsstellvertreter/in eingesetzt wurde. Um Inkohärenzen zu vermeiden, wurde für haushaltsbezogene Auswertungen ein separater Haushaltsfaktor aus dem ungewichteten arithmetischen Mittel der zum Haushalt gehörenden Personen gebildet.

2.3.1 Angepasste Hochrechnungsfaktoren für deskriptive Auswertungen

Bei angepassten Hochrechnungsfaktoren des Campus-Files des MZ-Panels wurden die jeweiligen Original-Hochrechnungsfaktoren 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 ($t = 1996, 1997, 1998, 1999$) (siehe hierzu Bihler/Herter-Eschweiler/Wirth 2006, S. 70-72).

Analog der Konstruktion des Scientific Use Files wurde zur Bestimmung der Auswahlwahrscheinlichkeiten das Bundesland, das Geschlecht, die Staatsangehörigkeit, das Alter sowie der Erwerbsstatus einer Person mit berücksichtigt. Insgesamt wurde zur Abgrenzung der Subpopulationen folgende Kreuzkombinationen verwendet:

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

Für einen angepassten Hochrechnungsfaktor $HF_{i,j,t}^{Anp}$ einer Person i in der j -ten Subpopulation des t -ten Erhebungsjahrs ergibt sich:

$$HF_{i,j,t}^{Anp} = HF_{i,j,t} \cdot \frac{n_{j,t}^{MZQ_g}}{n_{j,t}^{CF-MZP_g}} \quad (1)$$

mit:

- $HF_{i,j,t}^{Anp}$ Angepasster Hochrechnungsfaktor der i -ten Person in der j -ten Subpopulation des t -ten Erhebungsjahrs
- $HF_{i,j,t}$ Hochrechnungsfaktor der i -ten Person in der j -ten Subpopulation aus dem Mikrozensus-Originalfile des t -ten Erhebungsjahrs
- $n_{j,t}^{MZQ_g}$ Anzahl der hochgerechneten Personen im Mikrozensus-Originalfile in der j -ten Subpopulation des t -ten Erhebungsjahrs
- $n_{j,t}^{CF-MZP_g}$ Anzahl der hochgerechneten Personen in der anonymisierten Substichprobe des Campusfiles des MZ-Panel-Querschnitts in der j -ten Subpopulation des t -ten Erhebungsjahrs, wobei als Hochrechnungsfaktor der Hochrechnungsfaktor aus dem Mikrozensus-Originalfile übernommen wurde.

Der Quotient $n_{j,t}^{MZQ_s} / n_{j,t}^{CF-MZP_s}$ wird als „Korrekturfaktor“ bezeichnet. Da das Campusfile nur eine kleine Substichprobe darstellt, kann es in einigen Fällen zu leeren Feldern kommen. In diesen Fällen wurde, wie beim Scientific-Use-File auch, als Korrekturfaktor der Kehrwert der erwarteten Auswahlwahrscheinlichkeit, dass eine Person im k-ten Bundesland aus dem Mikrozensus-Originalfile in die Substichprobe des Campusfiles aufgenommen wurde, eingesetzt.

Auswahlwahrscheinlichkeiten

In dem Original-Mikrozensus ist die Stichprobenerhebung über Arbeitskräfte in der Europäischen Gemeinschaft mit einem reduzierten Auswahlatz integriert (Unterstichprobe). Im Bundesdurchschnitt liegt der Auswahlatz für die Erhebungsjahre 1996 bis 2004 im Original-Mikrozensus bei 0,45 Prozent. Aufgrund der Genauigkeitsanforderungen für die Stichprobe (siehe Statistisches Bundesamt 2006, S.28) kann der Auswahlatz auf der Ebene der Regierungsbezirke 0,4, 0,6, 0,8 oder auch 1 betragen. Die Unterstichprobe ist zudem so angelegt, dass eine Entlastung der zu befragenden Person in der ersten und vierten Befragung erreicht werden soll. So sind im ersten Erhebungsjahr eines neu in die Auswahl gelangten Rotationsviertels ca. 45 Prozent der zu befragenden Personen auch in das Unterstichprobenprogramm involviert. In den beiden darauf folgenden Jahren erhöht sich der Anteil auf ca. 84 Prozent. Die Merkmale der Unterstichprobe sind für das vierte Erhebungsjahr aufgrund der notwendigen Anonymisierungsmaßnahmen (siehe Abschnitt 2.1) nicht mit einbezogen worden.

Das Campus-File des MZ-Panels 1996-1999 ist eine 10-prozentige Substichprobe aus einem Rotationsviertel: Somit ergibt sich eine Auswahlwahrscheinlichkeit, dass eine Person aus einem Mikrozensus-Originalfile in den Campusfile gelangt, von $0,25 * 0,1 = 0,025$. Für die Berechnung der Auswahlwahrscheinlichkeit, dass eine Person auch in die Unterstichprobe einbezogen ist, muss der Auswahlatz der Unterstichprobe mit berücksichtigt werden.

Dieser fällt allerdings wie oben erwähnt für die einzelnen Bundesländer, Regierungsbezirke und in den einzelnen Erhebungsjahren in einem Rotationsviertel unterschiedlich aus (siehe hierzu Herter-Eschweiler 2006a, S. 29). Für den Bundesdurchschnitt sind die Auswahlwahrscheinlichkeiten in Tabelle 3 dargestellt.

Tabelle 3: Auswahlwahrscheinlichkeit für eine Person in der Unterstichprobe

Erhebungsjahr	Rotationsviertel	Unterstichprobe	Campus-File	Auswahlwahrscheinlichkeit			
1996	0,25	*	0,449	*	0,1	=	0,0112
1997	0,25	*	0,839	*	0,1	=	0,0210
1998	0,25	*	0,845	*	0,1	=	0,0211

2.3.2 Angepasste Hochrechnungsfaktoren für inferenzstatistische Auswertungen

Analog zu dem faktisch anonymisierten Mikrozensus-Panel 1996-1999 werden auch für den Campusfile Hochrechnungsfaktoren für inferenzstatistische Auswertungen bereitgestellt. Sie geben zwar die Struktur der gewichteten Verteilungen wieder, ihre

Fallzahlen werden allerdings auf die Stichprobengröße eines jeden Erhebungsjahres eines Mikrozensus-Panel-Querschnitts normiert. Dies ist wichtig, da im Rahmen von Teststatistiken z.B. mit SPSS generell die gewichteten Fallzahlen herangezogen werden. Bei der Normierung wird jeder Gewichtungsfaktor eines i -ten Datensatzes mit dem Verhältnis von ungewichteter (n_t^{CF-MZP}) und gewichteter Fallzahl ($n_t^{CF-MZP_g}$) multipliziert.

$$HF_{i,t,Stp} = HF_{i,t} \cdot k_{i,j,t} \cdot \frac{n_t^{CF-MZP}}{n_t^{CF-MZP_g}} \tag{2}$$

mit:

- $HF_{i,t,Stp}$ Auf den Stichprobenumfang des Campus-Files des MZ-Panel-Querschnitts normierter Hochrechnungsfaktor im t -ten Erhebungsjahr für die Person i
- $HF_{i,t}$ Aus dem Mikrozensus-Originalfile übernommener Hochrechnungsfaktor des t -ten Erhebungsjahrs für die Person i
- $k_{i,j,t}$ Korrekturfaktor zur Bildung angepasster Hochrechnungsfaktoren an das jeweilige Mikrozensus-Originalfile für die Person i
- n_t^{CF-MZP} Anzahl der befragten Personen in der Campus Substichprobe des Mikrozensus-Panel-Querschnitts im t -ten Erhebungsjahr
- $n_t^{CF-MZP_g}$ Anzahl der hochgerechneten Personen in der Campus Substichprobe des Mikrozensus-Panel-Querschnitts im t -ten Erhebungsjahr

Die Variablennamen und Labels aller zur Verfügung stehenden angepassten Hochrechnungsfaktoren lassen sich der folgenden Tabelle entnehmen:

Tabelle 4: Variablennamen und –labels der angepassten Hochrechnungsfaktoren in den einzelnen Erhebungsjahren des Campus-Files 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 –

Angepasster Hochrechnungsfaktor ...

	af755b	Unterstichprobe (korrigiert für deskriptive Statistiken) (EF755) MZ96: Personen-Hochrechnungsfaktor – Unterstichprobe (korrigiert für Inferenzstatistiken)
	af756d	(EF756) MZ96:HH- und Familien-Hochrechnungsfaktor – Unterstichprobe (korrigiert für deskriptive Statistiken)
	af756b	(EF756) MZ96: HH- und Familien-Hochrechnungsfaktor – Unterstichprobe (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 – Unterstichprobe (korrigiert für deskriptive Statistiken)
	bf755b	(EF755) MZ97: Personen-Hochrechnungsfaktor – Unterstichprobe (korrigiert für Inferenzstatistiken)
	bf756d	(EF756) MZ97:HH- und Familien-Hochrechnungsfaktor – Unterstichprobe (korrigiert für deskriptive Statistiken)
	bf756b	(EF756) MZ97: HH- und Familien-Hochrechnungsfaktor – Unterstichprobe (korrigiert für Inferenzstatistiken)
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 – Unterstichprobe (korrigiert für deskriptive Statistiken)
	cf755b	(EF755) MZ98: Personen-Hochrechnungsfaktor – Unterstichprobe (korrigiert für Inferenzstatistiken)
	cf756d	(EF756) MZ98:HH- und Familien-Hochrechnungsfaktor – Unterstichprobe (korrigiert für deskriptive Statistiken)
	cf756b	(EF756) MZ98: HH- und Familien-Hochrechnungsfaktor – Unterstichprobe (korrigiert für Inferenzstatistiken)

Angepasster Hochrechnungsfaktor ...

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)

2.4 Die im Campus File des Mikrozensus-Panels 1996 – 1999 bereitgestellten Längsschnitt- und Mobilitätsgewichte

Analog dem faktisch anonymisierten Mikrozensus-Panel (SUF), stehen im Campus-File des Mikrozensus-Panels 1996-1999 zwei unterschiedliche Längsschnittgewichte und ein Mobilitätsgewicht für jeweils unterschiedliche Zeiträume zur Verfügung:

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

Bei den Mobilitätsgewichten handelt es sich um den Kehrwert der mit Logit-Modellen geschätzten Wahrscheinlichkeiten räumlicher Immobilität. Die Berechnung der Gewichte erfolgte mit der gleichen Prozedur wie schon beim SUF. Nähere Erläuterungen zur Konstruktion und Anwendung können daher dem Handbuch des Mikrozensus-Panels 1996 -1999, Kapitel 4 und 5 entnommen werden. Die Längsschnittgewichte zur Hochrechnung ohne und mit Berücksichtigung der Mobilitätsgewichte sind – analog dem SUF - nur mit der Ausgangsbasis des Erhebungsjahres 1996 verfügbar.

Die Bezeichnungen der Gewichtungsvariablen setzen sich aus zwei Kennungen zusammen: der Gewichtungsart und der Übergangsdauer. Die nachfolgende Tabelle gibt einen Überblick über die zur Verfügung stehenden Längsschnitt- und Mobilitätsgewichte. Zu beachten ist, dass analog dem faktisch anonymisierten MZ-Panel (SUF) die Längsschnitthochrechnungsfaktoren auf die Fallzahlen der Zielpopulation hochrechnen, während die angepassten Hochrechnungsfaktoren für Querschnittsberechnungen auf die Fallzahlen des Original-Mikrozensus Querschnitt hochrechnen.

Wie schon im Rahmen von Querschnittsauswertungen unter Abschnitt 2.3.2 erläutert wurde, müssen bei Teststatistiken z.B. mit SPSS die Längsschnittgewichte auf die Anzahl der befragten Personen normiert werden (siehe hierzu Basic 2006, S. 119). Bei der Normierung wird jeder Gewichtungsfaktor eines i -ten Datensatzes (z.B. HGM_{9699_i}) mit dem Verhältnis von ungewichteter (n) und gewichteter Fallzahl ($n_{HGM_{9699}}$) multipliziert. Als Ergebnis erhält man z.B. einen normierten

Längsschnittgewichtungsfaktor $NHGM_{9699_i}$, bei dem die Struktur der Gewichtung beibehalten und die Fallzahl auf die Anzahl der Befragten korrigiert wurde.

$$NHGM_{9699_i} = HGM_{9699_i} * \frac{n}{n_{HGM_{9699}}} \quad (3)$$

Tabelle 5: Variablenamen der einzelnen Längsschnitt- und Mobilitätsgewichte im Campus-File des Mikrozensus-Panels 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

2.5 Randverteilungen ausgewählter Merkmale in den Mikrozensus-Originalfiles für ein Erhebungsjahr und den Querschnitten des Campus File des Mikrozensus-Panels 1996-1999

Für ausgewählte Merkmale werden die Randverteilungen für die Bevölkerung in Privathaushalten für das Erhebungsjahr 1996 (unter Berücksichtigung der jeweiligen Hochrechnung) dargestellt. Hierbei werden die Ergebnisse des Mikrozensus-Originalfiles für ein Erhebungsjahr sowie des Campus-Files des Mikrozensus-Panels gegenübergestellt. Die Randverteilungen für die Jahre 1997 – 1999 lassen sich den Tabellen im Anhang der Dokumentation entnehmen. Die ausgewiesenen Merkmale des Vergleichs zwischen Mikrozensus-Originalfile und Campus File des Mikrozensus Panel-Querschnitts sind:

- 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.

Tabelle 6: Bevölkerung in Privathaushalten nach ausgewählten Merkmalen im MZ-Originalfile und dem Querschnitt des Campus-Files im Erhebungsjahr 1996 mit Personengewichtung

Bevölkerung in Privathaushalten im Erhebungsjahr 1996 mit Personengewichtung						
Gegenstand der Nachweisung	MZ-Originalfile		Querschnitt des Campus-Files		Differenz	
	1000	Prozent	1000	Prozent	absolut	relativ
Insgesamt	82.069	100,0	82.069	100,0	0	0,00
davon						
im Bundesland						
Schleswig-Holstein	2.770	3,4	2.770	3,4	-1	-0,04
Hamburg	1.710	2,1	1.668	2,0	-43	-2,50
Niedersachsen	7.779	9,5	7.748	9,4	-32	-0,41
Bremen	678	0,8	629	0,8	-49	-7,23
Nordrhein-Westfalen	17.842	21,7	17.921	21,8	79	0,44
Hessen	6.055	7,4	6.061	7,4	6	0,09
Rheinland-Pfalz	4.012	4,9	3.973	4,8	-39	-0,98
Baden-Württemberg	10.413	12,7	10.444	12,7	31	0,30
Bayern	12.137	14,8	12.205	14,9	68	0,56
Saarland	1.076	1,3	1.040	1,3	-35	-3,28
Berlin	3.511	4,3	3.493	4,3	-18	-0,51
Brandenburg	2.533	3,1	2.533	3,1	-1	-0,02
Mecklenburg-Vorpommern	1.813	2,2	1.814	2,2	1	0,05
Sachsen	4.541	5,5	4.562	5,6	21	0,47
Sachsen-Anhalt	2.709	3,3	2.721	3,3	12	0,44
Thüringen	2.490	3,0	2.490	3,0	0	-0,01
nach dem Geschlecht						
männlich	40.133	48,9	40.020	48,8	-113	-0,28
Weiblich	41.937	51,1	42.049	51,2	113	0,27
nach dem Familienstand						
Ledig	31.365	38,2	31.723	38,7	358	1,14
verheiratet	40.745	49,6	40.129	48,9	-617	-1,51
verwitwet	6.278	7,6	6.486	7,9	208	3,31
geschieden	3.681	4,5	3.732	4,5	51	1,38
nach Altersgruppen						
0 bis unter 5 Jahre	3.843	4,7	3.780	4,6	-62	-1,61
5 bis unter 10 Jahre	4.620	5,6	4.573	5,6	-48	-1,03
10 bis unter 15 Jahre	4.537	5,5	4.605	5,6	67	1,48
15 bis unter 20 Jahre	4.479	5,5	4.806	5,9	326	7,29
20 bis unter 25 Jahre	4.739	5,8	4.630	5,6	-109	-2,30
25 bis unter 30 Jahre	6.306	7,7	6.427	7,8	121	1,92
30 bis unter 35 Jahre	6.918	8,4	6.617	8,1	-301	-4,35
35 bis unter 40 Jahre	6.305	7,7	6.288	7,7	-17	-0,27
40 bis unter 45 Jahre	5.803	7,1	5.681	6,9	-123	-2,11
45 bis unter 50 Jahre	5.368	6,5	5.265	6,4	-103	-1,92
50 bis unter 55 Jahre	5.127	6,2	5.389	6,6	262	5,12
55 bis unter 60 Jahre	6.308	7,7	6.256	7,6	-52	-0,82
60 bis unter 65 Jahre	4.740	5,8	4.466	5,4	-273	-5,77
65 bis unter 70 Jahre	4.309	5,3	4.365	5,3	57	1,31
70 bis unter 75 Jahre	3.583	4,4	3.762	4,6	179	4,99
75 bis unter 80 Jahre	2.178	2,7	2.347	2,9	169	7,75
80 Jahre und älter	2.905	3,5	2.811	3,4	-94	-3,23

Noch Tabelle 6: Bevölkerung in Privathaushalten nach ausgewählten Merkmalen im MZ-Originalfile und dem Querschnitt des Campus-Files im Erhebungsjahr 1996 mit Personengewichtung

Bevölkerung in Privathaushalten im Erhebungsjahr 1996 mit Personengewichtung						
Gegenstand der Nachweisung	MZ-Originalfile		Querschnitt des Campus-Files		Differenz	
	1000	Prozent	1000	Prozent	absolut	relativ
Insgesamt	82.069	100,0	82.069	100,0	0	0,00
Davon						
nach der Staatsangehörigkeit						
Deutsche(r)	74.914	91,3	75.217	91,7	303	0,40
Ausländer(in)						
aus einem EU-Staat (EU-15)	1.828	2,2	1.755	2,1	-73	-4,00
Ausländer(in)						
aus einem Nicht-EU-Staat	5.327	6,5	5.098	6,2	-230	-4,31
nach Art der Erwerbsbeteiligung						
erwerbstätig	36.459	44,4	36.237	44,2	-223	-0,61
erwerbslos nach ILO	3.476	4,2	3.378	4,1	-98	-2,82
sonstig erwerbslos	533	0,6	500	0,6	-33	-6,14
Nichterwerbsperson	41.601	50,7	41.955	51,1	354	0,85
nach der Stellung im Beruf in der gegenwärtigen Tätigkeit						
Selbstständig ohne Beschäftigte	1.685	2,1	1.629	2,0	-56	-3,31
Selbstständig mit Beschäftigten	1.818	2,2	1.777	2,2	-41	-2,24
Mithelfende(r)						
Familienangehörige(r)	391	0,5	320	0,4	-71	-18,16
Beamtin/Beamter, Richter(in)	2.110	2,6	2.183	2,7	73	3,47
Angestellte(r)	16.260	19,8	15.771	19,2	-490	-3,01
(Heim-)Arbeiter(in)	12.211	14,9	12.385	15,1	174	1,43
Kaufm./techn. Auszubildende(r)	707	0,9	769	0,9	62	8,70
Gewerbl. Auszubildende(r)	767	0,9	894	1,1	127	16,57
Zeit-/ Berufssoldat(in) (einschl. BGS und Bereitschaftspolizei)	250	0,3	247	0,3	-3	-1,06
Grundwehrdienstleistender	166	0,2	188	0,2	22	13,12
Zivildienstleistender	96	0,1	75	0,1	-21	-21,65
trifft nicht zu (Person unter 15 Jahre, nicht erwerbstätig)	45.610	55,6	45.833	55,8	223	0,49
nach Voll-/Teilzeittätigkeit						
Vollzeittätigkeit	30.417	37,1	30.072	36,6	-345	-1,13
Teilzeittätigkeit	6.042	7,4	6.164	7,5	122	2,02
trifft nicht zu (Person unter 15 Jahre, nicht erwerbstätig)	45.610	55,6	45.833	55,8	223	0,49

Noch Tabelle 6: Bevölkerung in Privathaushalten nach ausgewählten Merkmalen im MZ-Originalfile und dem Querschnitt des Campus-Files im Erhebungsjahr 1996 mit Personengewichtung

Bevölkerung in Privathaushalten im Erhebungsjahr 1996 mit Personengewichtung						
Gegenstand der Nachweisung	MZ-Originalfile		Querschnitt des Campus-Files		Differenz	
	1000	Prozent	1000	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	34.116	41,6	-185	-0,54
Arbeitslosengeld bzw. Arbeitslosenhilfe	2.878	3,5	2.776	3,4	-102	-3,54
Rente, Pension	16.408	20,0	16.787	20,5	379	2,31
Unterhalt durch Eltern, Ehemann/-frau oder andere Angehörige	25.243	30,8	25.266	30,8	23	0,09
eigenes Vermögen, Vermietung, Zinsen, Altenteil	383	0,5	376	0,5	-6	-1,63
Sozialhilfe	1.535	1,9	1.575	1,9	41	2,65
Leistungen aus einer Pflegeversicherung	86	0,1	83	0,1	-3	-3,40
Sonstige Unterstützungen	1.236	1,5	1.089	1,3	-147	-11,89
nach den Individualeinkommensklassen						
unter 300 DM	2.425	3,0	2.670	3,3	245	10,12
300 bis unter 600 DM	4.575	5,6	4.425	5,4	-150	-3,28
600 bis unter 1000 DM	6.123	7,5	5.904	7,2	-219	-3,58
1000 bis unter 1400 DM	6.984	8,5	7.137	8,7	153	2,19
1400 bis unter 1800 DM	6.925	8,4	7.232	8,8	307	4,44
1800 bis unter 2200 DM	8.020	9,8	8.041	9,8	21	0,27
2200 bis unter 2500 DM	6.012	7,3	6.042	7,4	29	0,49
2500 bis unter 3000 DM	6.067	7,4	5.821	7,1	-247	-4,06
3000 bis unter 3500 DM	3.939	4,8	4.031	4,9	93	2,35
3500 bis unter 4000 DM	2.276	2,8	2.108	2,6	-168	-7,40
4000 bis unter 4500 DM	1.527	1,9	1.443	1,8	-84	-5,49
4500 bis unter 5000 DM	1.070	1,3	1.044	1,3	-26	-2,41
5000 bis unter 5500 DM	742	0,9	771	0,9	28	3,82
5500 bis unter 6000 DM	496	0,6	441	0,5	-55	-11,02
6000 bis unter 6500 DM	372	0,5	406	0,5	34	9,11
6500 bis unter 7000 DM	248	0,3	237	0,3	-11	-4,42
7000 bis unter 7500 DM	182	0,2	230	0,3	48	26,55
7500 und mehr DM	794	1,0	697	0,8	-96	-12,14
selbstständige(r) Landwirt(in) in der Haupttätigkeit	338	0,4	328	0,4	-10	-2,97
kein Einkommen	20.069	24,5	20.106	24,5	37	0,18
keine Angabe	2.886	3,5	2.954	3,6	69	2,38

Noch Tabelle 6: Bevölkerung in Privathaushalten nach ausgewählten Merkmalen im MZ-Originalfile und dem Querschnitt des Campus-Files im Erhebungsjahr 1996 mit Personengewichtung

Bevölkerung in Privathaushalten im Erhebungsjahr 1996 mit Personengewichtung						
Gegenstand der Nachweisung	MZ-Originalfile		Querschnitt des Campus-Files		Differenz	
	1000	Prozent	1000	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	32.598	39,7	-91	-0,28
Abschluss der allgemeinbildenden POS der ehemaligen DDR	5.837	7,1	5.532	6,7	-306	-5,24
Realschul- oder gleichwertiger Abschluss	10.743	13,1	11.037	13,4	294	2,74
Fachhochschulreife	2.288	2,8	2.308	2,8	20	0,88
Allgemeine oder fachgebundene Hochschulreife	9.369	11,4	8.774	10,7	-594	-6,34
keine Angabe	375	0,5	528	0,6	153	40,83
trifft nicht zu (Person unter 15 Jahre, kein allgemeinbildender Abschluss bzw. keine Angabe hierzu)	20.768	25,3	21.292	25,9	524	2,52
nach dem höchsten beruflichen Ausbildungsabschluss						
Anlernausbildung, berufliches Praktikum	1.825	2,2	2.004	2,4	179	9,81
Abschluss einer Lehrausbildung oder gleichwertiger Berufsfachschulabschluss	31.382	38,2	31.195	38,0	-187	-0,60
Meister-/Technikerabschluss oder gleichwertiger Fachschulabschluss	3.659	4,5	3.639	4,4	-20	-0,54
Abschluss der Fachschule in der ehemaligen DDR	1.101	1,3	1.046	1,3	-55	-5,01
Fachhochschulabschluss (auch Ingenieurschulabschluss)	2.506	3,1	2.423	3,0	-83	-3,33
Hochschulabschluss	4.327	5,3	4.042	4,9	-285	-6,59
keine Angabe	524	0,6	592	0,7	68	13,08
trifft nicht zu (Person unter 15 Jahre, kein beruflicher Abschluss bzw. keine Angabe hierzu)	36.746	44,8	37.129	45,2	383	1,04

2.6 Berechnung der Stichprobenfehler und der Konfidenzintervalle für ausgewählte Merkmale

Das Mikrozensus-Panel Originalfile 1996-1999 stellt Stichprobeninformationen bereit, die es ermöglichen für das Campus-File des MZ-Panels näherungsweise die Varianz von Populationsschätzern für ausgewählte Merkmale zu berechnen. Dies ist zumindest für Querschnittsauswertungen anzunehmen. Aufgrund räumlicher Mobilität können selektive Ausfälle auftreten. Die damit verbundenen Schätzprobleme und Annahmen bei Verwendung von Korrekturgewichten werden im Folgenden nicht berücksichtigt.

Die nachfolgende Ermittlung der Stichprobenfehler orientiert sich an dem von Rendtel/Schimpi-Neimanns (2001) entwickelten methodischen Konzept, welches auf Särndal et al (1992) zurückgeht. Die konkrete Berechnung der Varianzen wird mit Hilfe des frei zugänglichen SAS-Programms „varmz_a.sas“, welches für das Campus-MZ-Panel entsprechend angepasst wurde, durchgeführt. Die Ermittlung der Varianz der Populationsschätzer erfolgt hierbei nach der Anpassung an die Bevölkerungsfortschreibung (s. u.).

In den folgenden Darstellungen wird zunächst die Anpassung an die Bevölkerungsfortschreibung nicht berücksichtigt, sondern nur die so genannte freie Hochrechnung bzw. Designgewichtung betrachtet.

Zu beachten ist, dass das Campus-File im Unterschied zu den Scientific Use Files einzelner Erhebungsjahre des Mikrozensus (Querschnittsdaten) keine Haushaltsstichprobe darstellt, sondern eine 10% Substichprobe der Auswahlbezirke des vollen Mikrozensus Panels 1996-1999 ist. Alle Haushalte eines gezogenen Auswahlbezirks sind auch in dem Campus-File enthalten. Das Campus File ist somit im strengen Sinne durch ein zweiphasiges Ziehungsverfahren auf Auswahlbezirksebene entstanden (1. Phase: Auswahl der Auswahlbezirke des vollen Mikrozensus Panels; 2. Phase: 10% Auswahl der Auswahlbezirke des Campus Files).

Das Verfahren ist näherungsweise auch durch ein zweistufiges Auswahlverfahren beschreibbar (siehe Särndal et al. 1992, Kap. 9.3). Die Darstellung der Varianzschätzung lässt sich weiter vereinfachen, denn aufgrund des geringen Auswahlsatzes der Auswahlbezirke des Mikrozensus sind die Varianzanteile nach der ersten Stufe (bzw. der zweiten Phase) vernachlässigbar. Es reicht i. d. R. aus, den Standardfehler für die erste Auswahlstufe zu berechnen.

Für die Schätzung wird unter Annahme einer einstufigen Klumpenauswahl die Ziehungswahrscheinlichkeit eines Auswahlbezirks im Campus File benötigt. Im Folgenden wird vereinfachend davon ausgegangen, dass diese Ziehungswahrscheinlichkeit dem Produkt der Ziehungswahrscheinlichkeiten (a) eines Auswahlbezirks im Original-Mikrozensus-Querschnitt (1 %), (b) einem Rotationsviertel (0,25 %) des Original-Mikrozensus und (c) der Substichprobe Campus File (10 %) entspricht. Das Produkt der ersten beiden Ziehungswahrscheinlichkeiten (a) und (b) ist dabei die Ziehungswahrscheinlichkeit des vollen Mikrozensus Panel, aus dem das Campus File gezogen wurde. Insofern kann auch das Hochrechnungsverfahren des Original Mikrozensus mit

entsprechender Gewichtung mit dem Kehrwert der Ziehungswahrscheinlichkeit für das Campus-File des MZ-Panels verwendet werden.

Damit kann die einstufige Schätzung eines Gesamtaufkommens (Total) t bzw. \hat{Y} eines Merkmals und der Varianz $\hat{V}(\hat{Y})$ in Anlehnung an das Stata Survey Data Reference Manual (2007: 151-152) wie folgt dargestellt werden:

$$\hat{Y} = \sum_{h=1}^L \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} w_{hij} y_{hij} \quad (4)$$

$$\hat{V}(\hat{Y}) = \sum_{h=1}^L (1 - f_h) \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} (y_{hi} - \bar{y}_h)^2 \quad (5)$$

mit $y_{hi} = \sum_{j=1}^{m_{hi}} w_{hij} y_{hij}$ gewichtetes Total der PSU (h, i)

$\bar{y}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} y_{hi}$ Mittelwert der PSU-Totals der Schicht h

f_h Auswahlsatz, Inklusionswahrscheinlichkeit

h Schichten $h = 1, \dots, L$; gebildet durch die Bundesländer und Gebäudegrößenklassen

i Primäreinheiten (PSUs) in Schicht h mit $i = 1, \dots, N_h$

n_h Anzahl der PSUs pro Schicht

Der mit $w_{hij} = N_h / n_h$ gewichtete Merkmalswert y_{hij} bezieht sich somit auf die Person j im Auswahlbezirk (PSU) i in Schicht h .

Die Varianz nach der Anpassung an die Bevölkerungsfortschreibung

Von den statistischen Ämtern werden die vorliegenden Mikrozensusergebnisse anhand von Eckzahlen der laufenden Bevölkerungsfortschreibung hochgerechnet. Diese Anpassung kann Einfluss auf die Varianzschätzung haben und sollte deshalb mit berücksichtigt werden.¹⁰ Hierbei können die verwendeten Gewichte als Regressionschätzer interpretiert werden. Dieser Ansatz zur Schätzung von Populationsparametern basiert auf dem Group Mean Modell und ist ausführlich in Rendtel/Schimpi-Neimanns (2001) dargestellt, hier wird er nur skizziert.

Der benutzte Regressionsschätzer \hat{t}_{reg} lässt sich für das Group Mean Modell wie folgt darstellen (vgl. Rendtel/Schimpi-Neimanns 2001):

$$\hat{t}_{reg} = \sum_{k \in U} \hat{y}_k = \sum_{g=1}^G \sum_{k \in U_g} \hat{B}_g = \sum_{g=1}^G N_g \hat{B}_g = \sum_{g=1}^G \sum_{k \in s_g} \frac{N_g}{\hat{N}_g} \cdot \frac{y_k}{\pi_k} = \sum_{g=1}^G \sum_{k \in s_g} w_g \frac{y_k}{\pi_k} \quad (9)$$

Der Faktor $w_g = \frac{N_g}{\hat{N}_g}$ beschreibt das Verhältnis von dem Umfang von Gruppe g in der Grundgesamtheit (N_g), welches der Soll-Vorgabe entspricht¹¹, zu dem geschätzten Umfang von Gruppe g (\hat{N}_g), bzw. dem Ist-Wert. Der Regressionswert ist dann ein gewichtetes Mittel:

$$\hat{t}_{reg} = \sum_{k \in s_g} w_g \frac{y_k}{\pi_k} \quad (10)$$

Für die Herleitung der Varianz wird eine Taylorentwicklung von \hat{t} verwendet. Der lineare Teil davon ist durch die folgende Hilfsgröße u_k gegeben (Rendtel/Schimpi-Neimanns 2001, Särndal et al. 1992):

$$u_k = \frac{N_g}{\hat{N}_g} (y_k - \hat{B}_g) = w_k (y_k - \bar{y}_{s_g}) \quad k \in s_g \quad (11)$$

Dieses Hilfsmerkmal ist die mit w_k gewichtete Abweichung des Merkmalswertes y_k von dem jeweiligen Gruppenmittelwert \bar{y}_{s_g} in der Stichprobe. Als asymptotische Näherung für die Varianz $V(\hat{t}_{reg})$ kann dann die Varianz des Hilfsmerkmals u_k angesehen werden. In der obigen Gleichung Nr. 5 wird daher der Wert y_k durch u_k ersetzt. Nachfolgend werden für einige ausgewählte Merkmale die Stichprobenfehler der ersten Auswahlstufe (sig_b) sowie die Konfidenzintervalle für das Campus-MZ-Panel berechnet und für 1996 in der Tabelle 7 sowie für die weiteren Jahre im Anhang in den Tabellen 15 bis 17 dargestellt. Die Konfidenzintervalle lassen sich den Spalten P und Q entnehmen. Wenn die hochgerechneten Werte des Mikrozensus Originalfiles außerhalb des Intervalls liegen, wird dies in der Spalte O durch eine „1“ gekennzeichnet. Dies trifft in den Tabellen 15 bis 17 nur in einem Fall zu.

¹⁰ Die Fehlerrechnungen der statistischen Ämter berücksichtigen dies bislang aber nicht, sondern schätzen den Stichprobenfehler lediglich designbasiert bzw. für die so genannte freie Hochrechnung.

¹¹ Bei der Anpassung an die Bevölkerungsfortschreibung wird für N_g jeweils der Wert für bestimmte Gruppen gemäß der Fortschreibung gewählt.

**Tabelle 7: Standardfehlerberechnung für das Campus-File des MZ-Panels 1996-1999
Bevölkerung in Privathaushalten im Erhebungsjahr 1996 (Hochrechnung mit Personengewichtung)**

Gegenstand der Nachweisung	Mikrozensus Originalfile								Campus-MZ-Panel								
	Hochrechnung (EF750)		Fälle		Hochrechnung (AF750D)		Fälle		Differenz zum MZ Originalfile (Hochrechnung)		t_w gewichtetes / angepasstes Total	sig_b Standardfehler Between-Teil	rel relativer Standardfehler (sig_b/t_w)	sig			
									Werte	% (G-C)				MZ_O <> CI (B<P oder B>Q)	Untergrenze (L - 2*M)	Obergrenze (L + 2*M)	
	1000	%	Anzahl	%	1000	%	Anzahl	%			1000	%	0 bzw. 1	1000			
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	
Insgesamt	82.069	100,0	719.854	100,0	82.069	100,0	17.565	100,0	-0,0								
nach dem Familienstand																	
ledig	31.365	38,2	273.453	38,0	31.723	38,7	6.722	38,3	+1,1	+0,4	31.723	362	1,1	0	30.998	32.447	
verheiratet	40.745	49,6	358.236	49,8	40.129	48,9	8.630	49,1	-1,5	-0,8	40.129	372	0,9	0	39.384	40.874	
verwitwet	6.278	7,6	55.743	7,7	6.486	7,9	1.391	7,9	+3,3	+0,3	6.486	187	2,9	0	6.112	6.859	
geschieden	3.681	4,5	32.422	4,5	3.732	4,5	822	4,7	+1,4	+0,1	3.732	148	4,0	0	3.436	4.028	
nach dem überwiegenden Lebensunterhalt																	
Erwerbs-/ Berufstätigkeit	34.301	41,8	300.602	41,8	34.116	41,6	7.416	42,2	-0,5	-0,2	34.116	380	1,1	0	33.357	34.876	
Arbeitslosengeld bzw. .Arbeitslosenhilfe	2.878	3,5	24.871	3,5	2.776	3,4	604	3,4	-3,5	-0,1	2.776	128	4,6	0	2.520	3.032	
Rente, Pension	16.408	20,0	146.429	20,3	16.787	20,5	3.584	20,4	+2,3	+0,5	16.787	367	2,2	0	16.052	17.522	
Unterhalt durch Eltern, Ehemann/-frau oder andere Angehörige	25.243	30,8	221.023	30,7	25.266	30,8	5.337	30,4	+0,1	+0,0	25.266	381	1,5	0	24.503	26.029	
eigenes Vermögen, Vermietung, Zinsen, Altenteil	383	0,5	3.393	0,5	376	0,5	79	0,4	-1,6	-0,0	376	54	14,4	0	268	485	
Sozialhilfe	1.535	1,9	11.943	1,7	1.575	1,9	306	1,7	+2,7	+0,0	1.575	167	10,6	0	1.240	1.910	
Leistungen aus einer Pflegeversicherung	86	0,1	767	0,1	83	0,1	17	0,1	-3,4	-0,0	83	21	25,6	0	41	126	
Sonstige Unterstützungen	1.236	1,5	10.826	1,5	1.089	1,3	222	1,3	-11,9	-0,2	1.089	91	8,3	0	908	1.270	
nach Art der Erwerbsbeteiligung																	
erwerbstätig	36.459	44,4	319.780	44,4	36.237	44,2	7.892	44,9	-0,6	-0,3	36.237	379	1,0	0	35.479	36.994	
erwerbslos nach ILO	3.476	4,2	29.748	4,1	3.378	4,1	734	4,2	-2,8	-0,1	3.378	146	4,3	0	3.086	3.670	
sonstig erwerbslos	533	0,6	4.616	0,6	500	0,6	103	0,6	-6,1	-0,0	500	57	11,3	0	387	613	
Nichterwerbsperson	41.601	50,7	365.708	50,8	41.955	51,1	8.836	50,3	+0,9	+0,4	41.955	374	0,9	0	41.206	42.703	

3. Beispiele für multivariate Längsschnittanalysen mit dem Campus File des MZ-Panels 1996 - 1999

Die nachstehenden beispielhaften Übergangsmodelle sollen nur die prinzipiellen Analysemöglichkeiten des Campus-Files aufzeigen. Es besteht kein Anspruch auf eine vollständige Spezifikation der Modelle bzw. die korrekte Behandlung des Ausfallgeschehens.

3.1 *Übergang aus abhängiger Beschäftigung in Erwerbslosigkeit*

Als Beispiel soll in der folgenden Analyse mit Hilfe des Campus-Files der Übergang aus dem Erwerbszustand „abhängig beschäftigt“ in „erwerbslos“ modelliert werden. Es werden als Übergänge drei Zeiträume betrachtet: von 1996 auf 1997, 1998 und 1999. In die Analyse gehen nur Personen ein, die im Jahr 1996 zwischen 18 und 65 Jahre alt waren, und welche in dem jeweiligen Beobachtungszeitraum in einem Privathaushalt in Deutschland gelebt haben und keine Erhebungslücke aufweisen. Betrachtet wird jeweils der Status der Personen zum Zeitpunkt der Erhebung im Jahr 1996 und zum Endzeitpunkt. Beamte und Wehr- sowie Zivildienstleistende werden ausgeschlossen. Die Analyse wird ungewichtet durchgeführt. Als unabhängige Variablen werden die folgenden Merkmale mit einbezogen:

- das Geschlecht,
- die Staatsangehörigkeit,
- das Alter,
- der höchste erreichte Berufsabschluss,
- der Beschäftigungssektor,
- die Beschäftigungsdauer sowie
- der Wohnort (Ost/West)

Alle Variablen wurden als 0/1-kodierte Dummy-Variablen in das Modell einbezogen. Für das Alter der Befragten wurden insgesamt drei Klassen gebildet. Beim höchsten erreichten Berufsabschluss wurde zwischen Personen mit (Fach-) Hochschulabschluss, Personen mit beruflichem Bildungsabschluss und Personen ohne berufsqualifizierendem Abschluss differenziert. Letztere Kategorie dient als Referenzkategorie. Die Schätzungen wurden mit ungewichteten Fällen (Tabelle 8) als auch mit normierten Längsschnittgewichten (Tabelle 9) durchgeführt. Im letzteren Fall wurden die im Campus-File enthaltenen Längsschnittgewichte (HGM_9697, HGM_9698, HGM_9699) auf die Stichprobengröße normiert (siehe hierzu Abschnitt 2.4).

Tabelle 8: Übergang von abhängiger Beschäftigung in Erwerbslosigkeit - ungewichtet

Logistisches Regressionsmodell für Übergänge aus abhängiger Beschäftigung in Erwerbslosigkeit: Ergebnisse von Analysen mit den Daten des CF-MZ-Panels für die Übergänge 1996 auf 1997, 1998 und 1999.

Betrachtung des Status von Personen im Jahr 1996, 1997, 1998 und 1999 (ungewichtet)

Variable (Status 1996)	1996 -> 1997		1996 -> 1998		1996 -> 1999	
	β	p-Wert	β	p-Wert	β	p-Wert
Regressionskonstante	-0,35	0,25	-0,62	0,05	-0,97	0,00
Geschlecht: männlich	-0,06	0,65	-0,05	0,73	0,16	0,27
Staatsangeh.:deutsch	-0,93	0,00	-0,61	0,02	-0,81	0,00
Alter: 31 – 50 (Ref.)						
Alter: 18 - 30	-0,51	0,00	-0,19	0,30	-0,07	0,72
Alter: 51 - 65	0,46	0,00	0,84	0,00	1,03	0,00
Beruflicher Bildungsabschluss	-0,17	0,35	-0,49	0,00	-0,01	0,94
(Fach-)Hochschulabschluss	-0,53	0,06	-0,90	0,00	-1,22	0,00
im öffentlichen Dienst	-0,23	0,22	-0,13	0,48	-0,42	0,05
Betriebszugehörigkeit > 36 Monate	-1,12	0,00	-0,99	0,00	-0,93	0,00
Wohnsitz: Westdeutschland	-1,19	0,00	-1,03	0,00	-1,05	0,00
N	5015		4496		4058	
Nagelkerke R ²	0,103		0,094		0,105	

Quelle: Campus-File des Mikrozensus-Panels 1996 – 1999

Gewichtete und ungewichtete Schätzungen weichen vor allem bei der Variablen „Geschlecht“ voneinander ab. Während die ungewichteten Schätzungen bei einem 10%igen Signifikanzniveau für diese Variable keinen signifikanten Einfluss erkennen lassen, führt die Längsschnittgewichtung mit Mobilitätsgewichten dazu, dass abhängig beschäftigte Männer zumindest bei den ersten beiden betrachteten Übergängen eine signifikant geringere Wahrscheinlichkeit aufweisen, erwerbslos zu werden. Dies bedeutet, dass die durch das Mobilitätsgewicht erfasste und korrigierte räumliche Mobilität offensichtlich einen Einfluss auf die Modellergebnisse hat. Die Werte der anderen geschätzten Koeffizienten entsprechen weitestgehend den inhaltlichen Erwartungen. Beispielsweise ergibt sich bei einer 1996 vorliegenden Betriebszugehörigkeitsdauer von über drei Jahren für alle drei betrachteten Zeiträume ein starker negativer Effekt, d.h. Personen, die schon über drei Jahre beim gleichen Arbeitgeber abhängig beschäftigt sind, weisen gegenüber Personen mit kürzerer Betriebszugehörigkeitsdauer eine deutlich geringe Wahrscheinlichkeit auf, erwerbslos zu werden. Das Risiko erwerbslos zu werden wird zudem durch das Vorhandensein eines (Fach-) Hochschulabschlusses signifikant reduziert. Ein höheres Risiko tragen dagegen Nichtdeutsche und ältere abhängig Beschäftigte. Zudem ist die Wahrscheinlichkeit erwerbslos zu werden in Ostdeutschland deutlich höher als in Westdeutschland.

Zu beachten ist, dass trotz der plausiblen Resultate bei der Interpretation der Ergebnisse berücksichtigt werden muss, dass Angaben zum Erwerbsstatus zwischen den Erhebungszeitpunkten nicht vorliegen und in die Modellierung nicht mit einfließen. Die Berechnung der Übergangswahrscheinlichkeiten erfolgt allein auf Basis des Erwerbsstatus der Personen zu den konkreten Erhebungszeitpunkten.

Tabelle 9: Übergang von abhängiger Beschäftigung in Erwerbslosigkeit - längsschnittgewichtet

Logistisches Regressionsmodell für Übergänge aus abhängiger Beschäftigung in Erwerbslosigkeit: Ergebnisse von Analysen mit den Daten des CF-MZ-Panels für die Übergänge 1996 auf 1997, 1998 und 1999.

Betrachtung des Status von Personen im Jahr 1996, 1997, 1998 und 1999 (gewichtet mit normierten Längsschnittgewichten unter Berücksichtigung der Mobilitätsgewichte)

Variable (Status 1996)	1996 -> 1997		1996 -> 1998		1996 -> 1999	
	β	p-Wert	β	p-Wert	β	p-Wert
Regressionskonstante	-0,35	0,18	-0,71	0,01	-0,98	0,00
Geschlecht: männlich	-0,24	0,07	-0,23	0,07	-0,01	0,96
Staatsangeh.:deutsch	-0,92	0,00	-0,45	0,02	-0,90	0,00
Alter: 18 - 30	-0,42	0,01	-0,09	0,59	0,23	0,17
Alter: 51 - 65	0,50	0,00	0,89	0,00	1,08	0,00
Beruflicher Bildungsabschluss (Fach-)Hochschulabschluss	-0,22	0,20	-0,65	0,00	-0,03	0,87
im öffentlichen Dienst	-0,32	0,21	-0,60	0,01	-0,75	0,02
Betriebszugehörigkeit > 36 Monate	-0,34	0,08	-0,03	0,88	-0,09	0,66
Wohnsitz: Westdeutschland	-1,08	0,00	-0,93	0,00	-0,77	0,00
N	5015		4496		4058	
Nagelkerke R ²	0,098		0,090		0,095	

Quelle: Campus-File des Mikrozensus-Panels 1996 – 1999

3.2 Übergang von Familienstand „verheiratet“ in den Familienstand „geschieden“

In dem folgenden Beispiel soll der Übergang von dem Zustand „verheiratet“ in den Zustand „geschieden“ mit Hilfe eines Logit-Modells untersucht werden. Als unabhängige Variablen werden die folgenden Merkmale mit einbezogen:

- die Ehedauer bis zum Ausgangszustand 1996 (4 Dummy-Variablen)
- die Staatsangehörigkeit,
- der Bildungsabschluss,
- Kinder unter 16 Jahren leben 1996 im Haushalt
- die Person war 1996 erwerbstätig

In die Analyse gehen nur Personen bzw. Ehepaare ein, die im Jahr 1996 älter als 18 Jahre alt und verheiratet waren, sowie in dem jeweiligen Beobachtungszeitraum in einem Privathaushalt in Deutschland gelebt haben und für den Beobachtungszeitraum keine Erhebungslücke aufweisen. Betrachtet wird wiederum der Status der Personen zum Zeitpunkt der Erhebung im Jahr 1996 und zum Endzeitpunkt. Ein Teil der statistischen Einheiten sind zusammenlebende Ehepaare. Insofern muss davon ausgegangen werden, dass die Annahme von identisch und unabhängig verteilten Residuen nicht ganz zutrifft und die Standardfehler unterschätzt werden können. Wir haben daher eine robuste Schätzung der Standardfehler vorgenommen.¹²

¹² Die Schätzung erfolgte mit STATA 8.2 und der Option „Cluster“, wobei als Clustervariable die Haushaltsnummer diente.

Die geschätzten Koeffizienten lassen sich der Tabelle 10 entnehmen. Bei der Analyse ist zu beachten, dass sich Ehepartner kurz vor oder nach einer Scheidung in der Regel auch räumlich trennen, was dazu führt, dass zumindest ein Partner wegen des Auszugs nicht mehr in der Stichprobe erscheint. Es wurden daher auch hier zusätzlich die Schätzungen mit Längsschnittgewichten inklusive Mobilitätsgewichten angegeben.

Tabelle 10: Übergang von dem Zustand „verheiratet“ in den Zustand „geschieden“ - ungewichtet

Logistisches Regressionsmodell für Übergänge aus dem Zustand "verheiratet" in den Zustand "geschieden": Ergebnisse von Analysen mit den Daten des CF-MZ-Panels für die Übergänge 1996 auf 1997, 1998 und 1999.
Betrachtung des Status von Personen im Jahr 1996, 1997, 1998 und 1999 (ungewichtet, robuste Schätzung)

Variable (Status 1996)	1996 -> 1997		1996 -> 1998		1996 -> 1999	
	β	p-Wert	β	p-Wert	β	p-Wert
Regressionskonstante	-4,78	0,00	-4,78	0,00	-5,27	0,00
Ehedauer <= 5 Jahre (Ref.)						
Ehedauer 6 – 10 Jahre	0,45	0,41	0,68	0,16	0,83	0,05
Ehedauer 11 - 15 Jahre	0,33	0,56	0,24	0,65	0,44	0,33
Ehedauer > 15 Jahre	-1,59	0,01	-1,29	0,01	-1,02	0,02
Staatsangeh.:deutsch	0,73	0,46	0,90	0,39	1,00	0,32
(Fach-)Hochschulabschluss	0,14	0,75	-0,15	0,73	0,00	0,99
mind. ein Kind < 16 Jahre im HH	-0,79	0,09	-0,45	0,24	-0,38	0,29
erwerbstätig	-0,07	0,87	0,00	1,00	0,59	0,08
Cluster (Haushalte)	3409		3199		3036	
N	6532		6067		5708	
Pseudo R ²	0,053		0,051		0,062	

Quelle: Campus-File des Mikrozensus-Panels 1996 - 1999

Tabelle 11: Übergang von dem Zustand „verheiratet“ in den Zustand „geschieden“ - gewichtet

Logistisches Regressionsmodell für Übergänge aus dem Zustand "verheiratet" in den Zustand "geschieden": Ergebnisse von Analysen mit den Daten des CF-MZ-Panels für die Übergänge 1996 auf 1997, 1998 und 1999.
Betrachtung des Status von Personen im Jahr 1996, 1997, 1998 und 1999 (gewichtet, robuste Schätzung)

Variable (Status 1996)	1996 -> 1997		1996 -> 1998		1996 -> 1999	
	β	p-Wert	β	p-Wert	β	p-Wert
Regressionskonstante	-4,60	0,00	-5,06	0,00	-4,98	0,00
Ehedauer <= 5 Jahre (Ref.)						
Ehedauer 6 - 10 Jahre	0,52	0,39	0,56	0,28	0,78	0,07
Ehedauer 11 - 15 Jahre	0,18	0,76	0,02	0,97	0,30	0,51
Ehedauer > 15 Jahre	-1,75	0,00	-1,57	0,00	-1,08	0,02
Staatsangeh.:deutsch	1,08	0,28	1,87	0,07	1,50	0,13
(Fach-)Hochschulabschluss	0,01	0,99	-0,24	0,59	-0,05	0,88
mind. ein Kind < 16 Jahre im HH	-0,88	0,07	-0,60	0,13	-0,51	0,17
erwerbstätig	-0,20	0,64	-0,02	0,95	0,53	0,13
Cluster (Haushalte)	3409		3199		3036	
N	6532		6067		5708	
Pseudo R ²	0,064		0,068		0,069	

Quelle: Campus-File des Mikrozensus-Panels 1996 - 1999

Alle Ergebnisse deuten darauf hin, dass das Scheidungsrisiko bei einer längeren Ehedauer von über 15 Jahren signifikant absinkt. Falls „Kinder unter 16 Jahren im Haushalt“ im Haushalt leben, reduziert sich das Scheidungsrisiko von 1996 auf 1997 im gewichteten als auch ungewichteten Fall deutlich. Gegenüber den ungewichteten weisen die gewichteten Schätzungen in Bezug auf die Ehedauer als auch der Kinder unter 16 Jahren im Haushalt etwas größere Logit-Koeffizienten auf. Hier übt anscheinend die durch das Mobilitätsgewicht erfasste räumliche Mobilität einen korrigierenden Einfluss aus.

Die anderen einbezogenen Merkmale wie die Staatsangehörigkeit sowie ein (Fach-) Hochschulabschluss und „1996 erwerbstätig“ haben keinen signifikanten Einfluss auf die Scheidungswahrscheinlichkeit.

Literatur

- Bihler, Wolf, Robert Herter-Eschweiler und Heike Wirth (2006): Das faktisch anonymisierte Mikrozensus-Panel 1996-1999. In: Statistisches Bundesamt (Hrsg.): Handbuch Mikrozensus-Panel 1996-1999, Kapitel 3. Seite 61 – 90.
- Herter-Eschweiler, Robert (2002): Der Mikrozensus als Datenquelle für Panelauswertungen. In: Statistisches Bundesamt (Hrsg.): Handbuch Mikrozensus-Panel 1996-1999, Kapitel 2. Seite 33 – 60.
- Herter-Eschweiler, Robert (2003): (SPSS-) Programmablauf zur Erstellung des faktisch anonymisierten Mikrozensus-Längsschnitts 1996-1999. Bonn: Statistisches Bundesamt (unveröffentlichtes Manuskript).
- Herter-Eschweiler, Robert (2006a): Der Mikrozensus als Panel: Längsschnittverknüpfung und Selektivitätsanalysen im Bereich der Art der Erwerbsbeteiligung und familialen Lebensformen. Methodenverbund „Aufbereitung und Bereitstellung des Mikrozensus als Panelstichprobe“. Arbeitspapier Nr.: 4. URL: <http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Content/Wissenschaftsforum/MethodenVerfahren/Mikrozensus/Arbeitspapiere/Arbeitspapier4,property=file.pdf>
- Herter-Eschweiler, Robert (2006b): SPSS-Syntax „MZ-Panel Substichprobenziehung“ (unveröffentlichte SPSS Syntax).
- Marek, Ivo und Sandra Rohloff (2006): Das Hochrechnungsverfahren für Längsschnittauswertungen im Mikrozensus-Panel 1996-1999. In: Statistisches Bundesamt (Hrsg.): Handbuch Mikrozensus-Panel 1996-1999, Kapitel 4. Seite 91 – 104.
- Särndal, Carl-Erik, Bengt Swensson und Jan Wretman (1992): Model Assisted Survey Sampling. New York: Springer.
- Schimpl-Neimanns, Bernhard (2006): Filekonzept zum Mikrozensuspanel. Methodenverbund „Aufbereitung und Bereitstellung des Mikrozensus als Panelstichprobe“. Arbeitspapier Nr.: 12. URL: <http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Content/Wissenschaftsforum/MethodenVerfahren/Mikrozensus/Arbeitspapiere/Arbeitspapier12,property=file.pdf>
- Schimpl-Neimanns, Bernhard und Ulrich Rendtel (2001): SAS-, SPSS- und STATA-Programme zur Berechnung der Varianz von Populationsschätzern im Mikrozensus ab 1996. ZUMA-Methodenbericht 2001/04.
- SAS-Programm: „VarMZ_A.SAS“ im Internet unter: http://www.gesis.org/Dauerbeobachtung/gml/Service/Mikrodaten-Tools/Varianz/varianz_tools.htm.
Programmautoren: Ulrich Rendtel und Bernhard Schimpl-Neimanns.

Stata, 2007: Stata Survey Data Reference Manual. Release 10. College Station, TX: Stata Press.

Statistisches Bundesamt (Hrsg.) (2006): Handbuch Mikrozensus-Panel 1996-1999. Methodenverbund „Aufbereitung und Bereitstellung des Mikrozensus als Panelstichprobe“. Version 0.2 – Juli 2006. URL: <http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Content/Wissenschaftsforum/MethodenVerfahren/Mikrozensus/Metadaten/HandbuchMikrozensuspanel,property=file.pdf>

4. Anhang

1. Randverteilungen ausgewählter Merkmale in den Mikrozensus-Originalfiles und den Querschnitten des Campus Files des Mikrozensus-Panels 1996-1999 für die Jahre 1997, 1998 und 1999

Tabelle 12: Bevölkerung in Privathaushalten im Erhebungsjahr 1997 mit Personengewichtung

Tabelle 13: Bevölkerung in Privathaushalten im Erhebungsjahr 1998 mit Personengewichtung

Tabelle 14: Bevölkerung in Privathaushalten im Erhebungsjahr 1999 mit Personengewichtung

2. Gesamtwerte und Standardfehler ausgewählter Merkmale im Mikrozensus-Panel 1996-1999 Originalfile und im Campus-File in den Erhebungsjahren 1997, 1998 und 1999

Tabelle 15: Gesamtwerte und Standardfehler ausgewählter Merkmale im Mikrozensus-Panel 1996-1999 Originalfile und im Campus-File. Bevölkerung in Privathaushalten im Erhebungsjahr 1997 (Hochrechnung mit Personengewichtung)

Tabelle 16: Gesamtwerte und Standardfehler ausgewählter Merkmale im Mikrozensus-Panel 1996-1999 Originalfile und im Campus-File. Bevölkerung in Privathaushalten im Erhebungsjahr 1998 (Hochrechnung mit Personengewichtung)

Tabelle 17: Gesamtwerte und Standardfehler ausgewählter Merkmale im Mikrozensus-Panel 1996-1999 Originalfile und Campus-File. Bevölkerung in Privathaushalten im Erhebungsjahr 1999 (Hochrechnung mit Personengewichtung)

3. Tabelle 18: Merkmalskatalog des Campus Files des MZ-Panels 1996 - 1999