

# Das Hochrechnungsverfahren für Längsschnittauswertungen aus dem Mikrozensus

Sandra Rohloff  
Statistisches Bundesamt

4. Nutzerkonferenz  
„Forschung mit dem Mikrozensus:  
Analysen zur Sozialstruktur und zum Arbeitsmarkt“

12. - 13. Oktober 2005, Mannheim

## **Zusammenfassung**

Im Rahmen des Projektes „Aufbereitung und Bereitstellung des Mikrozensus als Panelstichprobe“ befasst sich das Teilprojekt „Methoden“ unter anderem mit der Konstruktion und Bereitstellung von Hochrechnungsfaktoren für Längsschnittanalysen des Mikrozensus-Panels. Das Hochrechnungsverfahren wird im Folgenden detailliert erläutert. Unter anderem wird auf die Abgrenzung der Grundgesamtheit und der Stichprobe für die Betrachtung des Längsschnitts eingegangen. Anschließend wird das Kalibrierungsverfahren erklärt und die Eckwertvektoren beschrieben. Abschließend wird kurz auf Analysemöglichkeiten des Mikrozensuspanels unter Verwendung der berechneten Hochrechnungsfaktoren eingegangen.

## **1 Einleitung**

Ziel des Projektes „Aufbereitung und Bereitstellung des Mikrozensus als Panelstichprobe“ ist es, Analysemöglichkeiten des Mikrozensus als Längsschnittstichprobe zu untersuchen und anschließend einen Scientific Use File zur Verfügung zu stellen. An dem vom Bundesministerium für Bildung und Forschung geförderten Projekt arbeiten die Freie Universität Berlin, das Zentrum für Umfragen, Methoden und Analysen, das Landesamt für Datenverarbeitung und Statistik Nordrhein-Westfalen und das Statistische Bundesamt.

Der Mikrozensus stellt die größte Haushaltserhebung der amtlichen Statistik in Deutschland dar und wurde bis 2004 jährlich im April erhoben. Er

ist als rotierendes Panel konstruiert, wobei jedes Jahr ein Viertel der Stichprobenauswahlbezirke ausgetauscht werden und somit jedes Rotationsviertel im Erwartungswert den gleichen Stichprobenumfang an Haushalten besitzt. Jedes Rotationsviertel kann daher durch Zusammenfügen der entsprechenden Datensätze einen maximal vierjährigen Längsschnitt bilden. Zwar stellt der Mikrozensus eine Haushaltserhebung dar, doch werden für alle Haushaltsmitglieder der zu befragenden Haushalte Personendaten erhoben. Somit bietet der Mikrozensus Analysemöglichkeiten von Personendatensätzen eines maximal vierjährigen Längsschnitts eines Rotationsviertels. Die rechtlichen Grundlagen, um diese Möglichkeiten nutzen zu können, wurden erst mit dem Mikrozensusgesetz vom 17. Januar 1996 geschaffen. Erst mit diesem Gesetz ist es möglich, befragten Haushalten identische Ordnungsnummern über alle Erhebungsperioden zuzuweisen und diese bis zum Ende der Aufbereitung der letzten Erhebung zu speichern. Aus diesem Grund konzentriert sich das Projekt „Aufbereitung und Bereitstellung des Mikrozensus als Panelstichprobe“ auf den Längsschnitt der Jahre 1996 bis 1999.

Als vorteilhaft für Längsschnittuntersuchungen erweist sich der große Stichprobenumfang des Mikrozensus. Jährlich werden etwa ein Prozent aller Haushalte und Personen in Deutschland befragt. Die Bildung des Längsschnitts eines Rotationsviertels umfasst daher etwa 0,25 Prozent der Gesamtbevölkerung und stellt somit das datenmengenmäßig größte Haushaltspanel Deutschlands dar. Von Vorteil ist ebenfalls die Auskunftspflicht der Befragten, da sie zu einem geringen Unit-Nonresponse von etwa drei Prozent führt. Neben der großen Menge an Datensätzen bietet das Frageprogramm des Mikrozensus eine Vielzahl an Untersuchungsvariablen. Diese können nicht nur für Analysen zu Bereichen wie Bildung, berufliche Tätigkeit und Familie genutzt werden, sondern erweisen sich auch für die Hochrechnung des Mikrozensus als vorteilhaft.

Das Stichprobendesign des Mikrozensus bietet allerdings nicht nur Vorteile für Längsschnittanalysen. Die Tatsache, dass der Mikrozensus als Flächenstichprobe erhoben wird, stellt einen eindeutigen Nachteil für Panelanalysen dar. So werden Haushalte und Personen nicht direkt aus der Grundgesamtheit für die Stichprobe gezogen, sondern die Auswahlseinheiten werden über räumliche Zuordnungen gebildet. Dies hat zur Folge, dass Haushalte, die aus einem zu befragenden Auswahlbezirk ziehen, nicht weiterverfolgt werden. Sie werden bei der nächsten Erhebung nicht mehr befragt und ihre Angaben stehen somit für Längsschnittanalysen nicht weiter zur Verfügung. Dagegen werden Haushalte und Personen, die in einen zu befragenden Auswahlbezirk ziehen, bei der nächsten Befragung miterhoben. Auch von diesen Personen existieren keine vollständigen Angaben über den gesamten vierjährigen Mikrozensus-Längsschnitt. Dieser Informationsverlust, der durch Zuzug oder Fortzug einer Person oder eines Haushalts entsteht, kann

wegen der Betrachtung von nur räumlich immobilen Personen zu einem Bias führen. Ziel der Längsschnitthochrechnung des Mikrozensuspanels ist somit die Reduktion dieses Bias durch entsprechende Gewichtung der immobilen Personen.

Weitere Informationen zur Zusammenführung der Mikrozensusdatensätze zum Längsschnittmaterial und zu möglichen Auswertungen anhand des Mikrozensuslängsschnitts sind unter [3] zu finden.

## **2 Das Hochrechnungsverfahren des Mikrozensus im Querschnitt in den Jahren 1996 bis 1999**

In diesem Abschnitt wird kurz auf die Hochrechnung des Mikrozensus-Querschnitts während der Jahre 1996 bis 1999 eingegangen. Das Erhebungsdesign des Mikrozensus wurde ab 2005 geändert, somit sind nachfolgende Angaben bezüglich der Hochrechnung des Querschnitts nur für die Jahre zwischen 1996 und 2004 gültig.

Zwar unterliegen die Befragten im Mikrozensus der gesetzlichen Auskunftspflicht, dennoch kommt es in der Regel zu einem Unit-Nonresponse von etwa drei Prozent. Um mögliche Verzerrungen durch die Antwortausfälle auszugleichen, wird die Stichprobe an Zahlen aus der Bevölkerungsfortschreibung angepasst. Da das Antwortverhalten der zu befragenden Personen von soziodemographischen Merkmalen abhängig ist, erfolgt eine Anpassung gegliedert nach bestimmten Aspekten, wie der Staatsangehörigkeit und dem Geschlecht. Zusätzlich wird die räumliche Zuordnung berücksichtigt, wie die regionalen Anpassungsschichten. Diese Schichten werden lediglich für die Auswertungen des Mikrozensus über die Anzahl der Einwohner gebildet und umfassen etwa 650.000 Personen. Innerhalb dieser regionalen Anpassungsschichten erfolgt für die Anpassung an Eckwerte eine Trennung nach Deutschen und Ausländer in Kombination mit dem Geschlecht. Zusätzlich zu den Angaben auf regionaler Anpassungsschicht wird an die Zahl der Berufssoldaten und der Wehrpflichtigen innerhalb der Regierungsbezirke angepasst.

Neben der Reduktion eines durch Unit-Nonresponse verursachten möglichen Bias kann im Allgemeinen die Anpassung des Mikrozensus im Querschnitt an Zahlen aus der laufenden Bevölkerungsfortschreibung auch zur Verringerung des Zufallsfehlers beitragen. Ein weiteres Ziel dieser Anpassung ist es, identische Eckzahlen für die Bevölkerung aus Mikrozensus und laufender Bevölkerungsfortschreibung zu erhalten.

Die in den jeweiligen Jahren von 1996 bis 1999 berechneten Hochrechnungsfaktoren des Mikrozensus-Querschnittmaterials stehen auch im Längsschnitt für Querschnittsauswertungen zur Verfügung. Eine Analyse des jeweiligen Querschnitts ist daher problemlos möglich, es ist lediglich der verminderte Stichprobenumfang von einem Viertel zu berücksichtigen.

### **3 Das Hochrechnungsverfahren des Mikrozensus-Längsschnitts der Jahre 1996 bis 1999**

#### **3.1 Ziele der Hochrechnung des Mikrozensus-Längsschnitts**

Ziel der Gewichtung des Mikrozensus-Längsschnitts ist es, Aussagen über Veränderungen in den zu analysierenden Variablen von der Stichprobe auf die Gesamtbevölkerung Deutschlands zwischen 1996 und 1999 übertragen zu können. So möchte man beispielsweise die Anzahl der Personen ermitteln können, die im betrachteten Zeitraum ihren Erwerbsstatus wechseln. Ebenso wie bei der Querschnittshochrechnung sind daher Antwortausfälle zu kompensieren. Zusätzlich zu den Nonresponse-Fällen im Querschnitt kann es im Längsschnitt zu Ausfällen aufgrund räumlicher Mobilität kommen, da keine Weiterverfolgung fortziehender Personen erfolgt. Diese sind ebenfalls zu beachten.

Zusätzlich zu den Antwortausfällen sind bei der Längsschnittgewichtung Veränderungen der Bevölkerungszahl im Zeitverlauf zu berücksichtigen. Jährlich ändert sich die Zusammensetzung der Bevölkerung. Geburten beispielsweise lassen die Anzahl der Personen in Deutschland steigen. Wird ein Kind in einen zu befragenden Haushalt geboren, so wird es bei der nächsten Erhebung miterfasst. Daher zählt jedes Neugeborene zur Grundgesamtheit und muss bei einer Hochrechnung berücksichtigt werden. Sterbefälle dagegen vermindern die Grundgesamtheit im darauf folgenden Jahr und können daher ebenfalls nicht vernachlässigt werden. Neben diesen demographischen Änderungen können auch Umzüge zu Veränderungen in den Bevölkerungszusammensetzung führen. Zuzüge aus dem Ausland und Fortzüge ins Ausland sind zu berücksichtigen, um Aussagen über die Gesamtbevölkerung Deutschlands von 1996 bis 1999 treffen zu können. Die Berücksichtigung dieser „mobilen“ Personen stellt eine wesentliche Erweiterung zur Hochrechnung im Querschnitt dar. Im Querschnitt können keine Veränderungen der Bevölkerung im Zeitverlauf betrachtet werden.

### 3.2 Der Datensatz des Mikrozensus-Längsschnitts 1996 - 1999

Weiterhin ist für Analysen des vorliegenden Längsschnittmaterials von 1996 bis 1999 wichtig zu beachten, dass das Datenmaterial nur Personen in Privathaushalten beinhaltet. Im Mikrozensus werden jedoch alle in einem Stichprobenauswahlbezirk wohnenden Personen befragt. Somit sind im Querschnitt des Mikrozensus sowohl Personen in Privathaushalten als auch Personen in Gemeinschaftsunterkünften vertreten. Da sich jedoch in Gemeinschaftsunterkünften die Zusammensetzung der Haushalte häufiger ändern kann und die Anzahl der Proxy-Interviews größer ist als in Privathaushalten, war die Zusammenführung der einzelnen Personendatensätze in Gemeinschaftsunterkünften bisher nicht möglich. Es wurde daher auf die Datensätze der Gemeinschaftsunterkünfte im Längsschnittdatei verzichtet. Dies ist bei der Hochrechnung und bei späteren Analysen zu berücksichtigen.

Ein weiteres Problem des nachträglichen Zusammenführens der Einzeldatensätze stellen unplausible und fehlerhafte Datensätze dar. Zieht beispielsweise eine Person aus einem Haushalt aus und kommt im gleichen Jahr ein neues Haushaltsmitglied hinzu, so ist nicht in allen Fällen eindeutig zu erkennen, dass es sich hier um zwei verschiedene Personen handelt.

Neben möglichen Fehlern beim Zusammenfügen der Datensätze kann es bei der Übertragung der Daten in das Datenerfassungsprogramm oder dem Ausfüllen des Fragebogens zu Fehlern gekommen sein. Zum Beispiel ändert sich das Geburtsjahr in einigen Fällen. Hier können Zahlendreher beim Eingeben des Jahres ebenso wie beim Ausfüllen des Fragebogens aufgetreten sein, beziehungsweise falls es sich um ein Proxy-Interview handelt, kann es passieren, dass hier nicht das exakte Jahr angegeben wurde. Datensätze, die kein konsistentes Geburtsjahr aufweisen, wurden einer Plausibilitätsprüfung unterzogen und es wurde für alle Personen eine neue Variable generiert, die das korrekte Geburtsjahr beinhaltet. Nicht nur beim Geburtsjahr kann es zu solchen Fehlern gekommen sein. Da jedoch keine generelle Plausibilitätsprüfung im Längsschnitt erfolgte, sind fehlerhafte und unplausible Angaben nicht immer eindeutig zu erkennen. Um Auswertungen zu erleichtern und weiteren Fehlern vorzubeugen, wurden daher für die wichtigsten Variablen wie beispielsweise Geburtsjahr und Geschlecht neue Variablen konstruiert, die für allen vier Jahren gleich sind.

Für Ereignisse wie Geburten oder Sterbefälle einer Person, Zuzüge oder Fortzüge einer Person oder eines Haushaltes wurden ebenfalls neue Variablen definiert, die neben dem Ereignis das Jahr angeben, in dem es stattfand. Anhand dieser Variablen lassen sich mobile Personen leicht klassifizieren und bei der Hochrechnung entsprechend berücksichtigen. Zu beachten ist lediglich, dass bei den Zuzügen und Fortzügen sowohl Umzüge innerhalb

Deutschlands als auch ins und aus dem Ausland codiert wurden. Ein Zuzug aus dem Ausland kann über die Frage nach dem Wohnsitz im Jahr vor der Erhebung ermittelt werden. Die Beantwortung dieser Frage ist jedoch für den Befragten freiwillig und unterliegt nicht der Auskunftspflicht. Es kann daher passieren, dass eine Person, die im Jahr vor der Erhebung im Ausland lebte, dies nicht angibt und somit der Datensatz nicht als Zuzug aus dem Ausland gewertet werden kann. Aufgrund der freiwilligen Beantwortung dieser Frage kommt es zu einer Unterrepräsentation der Zuzüge aus dem Ausland in der Stichprobe. Weiter ist bei den Zuzügen aus dem Ausland zu beachten, dass die Angaben zum Wohnsitz im Jahr vor der Erhebung für das Jahr 1999 aus Plausibilitätsgründen nicht verfügbar sind. Für dieses Jahr können keine Zuzüge aus dem Ausland codiert werden. Dies ist bei der Hochrechnung zu berücksichtigen, für diese stehen nur die Informationen zu Zuzügen aus dem Ausland für die Übergänge 1996 – 1997 und 1997 – 1998 zur Verfügung.

Bei den Fortzügen ergibt sich das Problem, dass man in den meisten Fällen nicht weiß, ob der Haushalt oder die Person verzogen ist, da keine Weiterverfolgung erfolgt. In wenigen Fällen ist bekannt, dass eine Person oder ein Haushalt fortgezogen ist, jedoch nicht, wohin. Man kann somit einen Umzug innerhalb Deutschlands nicht von einem Fortzug ins Ausland unterscheiden. Fortzüge ins Ausland müssen zwar formal bei der Hochrechnung betrachtet werden, da dieses Ereignis aber keinem Datensatz der Stichprobe zugeordnet werden kann, werden diese im Weiteren zusammen mit den Sterbefällen betrachtet. Dies ist möglich, da beide Ereignisse die Grundgesamtheit verringern.

### **3.3 Grundgesamtheit und Stichprobe des Mikrozensus-Panels**

Neben allen mobilen Personen zählen auch alle immobilen Personen zur Grundgesamtheit der Erhebung. Insgesamt besteht die Grundgesamtheit, auf die hochgerechnet wird, für den Mikrozensus-Längsschnitt 1996 – 1999 aus allen Personen, die 1996 und 1999 oder in einem der beiden Jahre mit Hauptwohnsitz in einem Privathaushalt in Deutschland lebten. Falls nicht der gesamte vierjährige Längsschnitt betrachtet werden soll sondern nur ein Teil, zum Beispiel der zweijährige Längsschnitt von 1996 bis 1997, so umfasst die Grundgesamtheit alle Personen, die 1996 und 1997 oder in einem der beiden Jahre mit Hauptwohnsitz in einem Privathaushalt in Deutschland lebten. Für Längsschnittanalysen wird in der Regel eine Teilmenge der hier beschriebenen Grundgesamtheit betrachtet. Bei der Abgrenzung der Grundgesamtheit für die Hochrechnung ist zu beachten, dass nur Personen mit Hauptwohnsitz zu dieser Menge gehören. Personen am Nebenwohnsitz zählen nicht zur Grundgesamtheit, da sie über den Hauptwohnsitz erfasst werden und man somit eine Doppelerfassung vermeiden kann.

Im Gegensatz zur Grundgesamtheit wird die Stichprobe des Mikrozensus-Längsschnitts, für die Hochrechnungsfaktoren berechnet werden, deutlich differenzierter abgegrenzt. Zwar zählen zur Stichprobe des Längsschnitts 1996 – 1999 alle Personen, die 1996 und 1999 oder in einem der beiden Jahre in Deutschland lebten, doch um Doppelerfassungen aufgrund von räumlicher Mobilität zu vermeiden, wird die Stichprobe folgendermaßen abgegrenzt: Zur Stichprobe des Längsschnitts 1996 – 1999 zählen alle Personen, die

- in allen vier Jahren in einem Stichprobenauswahlbezirk lebten
- 1996 in einem Stichprobenauswahlbezirk lebten und vor der letzten Befragung im April 1999 verstarben
- 1996 in einem Stichprobenauswahlbezirk lebten und vor der letzten Befragung im April 1999 ins Ausland verzogen
- nach April 1996 geboren wurden und 1999 in einem befragten Haushalt lebten
- nach April 1996 aus dem Ausland in einen Stichprobenauswahlbezirk zuzogen und in diesem 1999 erfasst wurden

Für den Längsschnitt 1996 – 1997 und den Längsschnitt 1996 – 1998 erfolgt diese Abgrenzung analog, es ist lediglich die Zahl 1999 durch das entsprechende letzte Jahr des betrachteten Längsschnitt zu ersetzen.

Die Abgrenzung der Stichprobe erfolgt über die oben beschriebenen Variablen Geburten, Sterbefälle, Zuzüge und Fortzüge, Wohnsitz im Jahr vor der Erhebung und einer weiteren Variablen, die für jeden Datensatz angibt, in welchem Jahr die Person befragt werden konnte, bzw. in welchem Jahr ein Ausfall auftrat und dieser erklärbar war oder nicht geklärt werden konnte.

### **3.4 Kalibrierung des Mirkozensus-Panels**

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. So können beispielsweise

Ereignisse wie Eheschließungen und Ehescheidungen zu Umzügen und somit möglicherweise zum Ausfall einer oder mehrerer Stichprobenpersonen führen. Bei Umzügen innerhalb Deutschlands spielt das Alter der Personen eine große Rolle. Kinder, die bisher bei den Eltern wohnten, ziehen möglicherweise zum Studieren in eine andere Stadt oder gründen nach der Ausbildung einen eigenen Haushalt. Alte Menschen ziehen ins Altersheim oder zu Verwandten und geben somit ihren Haushalt auf. Um diese Veränderungen zu berücksichtigen, wird auch an Variablen wie Alter, Eheschließungen und Ehescheidungen angepasst.

Zur Bestimmung des Hochrechnungsfaktors für jede Person wird neben der Anpassung an Eckwerte aus den Bevölkerungsdaten die Wahrscheinlichkeit berechnet, dass die Person in der Stichprobe erfasst wurde. Diese wird teilweise über die Ziehungswahrscheinlichkeit  $\pi_k$  für jede Person  $k$  bestimmt. Die Ziehungswahrscheinlichkeit lässt sich über den Auswahlatz aus der Grundgesamtheit ermitteln. Betrachtet wird zunächst die Wahrscheinlichkeit im Mikrozensus-Querschnitt vertreten zu sein. Die Ziehungswahrscheinlichkeit des Mikrozensus-Querschnitts liegt aufgrund des Auswahlatzes bei exakt einem Prozent. Für den Mikrozensus-Längsschnitt wird jedoch nur ein Rotationsviertel betrachtet. Somit muss die Ziehungswahrscheinlichkeit für den Mikrozensus-Querschnitt durch vier dividiert werden, um die Ziehungswahrscheinlichkeit für den Längsschnitt zu erhalten. Die Wahrscheinlichkeit, im Mikrozensus-Längsschnitt zu sein, ergibt sich nun aus der Ziehungswahrscheinlichkeit für den Längsschnitt  $\pi_k$  für jede Person  $k$  und der Wahrscheinlichkeit für Immobilität. Da mobile Personen nicht weiterverfolgt werden und somit aus dem Längsschnitt ausfallen, müssen sie über die immobilen repräsentiert werden. Hierzu wird für jede immobile Person  $k$  ein so genannter Mobilitätsfaktor  $\hat{\theta}_k$  geschätzt, der die Wahrscheinlichkeit angibt, für die Zeitspanne des betrachteten Längsschnitts immobil zu sein und somit über alle betrachteten Jahre befragt zu werden. Um den Zufallsfehler und den möglichen durch Nonresponse entstehenden Bias auszugleichen, wird der Kehrwert dieses Produkts der Wahrscheinlichkeiten mit einem Korrekturfaktor  $g_k$  multipliziert, um so für jede Person  $k$  den Hochrechnungsfaktor  $w_k$  zu erhalten.

$$w_k = \frac{g_k}{\pi_k \cdot \hat{\theta}_k} \quad (1)$$

Die Berechnung von  $g_k$  für jede Person  $k$  geschieht über die Anpassung an die Eckwertvektoren aus der Bevölkerung. Hierbei geht die Differenz des Vektors  $\mathbf{t}_x$  der tatsächlichen Totalwerte der Hilfsmerkmale und des geschätzten Horwitz-Thompson-Schätzers der Totalwerte mit in die Berechnung ein.  $x_k$  bezeichnet für jede Person  $k$  den Vektor aller Ausprägungen der Hilfsmerkmale für die Person  $k$ .



$$g_k = 1 + (\mathbf{t}_x - \hat{\mathbf{t}}_{x,HT})' \left( \sum_{k=1}^n \frac{\mathbf{x}_k \mathbf{x}_k'}{\pi_k \hat{\theta}_k} \right)^{-1} \mathbf{x}_k \quad (2)$$

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.

$$\hat{\mathbf{t}}_{x,HT} = \sum_{k=1}^n \frac{\mathbf{x}_k}{\pi_k \hat{\theta}_k} \quad (3)$$

Insgesamt erhält man für die Berechnung des Regressionsschätzers  $\hat{\mathbf{t}}_y$  folgende Beziehungen:

$$\hat{\mathbf{t}}_y = \hat{\mathbf{t}}_{y,HT} + \hat{\mathbf{B}}(\mathbf{t}_x - \hat{\mathbf{t}}_{x,HT}) \quad (4)$$

$$= \sum_{k=1}^n \left( 1 + (\mathbf{t}_x - \hat{\mathbf{t}}_{x,HT})' \left( \sum_{k=1}^n \frac{\mathbf{x}_k \mathbf{x}_k'}{\pi_k \hat{\theta}_k} \right)^{-1} \mathbf{x}_k \right) \frac{1}{\pi_k \cdot \hat{\theta}_k} y_k \quad (5)$$

$$= \sum_{k=1}^n w_k y_k \quad (6)$$

### 3.5 Die geschätzten Mobilitätsfaktoren

Die Mobilitätsfaktoren sind mittels einer logistischer Regression geschätzte Wahrscheinlichkeiten, immobil zu sein. Für den Längsschnitt 1996 – 1999 erfolgt ein Abgleich aller in diesem Zeitraum immobilen Personen und aller Personen, die 1996 in der Stichprobe vertreten waren ohne Geburten, Sterbefälle, Zuzüge aus dem Ausland und Fortzüge ins Ausland. Analog werden die Mobilitätsfaktoren für die Längsschnitte 1996 – 1997 und 1996 – 1998 berechnet, hier werden alle immobilen Personen im betrachteten Längsschnitt allen Personen der Stichprobe von 1996 gegenübergestellt. Bei der Betrachtung der Mobilitätswahrscheinlichkeit versteht man unter immobil alle Personen, die während des gesamten Längsschnitts im Mikrozensus vertreten waren. Zu mobil zählen alle Personen, die zwar 1996 im Mikrozensus erhoben wurden jedoch beim letzten Interview des betrachteten Längsschnitts fehlten und nicht verstarben. Geburten, Zuzüge aus dem Ausland und Sterbefälle werden gesondert betrachtet und gehen nicht mit in die Berechnung des Mobilitätsfaktors ein. Dieser wird für alle drei Längsschnittperioden auf Haushaltsebene berechnet, so dass alle Personen innerhalb des gleichen Haushalts den gleichen Mobilitätsfaktor zugewiesen bekommen. Für die Beschreibung der abhängigen Variable Mobilität gehen mehrere haushaltsspezifische und personengebundene Merkmale als erklärende Variablen ein. Insgesamt werden folgende Merkmale als erklärende Variablen betrachtet:

- Haushaltsgröße  
1 Person, 2 Personen, 3 Personen, 4 Personen, 5 und mehr Personen
- Alter der Haushaltsmitglieder  
unter 30, zwischen 30 und 60 Jahren, älter als 60 Jahre
- Haushaltseinkommen  
unter 2200, 2200 bis unter 3000, 3000 bis unter 4000, 4000 bis unter 5500, 5500 und mehr, keine Angabe
- Region  
Ostdeutschland, Westdeutschland
- Schulabschluss
- beruflicher Abschluss
- Staatsangehörigkeit  
deutsch, nicht-deutsch

Die Berechnung der Mobilitätsfaktoren erfolgte im Rahmen des Projekts „Aufbereitung und Bereitstellung des Mikrozensus als Panelstichprobe“ als Teil der Arbeiten, die von den Mitarbeitern der Freien Universität Berlin durchgeführt wurden. Weitere Informationen zu den Faktoren und ihrem Vergleich zwischen Mikrozensus und Sozio-Ökonomischen Panel (SOEP) finden sich in [4].

### 3.6 Anpassung des Mikrozensus-Längsschnitts

Gemeinsam mit der Ziehungswahrscheinlichkeit  $\pi_k$  für jede Person  $k$  geht der geschätzte Mobilitätsfaktor  $\hat{\theta}_k$  in die Kalibrierung der Stichprobe an die Bevölkerungsdaten ein. Die Anpassung des Längsschnitts erfolgt parallel an die Bevölkerung von 1996 und an die Bevölkerung des im Längsschnitt zuletzt betrachteten Jahres. Für den Längsschnitt 1996 – 1999 bedeutet dies, es wird gleichzeitig an Angaben aus dem Jahr 1996 und von 1999 angepasst. Auf diese Weise können Änderungen in den betrachteten Merkmalsausprägungen berücksichtigt werden. Es 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 befragt, doch sie fehlen in der Menge der befragten Personen von 1999. Um dieses Defizit

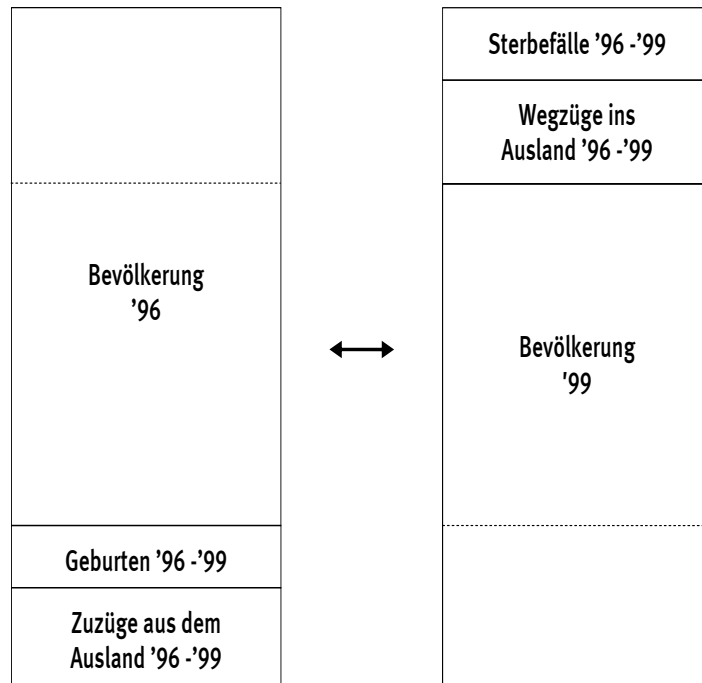


Abbildung 1: Definition der zwei betrachteten Mengen für die Randanpassung des Längsschnitts 1996 – 1999

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 1 abgebildet.

Nun erfolgt eine Randanpassung an beide Mengen gleichzeitig. Personen, die während des gesamten betrachteten Längsschnitts im Mikrozensus erhoben wurden, sind in den Mengen Bevölkerung '96 und Bevölkerung '99 vertreten. Geburten sind unter Geburten '96 - '99 und in der Menge Bevölkerung '99 zu finden. Analog gilt dies für Zuzüge aus dem Ausland. Sterbefälle und Fortzüge ins Ausland sind in der Menge Bevölkerung '96 und 1999 in der Menge Sterbefälle '96 - '99 bzw. Fortzüge ins Ausland '96 - '99 vertreten. Angepasst wird nun an jeweils zwei Vektoren gleichzeitig, die jedoch während der Hochrechnung zu einem Vektor  $\mathbf{t}_x$  zusammengefasst werden. Im

Folgenden jedoch sollen diese beiden Vektoren zunächst getrennt betrachtet werden. Der eine Vektor besteht aus den Bevölkerungsdaten von 1996, sie bilden die ersten Komponenten, und wird ergänzt durch zwei weitere Komponenten, die Anzahl der Geburten und die Anzahl der Zuzüge aus dem Ausland während des betrachteten Zeitraums. Wird an den Längsschnitt 1996 – 1999 angepasst, so besteht der zweite Eckwertvektor aus den Angaben der Bevölkerung von 1999 ergänzt um eine Komponente, die die Anzahl der Sterbefälle und Fortzüge ins Ausland zwischen 1996 und 1999 angibt. Wie bereits erwähnt, kann in der Stichprobe keine Trennung zwischen der Zahl der Sterbefälle und der Fortzüge ins Ausland vorgenommen werden. Dies führt zu einer Übergewichtung der Sterbefälle in der Stichprobe, da diese zusätzlich die Anzahl der Fortzüge repräsentieren.

### 3.7 Die Eckwerte der Randanpassung des Mikrozensus-Panels

Die Anzahl der Geburten stehen als monatliche Angaben zur Verfügung und können zusätzlich nach Geschlecht untergliedert werden. Ebenso stehen die Zahlen der Zuzüge aus dem Ausland nach Geschlecht gegliedert zur Verfügung, hier handelt es sich jedoch um jährliche Angaben. Da die Summe der Eckwertvektoren gleich sein muss, wird die Anzahl der Sterbefälle und Fortzüge ins Ausland über die Differenz des Eckwertvektors für 1996 und der Bevölkerungszahl des letzten Jahres berechnet. Zwar liegen zu den Sterbefällen und Fortzügen ins Ausland ebenfalls Zahlen aus der laufenden Bevölkerungsfortschreibung vor, doch da es sich bei den Bevölkerungszahlen jeweils um unterschiedliche Quellen handelt, kann es zu kleinen Abweichungen in der Summe kommen. Die Berechnung der Angaben zu Sterbefällen und Fortzügen ins Ausland sei hier am Beispiel 1996 – 1999 dargestellt:

$$N_{Weg9699} = (N_{Quer96} + N_{Geb9699} + N_{Zuzug9699}) - N_{Quer99} \quad (7)$$

mit:

$N_{Quer96}$ :	Bevölkerungszahl von 1996
$N_{Quer99}$ :	Bevölkerungszahl von 1999
$N_{Geb9699}$ :	Anzahl der Geburten zwischen Mai 1996 und April 1999
$N_{Zuzug9699}$ :	Anzahl der Zuzüge aus dem Ausland zwischen Mai 1996 und April 1999
$N_{Weg9699}$ :	Anzahl der Sterbefälle und Fortzüge ins Ausland zwischen Mai 1996 und April 1999

Sterbefälle und Wegzüge ins Ausland können aus Fallzahlgründen nicht nach Geschlecht unterteilt werden. Da die Angaben der Bevölkerungszahlen aus der laufenden Bevölkerungsfortschreibung differenziert nach beispie-

weise Alter und Geschlecht oder Staatsangehörigkeit vorliegen, können jedoch die Bevölkerungsangaben weiter untergliedert werden. In Anlehnung an die Hochrechnung im Querschnitt wird auch im Längsschnitt eine räumliche Unterteilung der Bevölkerungszahlen vorgenommen. Nicht in jedem Bundesland ist der durch Umzüge innerhalb Deutschlands entstehende Nonresponse gleichmäßig verteilt. Daher wird neben dem Alter und Geschlecht nach Bundesland differenziert angepasst. Betrachtet werden 8 Altersklassen, die in Kombination mit dem Geschlecht zu 16 Ausprägungen führen. Aufgeteilt auf die Bundesländer erhält man 256 Eckwerte, an die die Stichprobe angepasst wird. Aufgrund der Vielzahl an Eckwerten und der Fallzahlproblematik in der Stichprobe ist es nicht sinnvoll, die Kombination aus Alter und Geschlecht weiter regional zu unterteilen. Der räumliche Aspekt spielt jedoch bei der Betrachtung von Nonresponse eine große Rolle. Aus diesem Grund findet eine vereinfachte Anpassung auf Regierungsbezirksebene nach dem Geschlecht statt. Auch der Einfluss der Variable Staatsangehörigkeit ist bei der Nonresponse-Betrachtung nicht außer Acht zu lassen. Hier macht es ebenfalls Sinn, eine räumliche Unterteilung der Werte nach der Staatsangehörigkeit durchzuführen. Da die Fallzahlen der unterschiedlichen Nationalitäten zu gering sind, wird lediglich nach Deutschen und Nicht-Deutschen unterschieden. Für die Jahresübergänge 1996 – 1997 und 1996 – 1998 kann diese Staatsangehörigkeitsunterteilung auf Bundeslandebene vorgenommen werden. Da in der Stichprobe jedoch vergleichsweise wenig Ausländer aus ostdeutschen Bundesländern vertreten sind, wird die Anzahl der nicht-deutschen Personen für die ostdeutschen Länder zu einem Wert zusammengefasst. Für den Längsschnitt 1996 – 1999 ergibt sich hier eine größere Fallzahlproblematik, so dass für diesen Übergang nur an die Zahl der Deutschen und die Zahl der Ausländer angepasst wird. Insgesamt wird die Stichprobe an folgende acht Eckwertvektoren angepasst (hier für das Beispiel des Längsschnitts 1996 – 1998), die für die Hochrechnung zu dem Vektor  $\mathbf{t}_x$  zusammengefasst werden:

- Bundesland \* Alter \* Geschlecht '96, Geburten \* Geschlecht, Zuzüge aus dem Ausland \* Geschlecht
- Bundesland \* Alter \* Geschlecht '98, Sterbefälle und Fortzüge ins Ausland
- Regierungsbezirk \* Geschlecht '96, Geburten \* Geschlecht, Zuzüge aus dem Ausland \* Geschlecht
- Regierungsbezirk \* Geschlecht '98, Sterbefälle und Fortzüge ins Ausland
- Bundesland \* Staatsangehörigkeit (deutsch, nicht-deutsch) '96, Geburten \* Geschlecht, Zuzüge aus dem Ausland \* Geschlecht

- Bundesland \* Staatsangehörigkeit (deutsch, nicht-deutsch) '98, Sterbefälle und Fortzüge ins Ausland
- Eheschließungen, keine Eheschließungen
- Ehescheidungen, keine Ehescheidungen

Die Summe der hier beschriebenen Vektoren ist immer gleich. Der \*-Operator gibt an, dass hier die Kreuzkombination der Variablen gebildet wird.

Die Anzahl der Eheschließungen und Ehescheidungen gehen als Eckwerte in die Kalibrierung ein, um mögliche durch Nonresponse entstehende Verzerrungen zu verringern. Eheschließungen können im Mikrozensus-Längsschnittdatenmaterial über den Wechsel des Familienstandes von ledig auf verheiratet erfasst werden. Zusätzlich wird nach dem Eheschließungsjahr gefragt. Zwar ist die Beantwortung dieser Frage freiwillig, in den meisten Fällen wird jedoch in mindestens einem der vier Jahre zwischen 1996 und 1999 das Eheschließungsjahr angegeben. Somit ist die Codierung des Ereignisses Eheschließung problemlos möglich. Ehescheidungen dagegen werden nicht direkt erfragt. Sie können nur über den Wechsel des Familienstandes von verheiratet zu geschieden beobachtet werden. Alle anderen betrachteten Variablen werden direkt erfragt und somit ist eine entsprechende Zuordnung der Eckwerte kein Problem.

Bei der Konstruktion der Eckwerte ist jedoch zu bedenken, dass diese stets für die Gesamtbevölkerung Deutschlands erhoben werden. Da das Mikrozensuslängsschnittmaterial nur aus Personen in Privathaushalten besteht, ist die Bevölkerungszahl entsprechend anzupassen. Im Mikrozensusquerschnittmaterial stehen für alle vier Jahre die Hochrechnungsfaktoren für die Gesamtbevölkerung Deutschlands zur Verfügung. Die Zahl der Personen in Gemeinschaftsunterkünften kann somit anhand des Hochrechnungsfaktors für die Grundgesamtheit berechnet 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 von jedem Eckwerte die entsprechende Zahl der Personen in Gemeinschaftsunterkünften abgezogen werden. Bei den Eheschließungen, Ehescheidungen, Zuzügen aus dem Ausland und Sterbefällen 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 zwar

nicht vernachlässigbar, da die Anzahl der Sterbefälle und Fortzüge ins Ausland jedoch als Differenz der anderen Eckwerte erhoben wird, vgl. Gleichung 7, erhält man auch hier näherungsweise die Anzahl in Privathaushalten.

Ein weiteres Problem bei den Eckwerten besteht darin, dass nicht alle Daten monatlich erhoben werden. Teilweise stehen, wie zum Beispiel für Ehescheidungen, nur jährliche Angaben aus der Bevölkerungsforschung 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  $\frac{2}{3}$  gewichtet, die aus dem aktuellen Jahr mit  $\frac{1}{3}$ . 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.

## 4 Die Hochrechnungsfaktoren

Die berechneten Hochrechnungsgewichte berücksichtigen die von der Freien Universität Berlin erarbeiteten Mobilitätsfaktoren ebenso wie die Veränderungen der Bevölkerung innerhalb des Längsschnittzeitraums. Da in der Stichprobe Zuzüge aus dem Ausland unterrepräsentiert sind, erhalten diese Personen als Ergebnis der Hochrechnung ein deutlich höheres Gewicht als andere Personen, die im betrachteten Zeitraum räumlich immobil waren. Die männlichen Zuzüge sind im Vergleich zu den weiblichen weniger in der Stichprobe vertreten. Daher fallen die Hochrechnungsfaktoren der männlichen Zuzüge deutlich höher aus als die der Frauen.

Auch Sterbefälle treten im Vergleich zur Grundgesamtheit weniger häufig in der Stichprobe auf. Personen, die während des betrachteten Zeitraums des Längsschnitts versterben, erhalten somit ein vergleichsweise höheres Gewicht als andere Personen. Dieses fällt jedoch nicht ganz so hoch aus wie bei den männlichen Zuzügen.

Die Ereignisse Eheschließung oder Ehescheidung führen ebenfalls zu einem höheren Gewicht, doch diese Erhöhung fällt im Vergleich zu der der Zuzüge und Sterbefälle gering aus. Tritt jedoch das Ereignis Ehescheidung in Verbindung mit einem Zuzug auf, so vergrößert sich der Hochrechnungsfaktor der jeweiligen Person im Vergleich zu einer Scheidung ohne einen Zuzug aus dem Ausland. Dies gilt auch für Eheschließungen in Verbindung mit einem Zuzug aus dem Ausland.

Bei der Staatsangehörigkeit fällt auf, dass besonders Ausländer im Osten unterrepräsentiert sind. Dies führt zu einer Übergewichtung der in der Stichprobe vertretenen Ausländer in den ostdeutschen Bundesländern. Diesen Personen wird ein nicht ganz so hohes Gewicht zugeordnet wie Zuzügen aus dem Ausland und Sterbefällen, sie erhalten jedoch einen deutlich höheren Hochrechnungsfaktor als die restlichen Stichprobenpersonen.

Generell ist bei der Verwendung der Hochrechnungsfaktoren für Längsschnittanalysen die Abgrenzung der Stichprobe zu beachten. Diese kann von der für die Hochrechnung getroffenen abweichen, da für viele Analysen Geburten, Sterbefälle, Zuzüge aus dem Ausland und Fortzüge ins Ausland nicht mitberücksichtigt werden. Dennoch kann in den meisten Fällen auf die Gewichtungsfaktoren zurückgegriffen werden. Um Auswertungen zu erleichtern, wurden drei Variablen generiert, die für jeden Datensatz angeben, ob dieser für die Hochrechnung des betrachteten Längsschnitts herangezogen wurde und an welchen Eckwert er angepasst wurde. Diese Variablen sind mit *panel\_9697*, *panel\_9698* und *panel\_9699* bezeichnet, die genaue Codierung sei hier am Beispiel für *panel\_9699* beschrieben:

<i>panel_9699</i>	= 11	falls die Person 1996 – 1999 in der Stichprobe war
	= 12	falls die Person 1996 in der Stichprobe war und vor 1999 verstarb oder ins Ausland verzog
	= 21	falls die Person nach April 1996 geboren wurde und 1999 in der Stichprobe war
	= 31	falls die Person nach April 1996 aus dem Ausland zuzog und 1999 in der Stichprobe war
	= .	falls die Person für die Hochrechnung nicht berücksichtigt wurde

Eine Person wird für die Hochrechnung nicht berücksichtigt, wenn sie im betrachteten Zeitraum innerhalb Deutschlands umgezogen ist und somit in mindestens einem Jahr nicht weiter befragt werden konnte.

Anhand dieser Variablen ist es möglich, die verwendeten Eckwerte zu ermitteln. Die Bevölkerungszahl von 1996 erhält man beispielsweise, wenn man alle Hochrechnungsfaktoren der Personen, für die *panel\_9699* gleich 11



oder gleich 12 ist, aufsummiert. Die Summe der Gewichtungsfaktoren der Personen mit *panel\_9699* gleich 11, 21 oder gleich 31 ergibt die Bevölkerungszahl von 1999. Auch die Unterteilung nach beispielsweise Bundesland \* Alter \* Geschlecht lässt sich so ermitteln, hierzu ist lediglich die Summe über diese Kombination zu bilden.

Weiter erklären die Variablen *panel\_9697*, *panel\_9698* und *panel\_9699* auch, wieso für viele Auswertungen die Bevölkerungszahlen nicht getroffen werden können. Ist man beispielsweise an der Anzahl der Personen interessiert, die zwischen 1996 und 1999 ihren Erwerbsstatus wechselten, so können für diese Analyse nur Personen berücksichtigt werden, die in beiden Jahren in der Stichprobe vertreten waren. Sterbefälle und Fortzüge ins Ausland werden nicht mitberücksichtigt und somit fehlt deren Anzahl in der Summe der Gewichtungsfaktoren für die Zahl der Personen in der Grundgesamtheit von 1996. Die Grundgesamtheit für die Untersuchung der Übergänge stellt eine Teilmenge der Grundgesamtheit für die Hochrechnung dar. Ebenso bildet die Menge der Stichprobenpersonen der Analyse eine Teilmenge der Stichprobe für die Hochrechnung. Die Verwendung der Hochrechnungsfaktoren für diese Teilmengen repräsentiert dennoch die Grundgesamtheit für diese Untersuchung. Die Hochrechnungsfaktoren können unter Berücksichtigung der richtigen Abgrenzung der Mengen problemlos für die Analyse verwendet werden. Soll für eine Analyse die Gesamtbevölkerung Deutschlands zu dem betrachteten Zeitraum herangezogen werden und ist man an der Anzahl der Personen mit einem Wechsel in der Analysevariable interessiert, so ist die Verwendung der Hochrechnungsfaktoren problemlos möglich. Analysen wie etwa: „Wie viele Personen studieren am Wohnort der Eltern?“ oder „Bei wie vielen Personen fand ein Wechsel zwischen erwerbstätig und erwerbslos innerhalb des letzten Jahres statt?“, können anhand der Abgrenzung nach den Personen mit *panel\_96xx* gleich 11 durchgeführt werden. Umzüge innerhalb Deutschlands werden hier durch die in der Stichprobe verbleibenden Personen mitberücksichtigt, so dass sich die Summe der Hochrechnungsfaktoren der Personen mit den jeweiligen Übergänge in den betrachteten Analysevariablen immer auf die Anzahl der Personen in der Grundgesamtheit bezieht.

Dagegen sind keine Analysen möglich, die den Verlauf einer Analysevariablen bei räumlich mobilen Personen untersuchen. So können beispielsweise folgende Fragestellungen anhand des vorliegenden Mikrozensuslängsschnittmaterials nicht analysiert werden: „Findet ein Wohnortwechsel aufgrund einer neuen Beschäftigung statt?“, „Findet ein Auszug aus dem Elternhaus aufgrund eines Studiums in einer anderen Stadt statt?“ Um diese Fragestellungen untersuchen zu können, wäre eine Weiterverfolgung der räumlich mobilen Personen nötig, die im Mikrozensus jedoch nicht erfolgt. Für spezielle Untersuchungen auf nur der Menge der räumlich mobilen Personen oder nur der Menge der räumlich immobilen Personen können die be-

rechneten Hochrechnungsfaktoren nicht genutzt werden, sie repräsentieren stets sowohl räumlich mobile als auch räumlich immobile Personen innerhalb Deutschlands. Eine Differenzierung nach Mobilität kann nicht vorgenommen werden.

Für Analysen ist weiter wichtig zu beachten, dass im Rahmen dieses Projektes nur drei Längsschnitte 1996 – 1997, 1996 – 1998 und 1996 – 1999 betrachtet werden. Untersuchungen des Übergangs 1997 – 1998 sind auf diesem Datenmaterial nur bedingt möglich. Es ist nicht möglich zu untersuchen, wer 1996 nicht befragt wurde, aber zwischen 1997 – 1998 einen bestimmten Übergang miterlebte, beispielsweise einen Wechsel im Erwerbsstatus. Ist man dagegen an Personen interessiert, die 1996 befragt wurden und 1997 – 1998 ihren Erwerbsstatus wechselten, so sind diese Analysen möglich. Zusammenfassend kann gesagt werden, dass Personen, die keinen Hochrechnungsfaktor für den betrachteten Zeitraum zugewiesen bekommen haben, für Längsschnittanalysen nicht herangezogen werden können.

## 5 Schlussbemerkung

Die Hochrechnung des Mikrozensusdatenmaterials für Längsschnittuntersuchungen mittels einer Anpassung an Daten aus der laufenden Bevölkerungsfortschreibung ist möglich. Aufgrund des umfangreichen Frageprogramms kann bei der Hochrechnung an eine Vielzahl von Variablen angepasst werden. Hier sind jedoch die Fallzahlen und vorliegende Daten zur Bevölkerungsgrundgesamtheit zu beachten. Anhand der Hochrechnung kann nicht nur die Grundgesamtheit eines Zeitpunktes wiedergegeben werden, sondern sie spiegelt stets die Grundgesamtheit des Anfang- und Endjahres der betrachteten Zeitspanne wieder und berücksichtigt zusätzlich Informationen über Veränderungen in der Bevölkerungszusammensetzung im Zeitverlauf.

Zusammenfassend kann gesagt werden, dass mit dem hier beschriebenen Hochrechnungsverfahren bereits gute Ergebnisse erzielt werden. Es ist möglich, für die Mehrheit der Stichprobenpersonen einen Hochrechnungsfaktor zu bestimmen, der bei Analysen gute Ergebnisse erzielt. So werden im Rahmen des Projekts von den anderen Projektteilnehmern vergleichende Arbeiten zwischen dem hochgerechneten Längsschnitt und dem SOEP sowie Bevölkerungsdaten durchgeführt. Dennoch wären auch Alternativen des Hochrechnungsverfahrens möglich. Beispielsweise könnte das Längsschnittmaterial im Rahmen einer zweistufigen Anpassung an den vorliegenden Querschnitt und an die Daten aus der laufenden Bevölkerungsfortschreibung angepasst werden. Dies hätte den Vorteil, dass für Anpassungen ein weites Spektrum an Variablen aus dem Mikrozensusfrageprogramm des Querschnitts zur Verfügung steht. Zu beachten sind hier jedoch die Veränderun-

gen der Bevölkerung im Zeitverlauf. Diese erschweren eine genaue Abgrenzung der Populationen. In wie weit die Informationen aus dem Querschnitt dennoch für die Hochrechnung genutzt werden können, wird zur Zeit noch von den Mitarbeitern der Freien Universität Berlin untersucht.

Generell könnte die Qualität der Hochrechnung jedoch noch durch eine bessere Datenqualität verbessert werden. Hierzu ist eine Plausibilitätsprüfung im Längsschnitt notwendig. Sie ergänzt die Plausibilisierung im Querschnitt und achtet auf konsistente Angaben über den Zeitverlauf. Beispielsweise sollten Angaben wie das Alter und das Geschlecht einer Person konsistent sein. Auch die Angaben zur Staatsangehörigkeit, zum Familienstand und zur Bildung sollten plausibel sein. Ein Wechsel der Nationalität ist zwar durchaus möglich, er sollte jedoch nur maximal einmal auftreten. Dagegen sind bestimmte Wechsel im Schulabschluss nicht möglich. Wird im ersten Jahr eine allgemeine Hochschulreife als höchster Schulabschluss angegeben, so kann im nächsten Jahr nicht mittlere Reife angegeben werden. Generell würde eine Plausibilisierung des Längsschnitts die Arbeiten zur Hochrechnung erleichtern. Auch die Qualität und Zuverlässigkeit der Analysen würde zunehmen.

Zur Verbesserung der Hochrechnungsfaktoren können auch zeitnahe Bevölkerungsdaten beitragen. Diese sollten nicht nur zeitnah erhoben sein, sondern sich auf die gleiche Grundgesamtheit wie die Stichprobe beziehen. Somit könnten mögliche Rundungsfehler durch die lineare Interpolation und die Subtraktion der hochgerechneten Gemeinschaftsunterkünfte vermieden werden. In Hinblick auf Auswertungen für die Gesamtbevölkerung Deutschlands wäre es wünschenswert Gemeinschaftsunterkünfte in das Längsschnittfile zu integrieren.

## Literatur

- [1] Hans-Joachim Heidenreich, Robert Herter-Eschweiler: *Längsschnittdaten aus dem Mikrozensus: Basis für neue Analysemöglichkeiten*, Wirtschaft und Statistik, Heft 8/2002
- [2] Carl-Erik Särndal, Bengt Swensson, Jan Wretman: *Model Assisted Survey Sampling*, Springer-Verlag New-York, Inc. (1992)
- [3] Paul Lüttinger, Thomas Riede: *Der Mikrozensus - Amtliche Daten für die Sozialforschung*, ZUMA-Nachrichten (11/1997)
- [4] Edin Basic, Ivo Marek, Ulrich Rendtel: *The German Microcensus as a tool for longitudinal data analysis: An evaluation using SOEP data*, Arbeitspapier #3 ([http://www.destatis.de/mv/mzpanel\\_ap.htm](http://www.destatis.de/mv/mzpanel_ap.htm))