

Möglichkeiten der Verknüpfung von Mikrozensus-Querschnitterhebungen ab 2012 zu Panels

Januar 2018

Robert Herter-Eschweiler (Statistisches Bundesamt, H 306 Auswertung und Analyse der Haushaltserhebungen, Bonn)
Bernhard Schimpl-Neimanns (GESIS – Leibniz-Institut für Sozialwissenschaften, German Microdata Lab, Mannheim)

Einleitung

Im Rahmen des Methodenverbundes „Aufbereitung und Bereitstellung des Mikrozensus als Panelstichprobe“, ein vom Bundesministerium für Bildung und Forschung sowie der Deutschen Forschungsgemeinschaft geförderten Projektes zur Nutzung des Mikrozensus (MZ) ab 1996 als Panelstichprobe, wurden Vorschläge zur Bereitstellung von Scientific-Use-Files (SUF) erstellt, die es der Wissenschaft erlauben selbstständig aus den SUF der einzelnen Jahre Panels zu erstellen (siehe <https://www.destatis.de/DE/Methoden/Methodenpapiere/Mikrozensus/Mikrozensus.html>).

Das ab dem MZ 2012 geltende Konzept der Erstellung der SUF basiert weitgehend auf diesen Überlegungen des Methodenverbunds (siehe Schimpl-Neimanns 2006). Einschränkend ist zu erwähnen, dass ab 2016 eine neue MZ-Stichprobe mit der Auswahlgrundlage des Zensus 2011 gezogen wurde (siehe Bihler & Zimmermann 2016), sodass Verknüpfungen der Querschnittsdaten ab dem Erhebungszeitpunkt 2012 nur bis einschließlich der Erhebung 2015 möglich sind. Des Weiteren sind Verknüpfungen der Querschnitte ab 2016 auf den Zeitraum bis einschließlich 2019 beschränkt, da ab 2020 der MZ und weitere Haushaltsstatistiken auf ein grundlegend neues System umgestellt werden (siehe https://www.destatis.de/DE/Methoden/Rechtsgrundlagen/Statistikbereiche/Inhalte/1062_MZG.pdf?blob=publicationFile).

Wesentlich für Panelauswertungen sind der Stichprobenplan und das Konzept der partiellen Rotation der Auswahlbezirke im MZ sowie die damit verbundene Frage der Ziehung von Substichproben des SUF. Eine weitere elementare Voraussetzung für die Zusammenführung von Beobachtungs- und Stichprobeneinheiten zu einem Rotationspanel sind zeitkonsistente Identifikatoren (Ordnungsnummern). Die Angaben der befragten Personen beziehen sich in der Regel auf die der Befragung vorangegangene Woche (Berichtswoche). Aufgrund von Problemen der Erreichbarkeit kommt es aber zu zeitlichen Abweichungen der realisierten Interviews von den Befragungszeitpunkten gemäß Stichprobenplan. Es ist deshalb zu überlegen, wie bei der Zusammenführung mit den sogenannten „Jahresüberhängen“ zu verfahren ist.

Im Folgenden werden diese Punkte beschrieben und es wird gezeigt, wie die ab dem MZ SUF 2012 vorliegenden Daten zu einem Panel zusammengeführt werden können. Abschließend werden weitere methodische Fragen zur Panelmortalität angesprochen.

1. Stichprobendesign

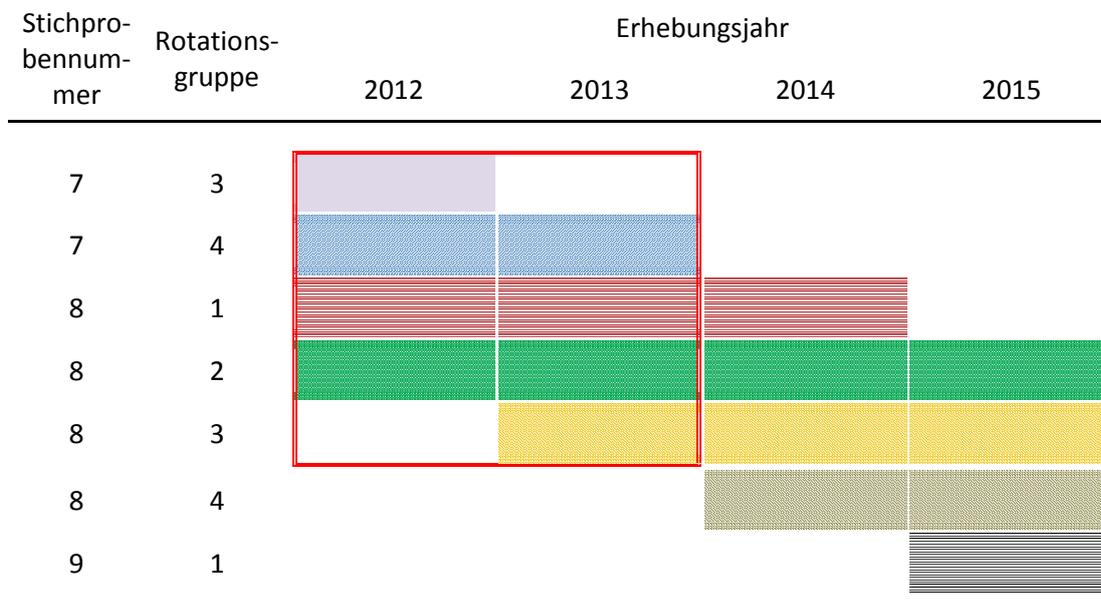
1.1 Rotationsschema

Der MZ ist eine einstufige Klumpenstichprobe, wobei die geografisch und fachlich abgegrenzten Flächenstücke (z. B. Straßenteil, Gebäude bzw. Gebäudeteile) benachbarte Wohnungen umfassen. Sie dienen als Interviewerbezirke und bilden die Primäreinheiten (Auswahlbezirke) bei der Stichprobenziehung.

Die Aufgabe des MZ besteht nicht nur in der Bereitstellung von Daten zur Bevölkerungsstruktur und zur sozialen und wirtschaftlichen Lage der Bevölkerung im Querschnitt, sondern es sollen auch möglichst präzise Daten über Strukturveränderungen ermittelt werden. Hierfür waren zwar bereits früh Wiederholungsbefragungen geplant (Deiningner 1960; Herberger 1973), wurden jedoch erst ab 1962 eingeführt.

Die Anforderungen an Längsschnittauswertungen werden ab 1977 im Stichprobenplan durch die partielle Rotation von Erhebungseinheiten wie folgt berücksichtigt. Die Auswahlbezirke und die darin wohnenden Haushalte und Personen verbleiben vier Jahre lang in der Befragung und bilden ein sogenanntes Rotationsviertel. Jährlich wird ein Viertel der Auswahlbezirke ausgetauscht. Der MZ ist somit eine Wiederholungsbefragung mit teilweiser Überlappung der Erhebungseinheiten. Abhängig vom Grad der Überlappung und der Korrelation der interessierenden Schätzwerte können sich bei Vergleichen von Querschnittsergebnissen (Netto-Veränderungen) deutliche Reduktionen des Stichprobenfehlers der Netto-Veränderungen gegenüber unabhängigen Stichproben ergeben.

Abb. 1: Rotation im Mikrozensus



Wie Abb. 1 zeigt, kann mit den Daten der Erhebungsjahre 2012 und 2013 ein Panel mit drei in beiden Jahren befragten Rotationsvierteln gebildet werden (Stichprobennummer \circ Rotationsgruppe: $7^{\circ}4$, $8^{\circ}1$ und $8^{\circ}2$). Infolge der Anlage des MZ als Flächenstichprobe werden die aus dem Auswahlbezirk wegziehenden Haushalte und Personen nicht weiter befragt, sondern durch die nachziehenden Haushalte und Personen ersetzt. Deshalb wird der Anteil identischer Auswahlbezirke von 75 Prozent nicht auf der Haushalts- und Personenebene erreicht. Zwar bietet dieses Verfahren der Ersetzung fortziehender durch zuziehende Einheiten den Vorteil der automatischen Erfassung von Populationsveränderungen. Für fortziehende Einheiten fehlen jedoch Angaben nach dem Fortzug und für zuziehende Einheiten liegen – bis

auf wenige Retrospektivangaben – keine Informationen vor dem Zuzug vor. Außerdem ist festzuhalten, dass zu den Merkmalen der im vierjährigen Abstand durchgeführten Zusatzerhebungen aufgrund des gegenwärtigen Rotationsschemas keine Panelangaben vorliegen.

1.2 Substichprobenziehung

Die Nutzbarkeit der Paneleigenschaft ist eng mit den Auswahlbezirken verbunden. Deshalb wurde die bis zum SUF 2011 angewendete Ziehung einer 70 %-Haushaltssubstichprobe ab 2012 durch ein neues Verfahren abgelöst. Die Substichprobenziehung orientiert sich weitgehend am Originalstichprobenplan, erweitert diesen aber um zwei weitere Aspekte, die Größe der Auswahlbezirke sowie die Kennung eines Auswahlbezirks zum Ad-hoc Modul der Arbeitskräfteerhebung.

In einem ersten Schritt werden die Auswahlbezirke (AWB) der Ausgangs-/Originaldaten sortiert nach:

1. Bundesland
2. Größenklasse der AWB (Dezile)
3. Grundausswahl/Aktualisierung der AWB
4. Unterstichprobenkennung (Ad-hoc-Modul)
5. NUTS-2 Raumeinheit
6. Regionale Anpassungsschicht
7. Regionale Schicht
8. Regionale Schichtuntergruppe
9. Gemeindegrößenklasse
10. Nummer der AWB

Zwar wurde bei der Grundausswahl bzw. wird bei der jährlichen Aktualisierung der Auswahlbezirke eine einheitliche Größe der Gebäudegrößenklassen angestrebt, sie unterscheiden sich jedoch fluktuationsbedingt erheblich. Insbesondere die Auswahlbezirke der Grundausswahl, deren Gebäudegrößenklassen auf der Grundlage der Volkszählungsdaten 1987 basieren, weisen mittlerweile zum Teil erhebliche Schwankungen auf. Um eine daraus resultierende Erhöhung der Varianz des Stichprobenfehlers zu begrenzen, erfolgte die zusätzliche Aufnahme einer Sortierung nach Größenklassen der Auswahlbezirke. Die Messung der Größenklassen erfolgte durch die Zahl der Personen in Privathaushalten bzw. Gemeinschaftsunterkünften in den Auswahlbezirken. Die jeweils getrennte Schichtung von Auswahlbezirken mit Privathaushalten und Auswahlbezirken mit Gemeinschaftsunterkünften wurde vorgenommen, um die unterschiedliche Struktur der jeweiligen Personengemeinschaft berücksichtigen zu können.

Mit der Aufnahme der Kennung über den Einbezug eines Auswahlbezirks in die Unterstichprobe soll die Unterstichprobeneigenschaft in der Substichprobenziehung kontrolliert werden. Beide Merkmale (Größenklasse und Unterstichprobenkennung eines Auswahlbezirks) wurden für so wichtig gehalten, dass sie in der Sortierung nach dem Bundesland an zweiter bzw. dritter Stelle platziert wurden.

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

Es wird ein einheitlicher Auswahlsatz von 70 % in allen Schichten angestrebt. Dazu werden in jeder Schicht mit Hilfe eines „Pseudozufallszahlengenerators“ sieben unterschiedliche Zahlen im Intervall a ($1 \leq a \leq 10$, $a \in \mathbb{N}$) gezogen. Jeder Auswahlbezirk innerhalb einer Schicht mit gezogener Zufallszahl gilt als ausgewählt. Bei den „unvollständigen“ Schichten gelten diejenigen Auswahlbezirke als gezogen, deren laufende Nummern innerhalb der Schicht sich unter den sieben zufällig aus den Zahlen eins bis zehn gezogenen Zahlen befindet. Der Stichprobenumfang einer solchen Schicht hängt also vom Zufall ab, die Ziehungswahrscheinlichkeit von 70 % bleibt aber erhalten.

2. System der Ordnungsnummern

Die Daten des MZ sind hierarchisch strukturiert, sodass die Erhebungseinheiten (Auswahlbezirke) und Analyseeinheiten (Wohnungen, Haushalte, Familien bzw. Lebensformen, Personen) aufeinander beziehbar und eindeutig abgegrenzt sind. D. h., alle einer Person übergeordneten Analyseeinheiten sind durch Identifikatoren (Ordnungsnummern) eindeutig definiert. Aus Datenschutzgründen müssen die Anordnung der Datensätze und somit die Ordnungsnummern systemfrei sein.

Für die Zusammenführung liegen folgende längsschnittkonsistente Identifikatoren (Variablen: EF „Eingabefeld“) vor:

| | |
|------|---|
| EF1 | Bundesland |
| EF3 | Nr. des Auswahlbezirks (systemfrei) |
| EF4 | Nr. des Haushalts im Auswahlbezirk (systemfrei) |
| EF63 | Personennummer |

Die Variable EF63 enthält im Unterschied zur Variable EF5a „Nr. der Person im Haushalt (systemfrei)“ eine über alle vier Berichtsjahre feste Personennummer. Sie wird ab dem Erhebungsjahr 2011 von den statistischen Landesämtern vergeben. Für die Ebene von Gebäuden (EF8 Nr. des Gebäudes im Auswahlbezirk (systemfrei)) und Wohnungen (EF9 Nr. der Wohnung im Auswahlbezirk (systemfrei)) sind die Identifikatoren zurzeit nicht längsschnittkonsistent¹.

Wie einleitend erwähnt, kommt es in der Praxis aufgrund von Problemen der Erreichbarkeit vor, dass die gemäß Stichprobenplan in einem bestimmten Zeitraum durchzuführenden Interviews erst später realisiert werden. Informationen dazu liegen in diesen Variablen vor:

| | |
|------|--------------------------------|
| EF5b | Jahresüberhang aus dem Vorjahr |
| EF12 | Berichtsquartal |

Zum Beispiel können bei der Zusammenführung der Querschnittsdaten 2012 und 2013 bis zu drei Beobachtungen für eine Befragungseinheit vorliegen: 2012: Jahresüberhang 2011; 2013: Jahresüberhang 2012 und reguläre Befragung 2013. Bei den Querschnittsauswertungen werden die Angaben aus den Jahresüberhängen wie Angaben eines regulären Interviews behandelt. Da von einer hohen Strukturgleichheit der Merkmale der in den einzelnen Jahren erst verspätet befragten Personen ausgegangen werden kann, ist das Vorgehen gerechtfertigt (Statistisches Bundesamt 2012: 2). Außerdem gilt auch bei Jahresüberhängen das Prinzip der Berichtswoche, d. h. die Angaben beziehen sich in der Regel auf die der Befragung vorangegangene Woche (Berichtswoche). Aus diesen Gründen und für Analysen, die eng mit den jeweiligen Querschnittsdaten verbunden sind, wäre somit eine Datenverknüpfung zweckmäßig, die sich eng an der Struktur der Querschnittsdaten orientiert. Beispielsweise betrifft dies die Untersu-

¹ Diese Identifikatoren sind nur in den Jahren sachgerecht verwendbar, in denen das 4-jährige Zusatzprogramm zur Wohnsituation durchgeführt wird.

chung von Differenzen von Schätzern (Nettoveränderungen) und die Kumulation von Daten mehrerer Erhebungszeitpunkte.

Dabei bleibt jedoch außer Acht, dass in den Angaben der Jahresüberhänge Angaben zum Vorjahr vorliegen, die den Unit-Nonresponse des Vorjahres reduzieren. Alternativ zum querschnittsorientierten Vorgehen können Datensätze der Jahresüberhänge dem Erhebungsjahr gemäß Stichprobenplan zugeordnet werden (Statistisches Bundesamt 2012: 7). Für die meisten Längsschnittanalysen dürfte diese Datenverknüpfung vorteilhaft sein. Damit werden Probleme potenzieller „Doppelerhebungen“ reduziert. Ergebnisse von Panel- und Querschnittsanalysen können sich aber unterscheiden. Auch fehlen für die jüngste Erhebung die erst im darauf folgenden Jahr verspätet realisierten Interviews (Jahresüberhänge). Des Weiteren wird das Prinzip der Berichtswoche verletzt, indem angenommen wird, dass die zum Zeitpunkt des verspäteten Interviews (t+1) erfolgten Angaben zum Zeitpunkt t erfolgt sind.

Welches Verfahren präferiert wird dürfte von den jeweiligen Analysezielen abhängen. Gesucht sind deshalb möglichst flexible und einfach umzusetzende Verfahren, wie Jahresüberhänge im Panel behandelt werden können. Im Folgenden werden zuerst die längsschnittorientierte und anschließend die querschnittsorientierte Version behandelt.

3. Panelbildung

3.1 Version 1 (Längsschnittorientierung)

In dieser Version werden die Angaben der Jahresüberhänge aus der Erhebung t+1 als Ersatz für die gemäß Stichprobenplan im Vorjahr ausgefallenen Interviews dem Panelzeitpunkt t zugeordnet. Für die Zusammenführung zweier Erhebungsjahre zu einem Panel sind weitere Vorbereitungen nötig. Zunächst werden die Befragungseinheiten pro Querschnittserhebung und dem zu bildenden Panelzeitpunkt separiert.

Die erste Datei enthält nur die Befragungseinheiten der Erhebung 2012 ohne Jahresüberhang (t = 1). Da Angaben der Befragungseinheiten der Erhebung 2012 mit Jahresüberhang gemäß Stichprobenplan zum Erhebungszeitpunkt 2011 (t - 1) gehören, werden sie für die Panelbildung ignoriert.

Die zweite (Teil-) Datei enthält die Befragungseinheiten der Erhebung 2013 mit Jahresüberhang, die gemäß Stichprobenplan zu 2012 (t = 1) zählen.

Die dritte (Teil-) Datei enthält Befragungseinheiten der Erhebung 2013 ohne Jahresüberhang (t = 2).

Für die Panelbildung werden die folgenden Ordnungsnummern benötigt:

- EF1 Bundesland
- EF3 Nr. des Auswahlbezirks (systemfrei)
- EF4 Nr. des Haushalts im Auswahlbezirk (systemfrei)
- EF63 Personennummer

Sie können zu einem alphanumerischen Personenidentifikator (z. B. „idpers“) verkettet werden. Die drei logischen Dateien bzw. die ausgewählten Daten aus den beiden Querschnittserhebungen 2012 und 2013 lassen sich mittels Personenidentifikator zusammenführen.

In nachfolgender Tabelle wird das Ergebnis dieser Panelbildung dargestellt. Ausgehend von den in den jeweiligen Erhebungsjahren vorliegenden Personenangaben zeigt sich, dass nach den Umbuchungen der Jahresüberhänge im Jahr 2012 die Anzahl der Personensätze im Vergleich zur Ausgangsbasis steigt und

im Jahr 2013 sinkt. Die Ursache hierfür liegt in den zur Verfügung stehenden Angaben zu den Jahresüberhängen. Werden in den Daten aus 2012 Jahresüberhänge aus 2011 entfernt und mit den Jahresüberhängen aus den Daten 2013 angereichert, kann in den Daten 2013 nur eine Rückbuchung der Jahresüberhänge aus 2012 umgesetzt werden. Die Informationen zu den Jahresüberhängen der Stichprobe von 2013 liegen erst im MZ 2014 vor. Mittels des Personenidentifikators lassen sich auf diese Weise 314 334 Personensätze zusammenführen. Die Zusammenführungsquote liegt damit bei 64,9 Prozent für die Daten aus 2012 bzw. 67,7 Prozent für die Daten aus 2013.

Tabelle 1: Zusammenführungsergebnisse – längsschnittorientiert

| | SUF 2012 | | SUF 2013 | |
|-------------------------------|----------|-------|----------|------|
| | n | % | n | % |
| Personensätze | | | | |
| Insgesamt | 482 127 | 100 | 478 176 | 100 |
| darunter nach „Umbuchungen“ | 484 412 | 100,5 | 464 065 | 97,0 |
| darunter nach Zusammenführung | 314 334 | 64,9 | 314 334 | 67,7 |

Die theoretisch zu erwartende Zusammenführungsquote von 75 Prozent (siehe Abb. 1) wird aufgrund der Bevölkerungsveränderung innerhalb der Auswahlbezirke (Fort- und Zuzüge, Geburten sowie Sterbefälle) und der ungleichen Möglichkeit, Jahresüberhänge zu berücksichtigen nicht erreicht.

Um die Konsistenz der Paneldaten zu prüfen und ggf. Daten zu korrigieren ist es sinnvoll, die Daten zunächst untereinander (long format) zu stellen. Pro Ausprägung des Personenidentifikators (idpers) und Panelzeitpunkt sollte maximal ein Personensatz vorliegen. Dies ist allerdings nicht immer gewährleistet. Zum Beispiel enthalten die Daten auch Sätze ausgefallener Befragungseinheiten im Auswahlbezirk, die für Ausfallanalysen hilfreich sein können. Diese Ausfallsätze sind in EF30 „Bevölkerung: Haupt- oder Nebenwohnsitz“ sowie weiteren Variablen mit -7 „Auswahlbezirk ohne befragten Haushalt“ und -8 „Entfällt (leerstehende Wohnung, ausgefallene Privathaushalte und Gemeinschaftsunterkünfte)“ codiert. Da diese Codes teilweise auch in der Variable EF5b „Jahresüberhang aus dem Vorjahr“ vorliegen, ist nicht sicher, ob sie ggf. zu Sätzen des Jahresüberhangs zählen. Diese wären bei den Querschnittsdaten 2012 gemäß Stichprobenplan bei der Panelbildung auszuschließen. Doppelte Sätze pro Personenidentifikator und Panelzeitpunkt sind vor weiteren Auswertungen zu bereinigen oder auszuschließen.

In Bezug auf Längsschnittauswertungen ist einschränkend auf verschiedene Punkte hinzuweisen. Im Fall von Jahresüberhängen lässt die Angabe zum Berichtsquartal (EF12) keine Aussagen über den zeitlichen Abstand zwischen zwei Befragungen zu. Im Unterschied zu dem früheren MZ-Panel 1996-1999 liegen für die Hochrechnung keine Längsschnittgewichte, sondern lediglich die Hochrechnungsfaktoren der Querschnittsdaten vor; siehe Kapitel 4-5 im Datenhandbuch zum MZ-Panel 1996-1999 (Statistisches Bundesamt 2006) und Bašić & Rendtel (2005) zu Fragen zur Konstruktion von Längsschnittgewichten.

In allen Panels treten zeitlich inkonsistente Angaben auf. Eine erste einfache Prüfung der vorliegenden Daten wird durch die Auswertung zeitlich übereinstimmender Angaben zum – im Prinzip – zeitkonstanten Geschlecht und Geburtsjahr möglich (siehe Tabelle 2). Für die längsschnittorientierte Panelbildung zeigt sich bei dem Geschlecht eine Inkonsistenz von 0,3 Prozent und bei dem Geburtsjahr von 2,7 Prozent. Eine nähere Betrachtung ergibt, dass inkonsistente Angaben zum Geburtsjahr häufig auf „Zahldreher“ zurückzuführen sind. Insgesamt deutet das Zusammenführungsergebnis zeitlich konstanter Angaben auf eine hohe Güte der Datenverknüpfung hin.

Tabelle 2: Konsistente Angaben in Bezug auf Geschlecht und Geburtsjahr

| | Konsistente Angaben | |
|--|---------------------|------|
| | n | % |
| Zusammenführungsergebnis | 314 334 | 100 |
| darunter konsistente Angaben bezüglich ... | | |
| ... des Geschlechts | 313 287 | 99,7 |
| ... des Geburtsjahres | 305 776 | 97,3 |
| ... des Geschlechts und des Geburtsjahres | 304 948 | 97,0 |

3.2 Version 2 (Querschnittsorientierung)

In den Querschnittsdaten sind alle Personen durch die Kombination der Ordnungsnummern (EF1, EF3, EF5b, EF12, EF4, EF5a) eindeutig zugeordnet. Für die Panelbildung muss die Ordnungsnummer der Person im Haushalt (EF5a) durch die längsschnittkonsistente Ordnungsnummer EF63 ersetzt werden. Die Daten können dann mittels Merkmalskombination der folgenden Ordnungsnummern zusammengeführt werden:

| | |
|------|---|
| EF1 | Bundesland |
| EF3 | Nr. des Auswahlbezirks (systemfrei) |
| EF5b | Jahresüberhang aus dem Vorjahr |
| EF12 | Berichtsquartal |
| EF4 | Nr. des Haushalts im Auswahlbezirk (systemfrei) |
| EF63 | Personennummer |

Alternativ können diese Ordnungsnummern auch zu einem alphanumerischen Indikator (z. B. „idpersx“) verkettet werden, mit dem dann die Querschnittsdaten zusammengeführt werden; entweder untereinander (long format) oder nebeneinander (wide format). Zu beachten ist, dass dabei die Angaben aus Jahresüberhängen wie bei Querschnittsauswertungen (s. o.) als eigenständige Personen- bzw. Haushaltssätze behandelt werden.

4. Panelmortalität

Aufgrund der Auskunftspflicht gibt es im MZ nur eine geringe Anzahl von Befragungsausfällen. Teilweise werden diese im Folgejahr befragt. Für die jüngste Erhebung (hier 2013) liegen allerdings Jahresüberhänge noch nicht vor. In Rotationspanels und Flächenstichproben treten spezielle Formen des Ausfalls auf. Fehlende Angaben aufgrund des Wechsels von Rotationsgruppen sind methodisch unproblematisch, da kein Zusammenhang zwischen den interessierenden Merkmalen und dem Ausfall besteht. Durch das Prinzip der Flächenstichprobe werden Haushalte und Personen, die zwischen zwei Erhebungszeitpunkten aus den Auswahlbezirken fortziehen, nicht weiter befragt. Räumlich mobile Haushalte und Personen können somit aufgrund fehlender Informationen zum Zeitpunkt vor bzw. nach dem Umzug nicht in die Längsschnittanalysen einbezogen werden. Im MZ gibt es für zugezogene Haushalte und Personen für die vorangegangenen Erhebungszeitpunkte nur eine geringe Anzahl an Informationen (z. B. Fragen zur Erwerbsbeteiligung vor 12 Monaten; ohne Auskunftspflicht). Das Prinzip der Flächenstichprobe kann somit zu methodischen Komplikationen führen, falls die räumliche Mobilität mit dem interessierenden Ereignis zusammenhängt.

In einem ersten Schritt ist deshalb zu klären, ob Ausfälle durch einen Wechsel der Rotationsgruppe designbedingt zustande gekommen sind oder ob andere Gründe vorliegen, wie beispielsweise Fort- und Zuzüge oder demographische Veränderungen (Geburten, Sterbefälle). Derzeit enthalten die Daten keine direkte Kennung der Rotationsgruppe. Die Ermittlung (potenziell) designbedingter Ausfälle ist näherungsweise dennoch gut möglich. Es kann geprüft werden, ob zu einem Befragungsausfall zum Panelzeitpunkt t bzw. $t + 1$ wenigstens ein Satz einer befragten Person oder ein Ausfallsatz (EF30=-7, -8) im gleichen Auswahlbezirk zum Zeitpunkt $t + 1$ bzw. t in den Längsschnittdaten vorliegt; d. h. der Auswahlbezirk Teil der Stichprobe war. Welche Kombinationsmöglichkeiten über das Vorliegen von Sätzen auf der Personen- und Auswahlbezirksebene sich hieraus ergeben und wie diese zu interpretieren sind, ist in nachfolgender Tabelle dargestellt.

Tabelle 3: Angaben zu Sätzen auf Personen- und Auswahlbezirksebene

| Angaben zur Panelzeit* für | | Sachverhalt |
|----------------------------|----------------|--|
| Sätze | Auswahlbezirk- | ~ Interpretation (Proxy) ~ |
| 12-- | 12-- | Sowohl für Sätze (Personen- und Ausfallsätze mit gültigen Angaben in EF4 und EF5b) als auch für Auswahlbezirke liegen nur zum Panelzeitpunkt 2012 Angaben vor. ~ Ausfälle 2013, da Auswahlbezirk 2012 letztmals in der Stichprobe war ~ |
| 12-- | 1213 | Für Sätze liegen nur zum Panelzeitpunkt 2012 Angaben vor. Für Auswahlbezirke liegen Angaben zu den Zeitpunkten 2012 und 2013 vor. ~ Ausfälle 2013 aufgrund von Wegzügen oder Todesfällen ~ |
| --13 | --13 | Sowohl für Sätze als auch für Auswahlbezirke liegen nur zum Panelzeitpunkt 2013 Angaben vor. ~ Ausfälle 2012, da Auswahlbezirk 2013 erstmals in der Stichprobe war ~ |
| --13 | 1213 | Für Sätze liegen nur zum Panelzeitpunkt 2013 Angaben vor. Für Auswahlbezirke liegen Angaben zu den Zeitpunkten 2012 und 2013 vor. ~ Ausfälle 2012 aufgrund von Zuzügen oder Geburten ~ |
| 1213 | 1213 | Sowohl für Sätze als auch für Auswahlbezirke liegen zu beiden Panelzeitpunkten Angaben vor. ~ Keine Panelmortalität ~ |

* 12 = 2012, 13 = 2013, -- = keine Angaben in Personensätzen bzw. keine gültigen Angaben in Ausfallsätzen

Die gewisse Unsicherheit hängt mit der oben erwähnten Einschränkung der Informationen zu den Ausfallsätzen zusammen. Es ist deshalb sinnvoll, die Prüfung auf Sätze befragter Personen und auf Ausfallsätze mit gültigen Angaben zum Jahresüberhang (EF5b) sowie zur Haushaltsnummer (EF4) in einem Auswahlbezirk zu beschränken. Der Identifikator des Auswahlbezirks (z. B. idpsu) kann durch die Verkettung von EF1 „Bundesland“ und EF3 „Nr. des Auswahlbezirks (systemfrei)“ erstellt werden. Wie Auswertungen mit den MZ-Originaldaten gezeigt haben, lassen sich auf diese Weise rund 98 % der durch einen Wechsel der Rotationsgruppe bedingten Ausfälle ermitteln.

Nähere Aufklärungen fehlender Angaben zum Zeitpunkt t (2012) können mit den Angaben der Erhebung $t + 1$ (2013) zum Zuzug in den letzten 12 Monaten (EF45) oder zum Wohnsitz vor 12 Monaten (EF451, ohne Auskunftspflicht) ermittelt werden. In Bezug auf den Umgang mit Ausfällen finden sich u. a. im Datenhandbuch zum MZ-Panel 1996-1999 (Statistisches Bundesamt 2006) weitere Hinweise.

5. Schluss

Ab der Erhebung 2012 sind die Scientific-Use-Files des Mikrozensus in Bezug auf die Substichprobenziehung und längsschnittkonsistente Ordnungsnummern so umgestaltet, dass Forscherinnen und Forscher selbstständig Längsschnittfiles erstellen können. Wie bisher können Querschnittsauswertungen durchgeführt werden. Die neu geschaffene Längsschnittverknüpfung bietet der Forschung weitergehende methodische und inhaltliche Analysemöglichkeiten. Beispielsweise können Verlaufsanalysen durchgeführt und Statusübergänge (Brutto-Veränderungen) untersucht werden. Darüber hinaus können für den Vergleich von Querschnittsergebnissen (Netto-Veränderungen) statistisch korrekte Stichprobenfehler geschätzt werden. Dabei sind Restriktionen infolge der Stichprobeneigenschaften (Klumpen-/ Flächenstichprobe und Rotation) und teilweise fehlender Informationen zu beachten. Die neuen Möglichkeiten sind für Forscherinnen und Forscher aber auch mit einem eigenen Aufwand bei der Panelerstellung und Datenprüfung verbunden. Um diesen möglichst gering zu halten, werden Beispiele zur Umsetzung der Datenzusammenführung und zu ersten Ausfallprüfungen mit den Statistikprogrammen SAS, SPSS und Stata bereitgestellt (siehe <https://www.gesis.org/missy/materials/MZ/tools/panel>).

Literatur

- Afentakis, A.; Bihler, W. (2005): Das Hochrechnungsverfahren beim unterjährigen Mikrozensus ab 2005. *Wirtschaft und Statistik*, 10/2005: 1039-1048.
- Bašić, E.; Rendtel, U. (2005): Estimation strategies in the presence of noncoverage in the German Microcensus-Panel: An evaluation using SOEP data. Arbeitspapier #8. Freie Universität Berlin, Fachbereich Wirtschaftswissenschaften.
https://www.destatis.de/DE/Methoden/Methodenpapiere/Mikrozensus/Arbeitspapiere/Arbeitspapier8.pdf?__blob=publicationFile.
- Bihler, W.; Zimmermann, D. (2016): Die neue Mikrozensusstichprobe ab 2016. S. 20-29 in: *Wirtschaft und Statistik* (6 / Dezember).
https://www.destatis.de/DE/Publikationen/WirtschaftStatistik/2016/06/DieNeueMikrozensusstichprobe_062016.pdf?__blob=publicationFile.
- Deininger, R. (1960): Repräsentativstatistik der Bevölkerung und des Erwerbslebens (Mikrozensus), S. 135-175 in: Statistisches Bundesamt (Hg.): *Stichproben in der Amtlichen Statistik*. Stuttgart/Mainz: Kohlhammer.
- Heidenreich, H.-J.; Herter-Eschweiler, R. (2002): Längsschnittdaten aus dem Mikrozensus. Basis für neue Analysemöglichkeiten. *Wirtschaft und Statistik*, Heft 8/2002.
https://www.destatis.de/DE/Publikationen/WirtschaftStatistik/AllgemeinesMethoden/Laengsschnittdaten.pdf?__blob=publicationFile.
- Herberger, L. (1973): Praktische Erfahrungen mit Verlaufsstatistiken, in: *Allgemeines Statistisches Archiv*, 57: 54-77.
- MZG 2016: Gesetz zur Durchführung einer Repräsentativstatistik über die Bevölkerung und die Arbeitsmarktbeteiligung sowie die Wohnsituation der Haushalte (Mikrozensusgesetz - MZG) vom 7. Dezember 2016.
https://www.destatis.de/DE/Methoden/Rechtsgrundlagen/Statistikbereiche/Inhalte/1062_MZG.pdf?__blob=publicationFile.
- Schimpl-Neimanns, B. (2006): Filekonzept zum Mikrozensuspanel. Arbeitspapier #12. Mannheim: ZUMA.
https://www.destatis.de/DE/Methoden/Methodenpapiere/Mikrozensus/Arbeitspapiere/Arbeitspapier12.pdf?__blob=publicationFile.
- Schimpl-Neimanns, B. (2011), Schätzung des Stichprobenfehlers in Mikrozensus Scientific Use Files ab 2005. *AStA Wirtschafts- und Sozialstatistisches Archiv* 5 (1): 19-38. DOI: [dx.doi.org/10.1007/s11943-011-0092-4](https://doi.org/10.1007/s11943-011-0092-4) (Preprint).
- Statistisches Bundesamt (2012): Mikrozensus: Haushaltszahlen ab 2005. Wiesbaden.
https://www.destatis.de/DE/Methoden/MikrozensusHaushaltszahlen.pdf;jsessionid=429244BAAA5838007B3394643420CA1C.InternetLive1?__blob=publicationFile.
- Statistisches Bundesamt (Hg.) (2006): *Handbuch des Mikrozensuspanel 1996-1999 - Methodenverbund "Aufbereitung und Bereitstellung des Mikrozensus als Panelstichprobe"*. Wiesbaden: Statistisches Bundesamt.
https://www.gesis.org/missy/files/documents/MZ/MZP9699_Handbuch_1_.pdf.
- Statistisches Bundesamt (Webseite): *Wissenschaftsforum – Methodenverbund „Aufbereitung und Bereitstellung des Mikrozensus als Panelstichprobe“*.
<https://www.destatis.de/DE/Methoden/Methodenpapiere/Mikrozensus/Mikrozensus.html>.