

Mikrozensus-Längsschnittdaten

Nutzung der Mikrozensus-Arbeitskräfteerhebung für die Analyse der Dynamik des österreichischen Arbeitsmarktes – Aufbau der zugrundeliegenden Datenbestände

DANIELA GUMPRECHT

ANGELIKA MERANER

ANDREAS BAIERL

Die österreichische Mikrozensus-Arbeitskräfteerhebung (MZ-AKE) besitzt neben der bekannten Querschnittsdimension, welche sich durch die Art der Erhebung auf Quartale bezieht, auch eine Längsschnittdimension. Diese ergibt sich aus der Dauer, für die ein ausgewählter Haushalt in der Stichprobe bleibt, nämlich für fünf aufeinanderfolgende Quartale. Schon eine Längsschnittbetrachtung über nur zwei aufeinanderfolgende Quartale eröffnet eine Vielzahl an zusätzlichen Analysemöglichkeiten im Vergleich zur Betrachtung von nur einem Quartal. Zusätzlich zur Abbildung von Veränderungen auf der Aggregatebene, welche mit Querschnittsdaten möglich ist, können auch Veränderungen auf Personenebene (sog. „flows“) betrachtet werden. Die Ströme von Personen aus einem Arbeitsmarktstatus in einen anderen rücken die Analyse der Dynamik des Arbeitsmarktes in den Fokus. Für die Analyse von Veränderungen von Quartal zu Quartal bedarf es geeigneter Datenbestände im Längsschnitt, deren Erstellung und Aufbau in diesem Artikel im Detail beschrieben werden. Eine Analyse der Stromgrößen und weitere Ergebnisse werden in einem späteren Artikel präsentiert.

Einleitung

Die Mikrozensus-Arbeitskräfteerhebung basiert auf einer rotierenden Stichprobe über fünf aufeinanderfolgende Quartale. Die Stichprobe ist eine nach Bundesländern geschichtete Zufallsauswahl von Haushalten. Da in einem Haushalt alle dort lebenden Personen gefragt werden, ist die Stichprobe darüber hinaus auch eine nach Bundesländern geschichtete Klumpenauswahl von Personen. Ein Haushalt bleibt für fünf aufeinanderfolgende Quartale in der Stichprobe. Unabhängig von der Zusammensetzung des Haushalts zum Zeitpunkt der Stichprobenziehung oder der ersten Befragung werden immer jene Personen befragt, die in der jeweiligen Referenzwoche, auf die sich die zentralen Fragen beziehen, im Haushalt leben. Die Referenzwochen sind gleichmäßig über alle Wochen eines Quartals verteilt. Die Respondenten und Respondentinnen sind gesetzlich zur Auskunftserteilung verpflichtet. Ziel ist es, jede (volljährige) Person¹⁾ persönlich zu befragen (Selbstauskunft). Ist das nicht möglich, dann kann auch eine andere erwachsene Person im Haushalt Auskunft erteilen (Fremdauskunft). Jedes Quartal scheidet ein Fünftel der Haushalte aus der Stichprobe aus und wird durch neue Haushalte ersetzt. In der Mehrheit der Fälle hat man über eine Person Informationen in allen fünf aufeinanderfolgenden Quartalen. Zieht eine Person um, dann fehlen die Informationen in einem oder mehreren Quartalen.

Die Analysemöglichkeiten und Anwendungsbereiche der Mikrozensus-Arbeitskräfteerhebung (MZ-AKE) werden deutlich erweitert, wenn neben der bekannten und weitverbreiteten Querschnittsdimension und der Analyse von Monaten, Quartalen und Jahren nun auch die Längsschnittdimension für jeweils zwei aufeinanderfolgende Quartale nutzer- und nutzerinnenfreundlicher aufbereitet und be-

trachtet wird. Damit sind dann nicht nur Veränderungen von Aggregaten, z.B. der Zahl der Arbeitslosen in zwei aufeinanderfolgenden Quartalen, sondern auch Veränderungen auf Personenebene sichtbar. Beispielsweise lässt sich feststellen, wie viele und v.a. auch welche Personen in die Arbeitslosigkeit wechseln, wie viele aus der Arbeitslosigkeit herauskamen, woher sie kamen bzw. wohin sie gingen. Damit wird die Dynamik auf dem Arbeitsmarkt wesentlich besser dargestellt. Denn auch wenn sich an den Gesamtzahlen keine Veränderung von einem Quartal aufs nächste zeigt, können trotzdem sehr viele Personen ihren Status gewechselt haben. Der Fokus liegt auf der individuellen Veränderung des Arbeitsmarktstatus von Personen. Somit ist die korrekte Abbildung von Veränderungen für die gesamte Datenaufbereitung zentral.

Zu dem Themenbereich „Übergänge auf dem Arbeitsmarkt“ gibt es auf nationaler wie auch internationaler Ebene verschiedene Arbeiten und Ansätze. Bezogen auf den österreichischen Arbeitsmarkt sollen hier nur drei Arbeiten exemplarisch erwähnt werden. *Schoiswohl und Wüger* (2016) haben aus den MZ-AKE Quartalsbeständen einen Längsschnittbestand von 2004 bis 2014 erstellt, neu gewichtet und Arbeitsmarktübergänge von einem Quartal zum nächsten analysiert. *Bacher et al.* (2013) haben ebenfalls einen Längsschnittbestand aus den MZ-Quartalsbeständen erzeugt und sich dann auf die Analyse der Arbeitsmarktlage von Jugendlichen konzentriert. Auf einer anderen Datengrundlage beruhen Flow-Statistiken basierend auf den registerbasierten Erwerbsverläufen (*Wanek-Zajic*, 2016). Hier werden aus Verwaltungsdaten überschneidungsfreie Erwerbskarrieren von Personen erstellt, die ebenfalls ein sehr breites Feld an Analysemöglichkeiten von Übergängen in den Erwerbsbiographien der Personen eröffnen. International gibt es ebenfalls erste Ansätze, so etwa von Eurostat Analysen auf Basis einfacher Gewichtungsmo-
delles (*Eurostat*, 2018).

¹⁾ Für Minderjährige wird die Auskunft durch ihren gesetzlichen Vertreter gegeben.

Der Mikrozensus-Flows-Datenbestand

Für die Analyse der Veränderungen auf Personenebene wurden für alle Quartalsübergänge ab 2008 neue, eigenständige Datenbestände erstellt. Sie werden in der Folge „MZ-Flows“-Bestände genannt und beinhalten Informationen über Personen in jeweils zwei aufeinanderfolgenden Quartalen. Personen und Merkmale stammen aus den MZ-AKE-Quartalsdaten.²⁾

Notation und Namenskonventionen

Aus der Benennung der Datenbestände geht hervor, dass es sich um spezielle Längsschnittbestände handelt, dass sie MZ-Daten enthalten, und auf welche beiden Quartale sie sich beziehen. Das aktuellere Quartal in den MZ-Flows-Beständen wird als **Referenzquartal** bezeichnet und nicht gesondert beschriftet. Das frühere Quartal wird allgemein als **Lag-Quartal** bezeichnet, in der Beschriftung wird auch noch zusätzlich durch eine Zahl angegeben, um welches frühere Quartal es sich handelt; z.B. „lag1“ steht für das Vorquartal, „lag4“ wäre das Vorjahresquartal.³⁾ Der Datenbestand für die Analyse der Veränderungen z.B. vom vierten Quartal 2016 auf das erste Quartal 2017 hat den Namen MZ_flow_2017q1_lag1.

Die **quartalsunabhängigen Merkmale** und jene, die sich auf das **Referenzquartal** beziehen, werden nicht gesondert gekennzeichnet, sondern haben denselben Namen wie in den MZ-AKE-Querschnittsbeständen; die Merkmale, die sich auf das Lag-Quartal beziehen, sind durch das Suffix „_lag“ erkennbar. Die Beschriftung der Labels ist analog dazu, Merkmale des Vorquartals sind auch im Variablenlabel mit dem Zusatz „(Lag-Quartal)“ gekennzeichnet. Beispiel: Das Geschlecht ist als zeitunabhängiges Merkmal mit „bsex“ be-

²⁾ Bei der Erstellung der Daten werden im Rahmen der Qualitätsprüfung auch Informationen verwendet, die nicht in den endgültigen, authentischen MZ-AKE-Quartalsbeständen enthalten sind (z.B. nicht befragte Personen oder der Erwerbsstatus aus Verwaltungsdaten).

³⁾ Diese Bestände werden derzeit nicht erzeugt.

nannt, der Erwerbsstatus nach internationaler Definition (kurz ILO⁴⁾-Erwerbsstatus genannt) im Referenzquartal (also im späteren Quartal, z.B. Q1 2017) heißt „xerwstat“, und der Erwerbsstatus im Lag-Quartal (also im Vorquartal Q4 2016) hat den Namen „xerwstat_lag“.

Personen

Die MZ-Flows-Bestände beinhalten generell alle Personen, die in zumindest einem der beiden Quartale im Mikrozensus erfolgreich befragt wurden und aufgrund des Rotationschemas in beiden Quartalen Teil der Befragung sein sollten. Damit sind auf jeden Fall alle Personen ausgeschlossen, die im Lag-Quartal bereits zum fünften und damit letzten Mal erhoben wurden. Ausgeschlossen sind auch alle, die im Referenzquartal erstmals erhoben wurden. Von den übrigen Personen werden jene Personen ausgeschlossen, für die es in beiden Quartalen keine Informationen aus der Befragung gibt. Das sind z.B. Personen, die bereits vor dem Lag-Quartal ausgezogen sind und in beiden Quartalen gar nicht mehr zum MZ-Haushalt gehörten, oder Personen, bei denen der gesamte Haushalt in beiden Quartalen nicht erreicht werden konnte. Sehr wohl enthalten sind aber Personen, die in einem der beiden Quartale befragt wurden,⁵⁾ im anderen aber nicht erhoben werden konnten, entweder weil sie in diesem Quartal gar nicht zum MZ-Haushalt gehörten (z.B. aufgrund eines Umzugs, oder aber auch, weil sie zwischen den beiden Referenzwochen verstorben sind oder erst geboren wurden), oder weil sie für ein Interview einfach nicht erreicht werden konnten. Selbstverständlich sind alle Personen enthalten, für die in beiden Quartalen Auskunft gegeben wurde. Um die Auswahl der Personen für die MZ-Flows-Bestände noch einmal zu verdeutlichen, wird auf die spezielle Form der MZ-AKE-Stichprobe bzw. -Erhebung verwiesen: *siehe Infobox und Abbildung 1.*

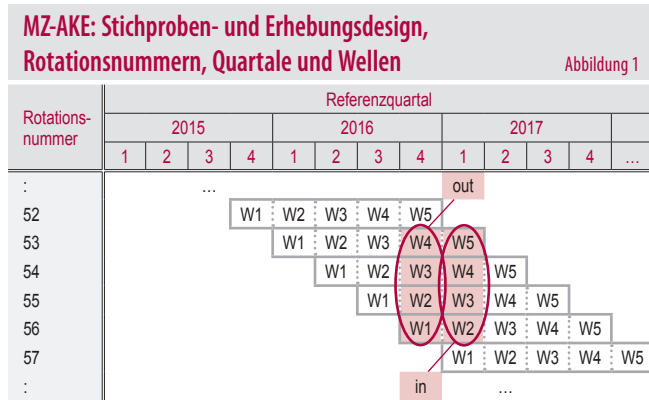
⁴⁾ ILO= International Labour Organisation.

⁵⁾ ... oder über die eine andere im Haushalt lebende Person Auskunft gab, sogenannte Proxy-Fälle - sie werden hier wie Selbstauskünfte behandelt.

Infobox

Mikrozensus-Arbeitskräfteerhebung Stichprobe und Erhebung

Die MZ-AKE ist schematisch und beispielhaft für einige Quartale in *Abbildung 1* dargestellt. Gezogen wird eine Adresse, befragt werden alle Personen, die an dieser Adresse bzw. in diesem Haushalt leben – Adresse und Haushalt werden synonym verwendet. Alle Personen eines MZ-Haushalts sind für fünf Quartale in der Erhebung. Verlässt eine Person einen MZ-Haushalt, so verlässt sie auch die Stichprobe. Personen, die wegziehen, werden nicht weiter verfolgt. Andererseits werden Personen, die in einen MZ-Haushalt einziehen, in die Befragung aufgenommen. Bei der Stichprobenziehung erhält jeder Haushalt (und damit auch alle ihm zugehörigen Personen) eine Nummer, die sogenannte Rotationsnummer. Diese Nummer behält er bis zum Ausscheiden aus der Stichprobe nach fünf Quartalen. Die Rotationsnummer gibt an, wann ein Haushalt erstmals in der Stichprobe aufscheint. Haushalte, die zum selben Zeitpunkt in die Stichprobe aufgenommen wurden, haben die gleiche Rotationsnummer. Die fünf aufeinanderfolgenden Befragungen werden auch fünf Wellen (Erstbefragung 1. Welle, Zweitbefragung 2. Welle usw.) genannt. In *Abbildung 1* sind sie mit W1 bis W5 beschriftet. In einem Quartal werden fünf Rotationsnummern befragt, eine Rotationsnummer wird in fünf aufeinanderfolgenden Quartalen befragt. In einem Quartal gibt es eine 1., eine 2., eine 3., eine 4. und eine 5. Welle. Eine Rotationsnummer hat fünf Wellen. In einem Jahr gibt es vier erste Wellen, vier 2. Wellen, vier 3., 4. und 5. Wellen.



Die Personen im MZ-Flows-Bestand 2017q1_lag1 sind in *Abbildung 1* farbig markiert. Neben den Personen, die in beiden Quartalen zu einem MZ-Haushalt gehören, 1. bis 4. Welle (W1-W4) im Lag-Quartal und 2. bis 5. Welle (W2-W5) im Referenzquartal, das sind Personen mit den Rotationsnummern 53, 54, 55 und 56, zählen auch jene dazu, die einen Haushalt vorzeitig verlassen („out“) und jene, die neu hinzukommen („in“); diese haben bzw. erhalten ebenfalls eine Rotationsnummer 53, 54, 55 oder 56 – je nach Haushaltzugehörigkeit.

Diese Personen stellen die größte Herausforderung bei der Erstellung der Datenbestände dar, da es für sie nur in einem Quartal Informationen über ihre (Erwerbs-) Lebensumstände aus der Befragung gibt. Die **Information aus dem Voroder Folgequartal fehlt**. Doch genau diese ist von großem Interesse für die Analyse der Veränderungen. Würde man die Personen einfach aus dem Flows-Bestand ausschließen, könnte dadurch das Ausmaß der Veränderungen unterschätzt werden (z.B. dann, wenn sehr viele Erwerbsstatuswechsel auch mit einer Wohnsitzänderung einhergehen, sich aber die Quartalseckzahlen für die einzelnen Erwerbsstatus wenig unterscheiden). Es ist anzunehmen, dass sich die mobilen Personen strukturell von jenen unterscheiden, die ihren Wohnsitz nicht wechseln. Fehlende Personen des Lag- oder Referenz-Quartals sollen daher im Datenbestand erhalten bleiben, und die fehlende Information wird imputiert (wobei wenn möglich, auch externe Informationen über den Erwerbsstatus des fehlenden Quartals in die Schätzung miteinbezogen werden). Wie diese Imputation im Detail funktioniert, ist im Abschnitt „*Imputation*“ beschrieben.

Merkmale

In einem regulären MZ-AKE-Quartalsbestand sind weit über 300 Merkmale enthalten. Die für eine Analyse von Arbeitsmarktdynamiken wichtigsten arbeitsmarktbezogenen und soziodemographischen Merkmale wurden ausgewählt und für diesen speziellen Zweck aufbereitet. Generell kann man verschiedene Merkmalstypen unterscheiden. Es gibt Merkmale, die sich nie (oder beinahe nie) verändern (z.B. Geburtsland, Geschlecht, etc.) – sie sind im MZ-Flows-Bestand nur einmal enthalten. Es gibt Merkmale, die sich per Definition ändern müssen und unabhängig vom Verhalten der

Personen sind (z.B. Referenzwoche oder das Alter einer Person) – sie gibt es zweimal, einmal für das Referenz- und einmal für das Lag-Quartal. Schließlich gibt es Merkmale, die sich ändern können und vom Verhalten der Personen abhängen (z.B. Erwerbsstatus, Arbeitszeit, etc.) – sie sind eigentlich von Interesse und natürlich ebenfalls für das Referenz- sowie das Lag-Quartal enthalten. Außerdem kann man zwischen personen- und haushaltsbezogenen Merkmalen unterscheiden. Merkmale die sich explizit auf den MZ-Haushalt beziehen, sind mit dem Suffix „_hh“ gekennzeichnet. Ein Beispiel für so ein haushaltsbezogenes Merkmal ist das Bundesland des MZ-Haushalts (Variablenname „xnuts2_hh“). Egal, ob eine Person umgezogen ist oder nicht, das Bundesland gibt immer das Bundesland des MZ-Haushalts, dem die Person in zumindest einem der beiden Quartale angehörte, an.

Die MZ-Flows-Bestände sind standardisiert, aber dennoch nicht über alle Quartale hinweg komplett gleich. Einerseits gab es Änderungen bei der Klassifikation der Berufe (ISCO).⁶⁾ Die Flows-Bestände von 2008q2_lag1 bis 2011q1_lag1 zeigen die Berufshauptgruppen nach der Klassifikation ISCO-88, ab dem Bestand 2011q2_lag1 entsprechen die Berufshauptgruppen der Klassifikation ISCO-08. Diese Änderung ist in den Variablennamen sichtbar. Andererseits gab es Änderungen bei der Gruppierung von Staaten der Merkmale Staatsangehörigkeit und Geburtsland. Die Merkmale haben zwar über alle Bestände hinweg die gleichen Variablennamen, aber ab dem Jahr 2014 gibt es eine eigene (zusätzliche) Ausprägung für Kroatien.

Neben bekannten MZ-AKE-Merkmalen gibt es auch neue, eigens für den MZ-Flows-Bestand erstellte Merkmale. Dies sind: eine neue Personenkennzahl (sie ersetzt die Personenkennzahl „asbper“ des Querschnittsbestandes), ein weiteres Merkmal, das angibt, ob eine Person ihren Wohnort gewechselt hat, geboren wurde oder gestorben ist, und ein Merkmal, das hilft, jene Personen auszuwählen, die in beiden Quartalen am Leben waren.⁷⁾ Neu sind auch zwei Indikatormerkmale, die zeigen, ob eine Person in der MZ-AKE im Referenz- bzw. Lag-Quartal erfolgreich befragt wurde.⁸⁾ Sie zeigen somit an, ob die veränderlichen personenbezogenen Merkmale des entsprechenden Quartals im Zuge der Erstellung des MZ-Flows-Bestandes imputiert wurden oder aus den authentischen MZ-AKE-Beständen stammen.

Insgesamt beinhaltet der MZ-Flows-Bestand zirka 80 Merkmale. Zu den **personenbezogenen unveränderlichen**, d.h. referenzzeitunabhängigen Merkmalen, die nur einmal im Bestand enthalten sind, gehören u.a. das Geschlecht, die

⁶⁾ ISCO= International Standard Classification of Occupations. In der MZ-AKE wurde bis zum Berichtszeitraum Q4 2010 ausschließlich die ISCO-88 verwendet; im Jahr 2011 wurde die neue ISCO-08 eingeführt, wobei für 2011 beide Klassifikationen parallel verwendet wurden, ab 2012 nur noch die neue.
⁷⁾ Ungeborene und Verstorbene können damit einfach aus den Analysen ausgeschlossen werden.
⁸⁾ Selbst- oder Fremdauskunft.

Staatsangehörigkeit⁹⁾ (gruppiert), das Geburtsland (gruppiert), der Migrationshintergrund und die Migrationsgeneration. Die **haushaltsbezogenen** Merkmale sind ebenfalls referenzzeitunabhängig und nur einmal im Bestand enthalten: das Bundesland, der Urbanisierungsgrad und die Gemeindegrößenklasse.

Die Mehrheit sind aber **personenbezogene veränderliche**, d.h. referenzzeitabhängige Merkmale. Sie sind zweimal im Bestand enthalten, jeweils für das Referenz- und das Lag-Quartal. Dazu zählen Referenzwoche, -quartal und -jahr, das Alter und die Haushaltsgröße. Letztere ist das einzige Merkmal, das auch fehlende Werte enthält, die im Gegensatz zu allen anderen Merkmalen nicht imputiert werden. Diese fehlenden Werte treten bei Personen auf, die nicht mehr oder noch nicht im MZ-Haushalt leben. In diesem Fall ist die Haushaltsgröße im entsprechenden Quartal nicht bekannt (Code: -2). Es gibt eine Reihe bildungsbezogener Merkmale, wie die höchste abgeschlossene Bildung (nationale Darstellung), das Jahr des höchsten Ausbildungsabschlusses und einen Indikator, ob eine Person in (formaler)¹⁰⁾ Ausbildung war, sowie den Zweck der letzten Ausbildung. Die meisten Merkmale aber stammen aus dem Bereich der Erwerbstätigkeit: der Erwerbsstatus nach internationaler Definition selbst (kurz auch ILO-Status genannt), die Berufshauptgruppen nach ISCO-88/ISCO-08, die berufliche Stellung, die Tätigkeit und der Wirtschaftszweig nach ÖNACE 2008. Des Weiteren gibt es noch die ergänzenden Merkmale zum Erwerbsstatus: stille Reserve und die zeitbezogene Unterbeschäftigung sowie eine Reihe job- und arbeitszeitbezogener Merkmale (Dauer der jetzigen Arbeit, Arbeitszeit) und Information zu Teilzeitarbeit (Grund), Leiharbeit und Befristung (Grund und Dauer der Befristung).

Wie bei den Querschnittsbeständen gibt es auch für die Längsschnittdaten ein **Hochrechnungsgewicht**, um von der Stichprobe wieder auf die Grundgesamtheit zu kommen. Diese Gewichte stammen nicht aus den authentischen MZ-AKE-Quartalsbeständen, sondern werden eigens für die MZ-Flows-Bestände erstellt. Details dazu sind im Kapitel „Gewichtung“ beschrieben.

Falsche und fehlende Werte

Durch die Längsschnittverknüpfung der MZ-AKE-Quartalsbestände kann es immer wieder Fälle geben, bei denen Merkmale aus Lag- und Referenzquartal ein und derselben Person nicht zusammenpassen, oder dass sogar zwei unterschiedliche Personen hinter derselben Personen-ID stehen. Beides tritt infolge irgendeines Fehlers, entweder auf Seite der Erhebungspersonen oder der Befragten an zumindest einem Erhebungszeitpunkt, auf. Solche Fehler sind in man-

chen Fällen sehr einfach und eindeutig als solche zu erkennen, z.B. wenn eine Person im späteren Quartal jünger ist, dann stimmt zumindest eine der Altersangaben in Lag- und Referenzquartal sicher nicht. In anderen Fällen sind Fehler gar nicht zu erkennen, da es sich um mögliche Änderungen handelt. Z.B. kann eine Veränderung der Arbeitszeit einer realen Änderung entsprechen oder aber Ergebnis einer Falschauskunft oder eines Eingabefehlers o.ä. sein. Die authentischen MZ-AKE-Bestände, die die Basis für die MZ-Flows-Bestände sind, sind von hoher Qualität. Sie durchlaufen eine Vielzahl an Kontrollen und Maßnahmen zur Qualitätssicherung, angefangen vom Fragebogendesign und der Schulung des Erhebungspersonals, über eine Vielzahl an Plausibilitätsprüfungen an unterschiedlichsten Stellen während der Erstellung der Datenbestände bis hin zum Vergleich von Ergebnistabellen mit Statistiken anderer Institutionen.

Nichtsdestoweniger kann es Fehler und Implausibilitäten geben, und gerade bei der Längsschnittbetrachtung können sich andere Probleme oder Inkonsistenzen als bei einem reinen Querschnittsbestand zeigen. Sind Fehler oder Inkonsistenzen einmal identifiziert, müssen sie behandelt werden. Werden Werte als falsch erkannt, werden sie gelöscht. Infolgedessen steht man dann vor dem Problem fehlender Werte.

Fehlende Werte gibt es auch dann – und das ist die wesentlich häufigere Ursache –, wenn für eine Person für eines der beiden Quartale keine Informationen aus der Befragung vorliegen. Dies betrifft auf jeden Fall alle Personen, die zwischen Lag- und Referenzquartal ihren Wohnsitz gewechselt haben. Außerdem trifft es Personen, mit denen aus unterschiedlichsten Gründen kein Interview stattfinden konnte (und für die auch keine andere Person Auskunft geben konnte). Um Aussagen über die Dynamik des Arbeitsmarktes und die Veränderungen im Erwerbsverhalten machen zu können, müssen die fehlenden Merkmalsausprägungen geschätzt und ergänzt werden. Dabei kann man auch auf Informationen derselben Person aus dem Vor- oder Folgequartal zurückgreifen. Außerdem gibt es für die große Mehrheit der Personen (etwa 95%) Informationen über ihren Erwerbsstatus aus administrativen Datenquellen, nämlich des Hauptverbandes der österreichischen Sozialversicherungsträger (HV) und des Arbeitsmarktservice Österreich (AMS). Die Daten der MZ-AKE, des HV und des AMS können über das bereichsspezifische Personenkennzeichen (bPK) verknüpft werden.¹¹⁾ Die Definitionen des Erwerbsstatus nach ILO, wie er in der MZ-AKE verwendet wird, und des Erwerbsstatus aus administrativen Daten des HV und AMS unterscheiden sich zwar, dennoch stellen diese Administrativdaten eine sehr wertvolle Quelle für die Imputation der Erwerbsmerkmale nicht

⁹⁾ Die Staatsangehörigkeit wird als unveränderlich definiert, da ein Wechsel zwischen zwei Quartalen sehr selten auftritt.

¹⁰⁾ Formale Ausbildung, d.h. Schüler/-in oder Student/-in im regulären Schul- oder Hochschulwesen oder - für Personen unter 40 Jahren - Lehrling.

¹¹⁾ Hat eine Person ein bPK und ist sie nicht in den HV- oder AMS-Daten zu finden, so ist die Person nicht in Österreich als erwerbstätig oder arbeitslos gemeldet. Die Person ist entweder Nicht-Erwerbsperson oder im Ausland beschäftigt.

befragter Personen dar. Außerdem werden die Personen der MZ-Flows-Bestände mit der österreichischen Wohnbevölkerung laut dem Register des Bevölkerungsstandes (POPREG) abgeglichen; der Abgleich erfolgt wieder über das bPK. Damit sind z.B. Binnenwanderungsfälle erkennbar.

Plausibilitätskontrolle

Bei den Plausibilitätskontrollen (kurz „Plaus“ genannt) wird hier zwischen Personen- und Merkmals-Plaus unterschieden. Bei der Personen-Plaus wird zunächst geprüft, ob im Lag- und Referenzquartal tatsächlich dieselbe Person hinter derselben Personen-ID steht. Sind die Personen geprüft, folgt die Merkmals-Plaus. Hier werden die Merkmalsausprägungen in Bezug auf die zeitliche Abfolge geprüft.

Personen-Plaus

Hier unterscheidet man generell Personen mit und ohne bereichsspezifisches Personenkennzeichen. Die bPKs werden im Zuge der Erstellung der authentischen MZ-AKE-Quartalsbestände für alle befragten Personen gesucht. Für Personen, die nicht im authentischen Quartalsbestand aufscheinen, z.B. weil sie ausgezogen sind, wird in diesem Quartal nicht nach einem bPK gesucht. Allerdings wurde in so einem Fall in zumindest einem anderen früheren Quartal gesucht, nämlich in jenem, in dem sie noch im MZ-Haushalt lebten und erhoben wurden.¹²⁾ Für die Personen der Flows-Bestände wird daher im Referenzquartal und in allen früheren Quartalen nach einem bPK gesucht. Je nach Quartal (Lag- oder Referenz) und Befragungswelle einer Person sind das null bis vier Vorquartale. Durch diese Suche in allen Vorquartalen kann es auch sein, dass für eine Person unterschiedliche bPKs gefunden werden – das ist ein eindeutiger Hinweis auf einen Fehler. Dies kann passieren, wenn bei der Erhebung bereits vorhandene Personen-IDs fälschlicherweise anderen Personen zugewiesen werden, z.B. wenn in einem Zweipersonenhaushalt die beiden Personen vertauscht werden.

Der erste Schritt bei der Personen-Plaus ist daher die Prüfung, ob die Personen, für die im Lag- und Referenzquartal ein bPK gefunden wurde, auch tatsächlich zweimal das gleiche bPK haben. Unterschiedliche bPKs bei ein und derselben Personen-ID treten sehr selten (unterer einstelliger Bereich) auf. Wenn doch, dann wird geprüft, ob tatsächlich nur die Personenkennzahlen vertauscht wurden. Wenn ja, werden die Personen-IDs korrigiert. Wenn nicht, werden die Zeilen gelöscht.

Kann in allen Quartalen nur ein oder überhaupt kein bPK für eine Person gefunden werden, dann werden für die Prüfung, ob es sich um dieselbe Person handelt, Merkmale aus der MZ-AKE-Erhebung selbst herangezogen, nämlich das Geschlecht, das Geburtsdatum und das Alter aus dem Lag- und dem Referenzquartal. Je nach Vorhandensein und Über-

einstimmung der Merkmale wird ein Merkmal für die Güte der Übereinstimmung erstellt. Die Ausprägungen gehen von „vollständige Übereinstimmung“¹³⁾ bis „sehr geringe Übereinstimmung“. Personen mit sehr geringer Übereinstimmung werden aus dem Flows-Bestand entfernt, da es sich dabei wahrscheinlich um zwei verschiedene Personen handelt. Anderenfalls würde man die Dynamik am Arbeitsmarkt überschätzen, da man auch Unterschiede zwischen Personen fälschlicherweise als Veränderungen einer Person betrachten würde. Entfernt werden auch Personen, bei denen man in beiden Quartalen keine Informationen hat – es wird damit verhindert, dass beide Quartale imputiert werden müssen. Eine Personen-Plaus in dieser Art ist bei all jenen Personen nicht möglich, bei denen es ausschließlich für ein Quartal Informationen gibt. Sie bleiben im Bestand enthalten.

Nach dem Entfernen aller unplausiblen Personen und der Berechnung der Haushaltsgröße (Anzahl der plausiblen Personen, die in einem Quartal zu einem Haushalt gehören) folgt noch eine weitere Plaus-Prüfung, und zwar werden alle Personen unter 15 Jahren, die ganz allein in einem Haushalt leben, entfernt. Solche Fälle treten aber extrem selten auf.

Ab dem ersten Quartal 2015 steht das Zentrale Personenstandsregister (ZPR) zur Verfügung. Vom ersten Quartal 2015 bis zum vierten Quartal 2017 werden alle laut Befragung verstorbenen Personen mit bPK mit den Sterbefällen des ZPR abgeglichen. Nur wenn eine laut MZ-AKE-Erhebung verstorbene Person auch im ZPR ein Todesdatum vor dem Quartalsende aufweist, wird sie als tatsächlich verstorben geführt. Alle anderen gelten als nicht verstorben. Damit wird ausgeschlossen, dass Personen in einem Quartal (rein aufgrund der Auskunft einer anderen Person in der Befragung oder eines Fehlers bei der Erhebung) als tot gelten und im nächsten wieder aufscheinen. Für die Jahre bis 2014 ist der Vergleich mit dem ZPR nicht möglich; tauchen hier vermeintlich Verstorbene wieder auf, so werden sie aus dem Datenbestand entfernt. Ab dem Jahr 2018 werden alle im Lag- oder Referenzquartal nicht befragten Personen mit dem ZPR abgeglichen. Ist z.B. eine im MZ nicht mehr befragte Person laut ZPR gestorben, so gilt sie als tot; findet sich im ZPR kein Todesdatum für diese Person, so gilt sie als verzogen. Für Personen ohne bPK ist ein Abgleich mit dem ZPR nicht möglich. Ihr Status muss geschätzt werden, wofür das Alter mit dem durchschnittlichen Sterbealter verglichen wird. Liegt das Alter der Person über diesem Wert, wird sie als verstorben angesehen, ansonsten als verzogen.

Am Ende bleiben nur jene Personen im MZ-Flows-Bestand, bei denen es als hinreichend sicher angesehen wird, dass die Informationen des Lag- und Referenzquartals zur selben Person gehören.

¹²⁾ Derzeit gilt: Für Personen, die zum Zeitpunkt der Stichprobenziehung noch nicht dem MZ-Haushalt angehört haben, ist die Wahrscheinlichkeit, ein bPK zu finden, wesentlich geringer.

¹³⁾ D.h. z.B. Geschlecht, Geburtsdatum und Alter sind in beiden Quartalen vorhanden und gleich. Generell wird beim Vergleich des Alters berücksichtigt, dass eine Person vom Lag- auf das Referenzquartal älter werden kann.

Merkmals-Plaus

Danach folgt die Prüfung der Merkmale ein und derselben Person, die Merkmals-Plaus. Dass die Kombinationen der Merkmalsausprägungen einer Person innerhalb eines Quartals plausibel sind, muss nicht gesondert geprüft werden, da dies im Zuge der Erstellung der authentischen MZ-AKE-Quartalsbestände sichergestellt wird. Die Prüfung der Merkmale bezogen auf die Längsschnittdimension findet bei der Erstellung der Flows-Bestände statt. Hinsichtlich der sogenannten unveränderlichen Merkmale wird dabei der Ansatz verfolgt, dass die aktuellere Information die richtigere ist. Eine Änderung solcher statischer Merkmale in der Interviewsituation ist ein aktiver Vorgang; es wird daher angenommen, dass dieser Aufwand nur dann betrieben wird, wenn eine Korrektur eines früheren (falschen) Wertes notwendig ist. Bei den veränderlichen Merkmalen ist es nicht so einfach, da – abgesehen von einer Herabstufung der höchsten abgeschlossenen Ausbildung – tatsächlich alles möglich ist, auch z.B. eine wenngleich unwahrscheinliche Änderung der Tätigkeit von hochqualifiziert in einem Quartal zu Hilfstätigkeit im folgenden Quartal. Bei der höchsten abgeschlossenen Ausbildung ist es anders; ist die Ausprägung im Referenzquartal „höchstens Pflichtschule“, dann wird der Wert des Lag-Quartals in jedem Fall auch auf „höchstens Pflichtschule“ gesetzt – eine Herabstufung ist nicht möglich und daher auch nicht zulässig.

Am Ende wird noch sichergestellt, dass alle noch nicht Geborenen im Lag-Quartal und alle Verstorbenen im Referenzquartal einen eigenen Missing-Code (-11) bei allen veränderlichen Merkmalen erhalten.¹⁴⁾

Imputation

Der MZ-Flows-Gesamtbestand umfasst auch Personen, die in einem der beide Quartale nicht befragt wurden, d.h. die Informationen über die veränderlichen Merkmale fehlen für ein Quartal vollständig. Kann das Hauptmerkmal ILO-Erwerbsstatus noch relativ leicht mit Hilfe des administrativen Erwerbsstatus geschätzt werden, so ist dies bei den meisten anderen Merkmalen wie z.B. der Arbeitszeit wesentlich schwieriger. Je nach Situation wird unterschiedlich vorgegangen, wie und wodurch ein fehlender Wert ersetzt wird.

Wenn es möglich ist, Informationen aus dem fehlenden Quartal mit Hilfe der Informationen über dieselbe Person aus dem anderen Quartal zu rekonstruieren, dann wird dies gemacht. Ist das nicht der Fall, wird eine möglichst ähnliche Person – ein Spender – ohne fehlende Werte gesucht. Die Merkmalsausprägungen dieser ähnlichen Person werden dann (für den Empfänger) übernommen. Generell wird die Variante der Rekonstruktion bevorzugt. So sind z.B. Bildungsmerkmale relativ stabil über die Zeit: Kennt man also den höchsten Bildungsabschluss in einem Quartal, dann ist es meist besser, basierend darauf den des anderen Quartals zu

¹⁴⁾Durch die Imputation der fehlenden Werte passiert es, dass auch Verstorbene z.B. einen Erwerbsstatus, eine berufliche Stellung usw. imputiert bekommen. Diese Werte werden anschließend wieder gelöscht.

schätzen, als den Bildungsabschluss einer fremden Person zu übernehmen. Nur wenn eine Rekonstruktion nicht möglich ist, wird ein Spender gesucht. Spender und Empfänger sollen einander möglichst ähnlich sein. Bei der Suche eines geeigneten Spenders werden je nach Situation verschiedene Kriterien herangezogen. Zumeist sind dies Geschlecht, Alter, Erwerbsstatus und, je nach zu imputierendem Merkmal, noch weitere Kriterien wie Bildung, berufliche Stellung oder Dauer des Jobs. Kann so kein Spender gefunden werden, dann werden die Kriterien erweitert, d.h. sie werden gelockert, entweder indem Merkmalsausprägungen in gröbere Klassen eingeteilt werden, oder dass auf einzelne Merkmale ganz verzichtet wird. Kommen mehrere Spender in Frage, so wird jener ausgewählt, der dem Empfänger am ähnlichsten hinsichtlich Alter, Bildung und Arbeitsausmaß (Vollzeit/Teilzeit, kurz VZ/TZ genannt) ist.

Für die Imputation werden die veränderlichen Merkmale in vier Gruppen geteilt:

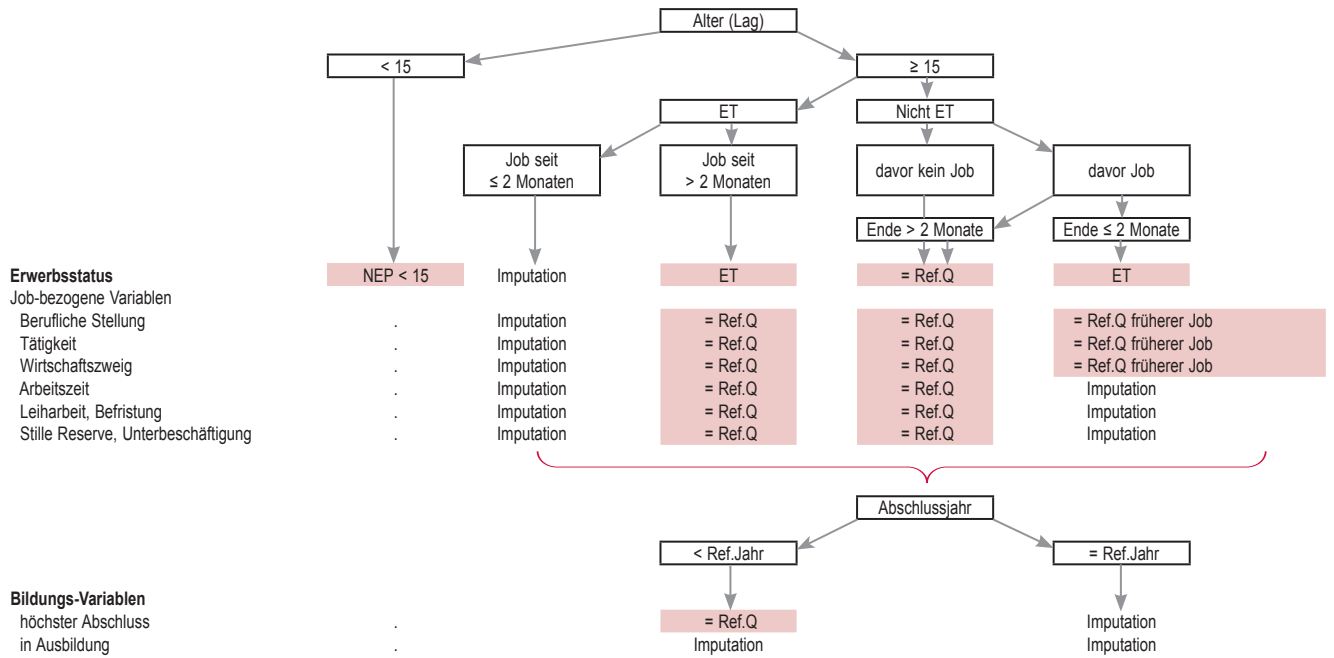
- **Gruppe 1:** übergeordnete Erwerbsmerkmale (ILO-Erwerbsstatus, Berufshauptgruppen, berufliche Stellung, Tätigkeit und Wirtschaftszweig),
- **Gruppe 2:** Jobbezogene und ergänzende Merkmale (VZ/TZ, Stundenausmaß, stille Reserve, Unterbeschäftigung, Leiharbeit, Befristung),
- **Gruppe 3:** höchster (formaler) Bildungsabschluss¹⁵⁾ und
- **Gruppe 4:** in Ausbildung / nicht in Ausbildung.

Sämtliche noch fehlenden Werte einer dieser Merkmalsgruppen werden durch die Werte ein und desselben Spenders ersetzt. Damit ist sichergestellt, dass die Merkmale innerhalb des Quartals stimmig sind – da die vorhandenen Befragungsinformationen aus den authentischen MZ-AKE-Quartalsbeständen stammen, können nur zueinander passende Kombinationen von Merkmalsausprägungen übernommen werden. Das Befüllen der fehlenden Werte erfolgt schrittweise, bereits festgelegte Ausprägungen werden für die Bestimmung weiterer noch fehlender Werte verwendet.

Lag-Quartal Missings

Hier gibt es alle Informationen für das Referenzquartal, für das vergangene Lag-Quartal jedoch keine Information aus der Befragung. Bei einigen Personen gibt es aber für das Lag-Quartal Information aus Verwaltungsdaten (Geschlecht, Alter, Wohnbundesland, administrativer Erwerbsstatus). Bei einigen Merkmalen ist es möglich, mit Hilfe des Wissens über das spätere Quartal auf das vorherige zu schließen – die fehlenden Merkmalsausprägungen werden „rekonstruiert“. Hilft das Wissen über das spätere Quartal nicht, um auf die Vergangenheit zu schließen, so wird die Information von einer anderen (ähnlichen) Person übernommen – die fehlenden Merkmalsausprägungen werden „imputiert“.

¹⁵⁾Ausprägungen: Pflichtschule / keine Pflichtschule; Lehrabschluss (Berufsschule); Berufsbildende Mittlere Schule (ohne Berufsschule); Allgemeinbildende höhere Schule; Berufsbildende höhere Schule; BHS-Abiturientenlehrgänge, Kolleg; Hochschulverwandte Lehranstalt, Universitätslehrgänge; Universität, Fachhochschule.



Rekonstruktion Lag-Quartal

Es kann aus dem späteren Quartal auf das frühere geschlossen werden. Dafür wichtig sind das Alter sowie verschiedene Merkmale der MZ-AKE, wobei einige davon im endgültigen MZ-Flows-Bestand selbst nicht mehr enthalten sind und nur als Hilfsmittel zur Rekonstruktion dienen. Diese Merkmale enthalten z.B. Informationen zum früheren Job. Das konkrete Vorgehen ist an die konkrete Situation angepasst. Eine Übersicht ist in *Abbildung 2* zu finden.

- Personen bis zum Alter von 14 Jahren im Lag-Quartal werden immer als Nicht-Erwerbspersonen (NEP) klassifiziert, und alle ergänzenden und jobbezogenen Merkmale werden auf -3 (per Definition fehlend) gesetzt.
- Ist eine Person im Lag-Quartal 15 Jahre oder älter, im Referenzquartal erwerbstätig (ET) und gibt an, dass sie ihren jetzigen Job seit mehr als zwei Monaten ausübt, wird angenommen, dass die Erwerbs- und Jobsituation (ILO-Status, Arbeitszeit, Tätigkeit, berufliche Stellung, Wirtschaftszweig, usw.) im Lag-Quartal die gleiche war wie im Referenzquartal. Der ILO-Status wird als erwerbstätig angenommen, die Ausprägungen der anderen erwerbs- und jobbezogenen Merkmale des Referenzquartals werden für das Lag-Quartal übernommen.
- Ist eine Person im Lag-Quartal 15 Jahre oder älter und im Referenzquartal nicht erwerbstätig und hatte auch davor keinen Job oder endete dieser vor mehr als zwei Monaten (Information aus der Befragung im Referenzquartal), werden die Informationen aus dem Referenzquartal übernommen (Annahme: die Situation ist unverändert). Hatte die Person aber davor einen Job, der vor höchstens zwei Monaten beendet wurde, dann werden die Informationen, die über diesen früheren Job in der Befragung des Referenzquartals vorhanden sind, so weit wie möglich übernommen, und nicht vorhandene Information werden imputiert.

- Das Bildungsmerkmal „höchste abgeschlossene Bildung“ wird übernommen, wenn das Jahr des Bildungsabschlusses vor dem Referenzjahr liegt. Das Merkmal „in Ausbildung“ wird immer imputiert.

Imputation Lag-Quartal

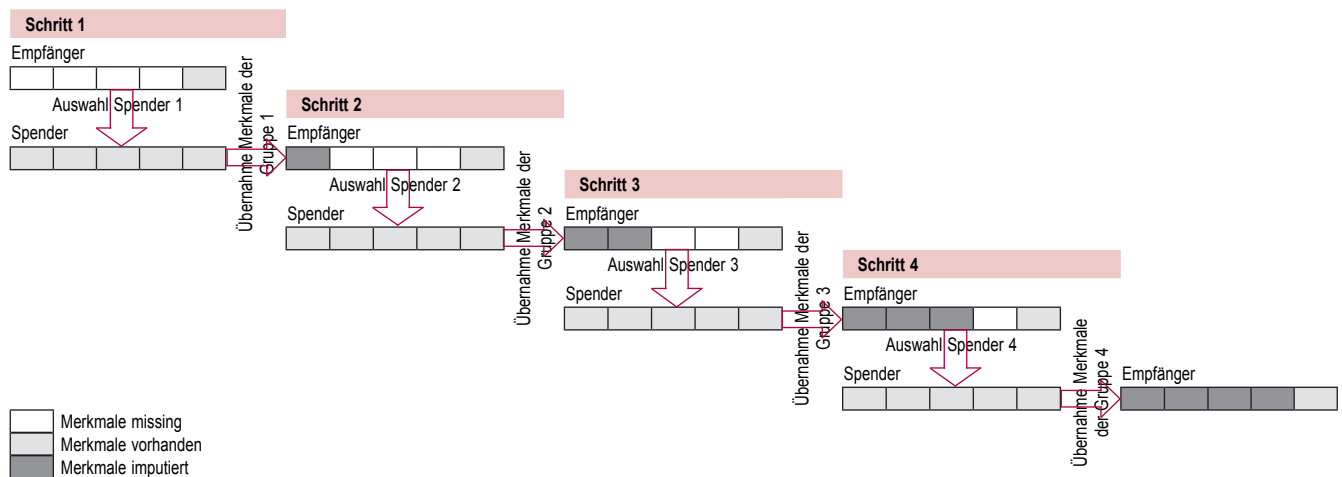
Für alle anderen Personen bzw. Situationen wird ein Spender gesucht, wobei je nach Merkmal bzw. Merkmalsgruppe die Spender anderen Kriterien entsprechen müssen. Gleich ist allerdings immer, dass ein Spender mindestens 15 Jahre alt sein muss, einen Erwerbsstatus aus Verwaltungsdaten (im Referenzquartal) haben muss und auch sonst keine fehlenden Werte haben darf. Eine Übersicht über die Imputation von fehlenden Werten im Lag-Quartal ist in *Abbildung 3* zu finden, die Merkmalsgruppen 1, 2, 3 und 4 sind am Beginn des Kapitels „Imputation“ beschrieben.

- Spender für übergeordnete Erwerbsmerkmale müssen das gleiche Geschlecht, die gleiche Altersklasse¹⁶⁾ und – wenn vorhanden – auch den gleichen administrativen Erwerbsstatus im Lag-Quartal haben. Findet man so keine Spender, wird das Kriterium abgeschwächt, es reicht dann, dass ein aggregierter Erwerbsstatus (mit nur zwei Ausprägungen „erwerbstätig“, „nicht erwerbstätig“) aus Verwaltungsdaten übereinstimmt. Fehlt der Erwerbsstatus aus den Verwaltungsdaten, muss stattdessen der ILO-Erwerbsstatus im Referenzquartal übereinstimmen.

¹⁶⁾ Es werden drei Altersgruppen verwendet; unterschieden werden nur Kinder (bis 14), Personen im Erwerbsalter (15-64) und Ältere (65+).

Imputation fehlender Werte des Lag-Quartals

Abbildung 3



- Spender für jobbezogene und ergänzende Merkmale müssen das gleiche Geschlecht, die gleiche Altersklasse, den gleichen ILO-Erwerbsstatus im Lag-Quartal und die gleiche berufliche Stellung im Lag-Quartal¹⁷⁾ haben. Findet man so niemanden, wird auf die Übereinstimmung der beruflichen Stellung verzichtet.
- Spender für das bildungsbezogene Merkmal „höchste abgeschlossene Ausbildung“ müssen im besten Fall das gleiche Geschlecht, die gleiche Altersklasse, die gleiche höchste abgeschlossene Ausbildung im Referenzquartal und den gleichen ILO-Erwerbsstatus im Lag-Quartal haben. Wenn nötig, wird auf die Übereinstimmung des ILO-Status verzichtet.
- Spender für das bildungsbezogene Merkmal „in Ausbildung / nicht in Ausbildung“ müssen das gleiche Geschlecht, die gleiche Altersgruppe, die gleiche Ausprägung des Merkmals „in Ausbildung“ im Referenzquartal und die gleiche berufliche Stellung im Lag-Quartal haben. Gibt es keine Person, die diesen Kriterien entspricht, wird die berufliche Stellung im Lag-Quartal ignoriert.

Referenzquartal Missings

Hier gibt es alle Informationen für das Lag-Quartal, für das Referenzquartal gibt es keine aus der Befragung, bei einigen Personen gibt es aber zumindest Informationen aus den Verwaltungsdaten.

Rekonstruktion Referenzquartal

Im Gegensatz zum Fall, bei dem die Information der Vergangenheit fehlt, ist eine Rekonstruktion des späteren Quartals mittels Informationen aus der früheren Befragung nur äußerst selten möglich. Eine Übersicht darüber gibt *Abbildung 4*.

- Personen die im Referenzquartal jünger als 15 Jahre sind, sind auch im Referenzquartal Nicht-Erwerbsperson unter 15, alle anderen Merkmale sind per Definition Missing (-3).

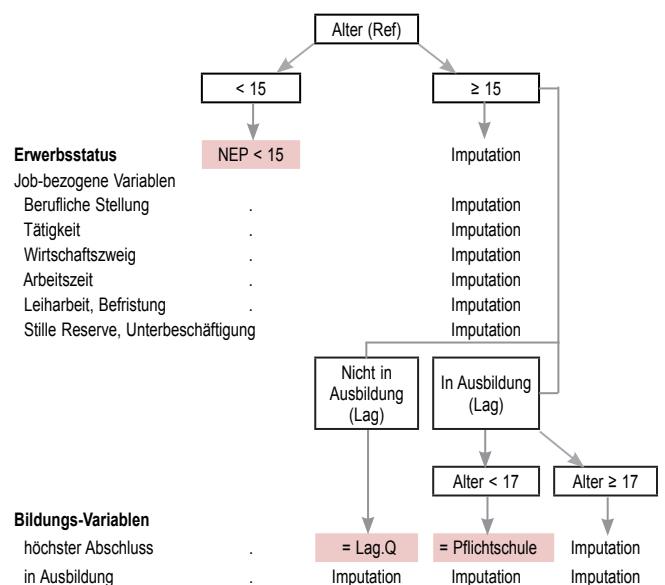
¹⁷⁾ ILO-Status und berufliche Stellung wurden im vorherigen Schritt imputiert und können nun für die Imputation dieser Gruppe von Merkmalen verwendet werden.

- Ist eine Person 15 Jahre oder älter und war sie im Lag-Quartal nicht in Ausbildung, erhält sie im Referenzquartal den höchsten Bildungsabschluss des Lag-Quartals. War sie in Ausbildung und ist sie im Referenzquartal noch jünger als 17, wird ihr als höchste abgeschlossene Ausbildung der Wert „Pflichtschule“ zugewiesen.

In allen anderen Fällen muss auf einen Spender zurückgegriffen werden.

Rekonstruktion fehlender Werte des Referenzquartals

Abbildung 4



Imputation Referenzquartal

Ein Spender kann auch hier nur aus der Gruppe der Personen stammen, die mindestens 15 Jahre alt sind, einen Erwerbsstatus aus Verwaltungsdaten (im Referenzquartal) haben und auch sonst keine Missings aufweisen. Aus allen gefundenen Spendern wird wieder derjenige ausgewählt, der dem Empfänger hinsichtlich Alter, höchster abgeschlossener Ausbildung und Teilzeitarbeit am ähnlichsten ist. Ein we-

sentlicher Unterschied zum Fall der Imputation der Werte in der Vergangenheit ist, dass hier versucht wird, die Wahrscheinlichkeit, dass eine Person ihren Job gewechselt hat, zu imputieren. Je nachdem, ob ein Jobwechsel erwartet wird oder nicht, wird entweder die Information des Spenders aus dem Referenzquartal oder die des Empfängers selbst aus dem Lag-Quartal verwendet. Die Auswahl eines geeigneten Spenders erfolgt im Prinzip gleich wie im Fall der fehlenden Werte im Lag-Quartal, allerdings sind im Fall der Missings im Referenzquartal deutlich weniger Fälle zu rekonstruieren und mehr Fälle zu imputieren, und damit ist auch die Auswahl eines geeigneten Spenders schwieriger. Die Auswahlkriterien für die verschiedenen Merkmalsgruppen werden je nach Erfolg der Spendersuche im Laufe des Prozesses immer weiter ausgeweitet. Eine Übersicht über die Imputation fehlender Werte im Referenzquartal ist in *Abbildung 5* gegeben; die Merkmalsgruppen 1, 2, 3 und 4 sind am Beginn des Kapitels „*Imputation*“ beschrieben.

- Spender für alle erwerbs- und jobbezogenen Merkmale, die ergänzenden Merkmale zur Erwerbstätigkeit und das Bildungsmerkmal „in Ausbildung / nicht in Ausbildung“ müssen für Empfänger **mit Erwerbsstatus aus Verwaltungsdaten** im Referenzquartal das gleiche Geschlecht, die gleiche Altersklasse, den gleichen Erwerbsstatus aus Verwaltungsdaten im Referenzquartal, den gleichen aggregierten ILO-Erwerbsstatus¹⁸⁾ im Lag-Quartal und die gleiche Verweildauergruppe im Lag-Quartal haben. Die Dauer des Erwerbsstatus im Lag-Quartal ergibt sich aus der Dauer des Jobs (falls die Person im Lag-Quartal erwerbstätig ist) bzw. dem Ende des vorherigen Jobs (falls die Person im Lag-Quartal nicht erwerbstätig ist, es davor aber einmal war). Diese Dauer wird in Gruppen (weniger als drei Monate, 3 Monate bis weniger als 1 Jahr, 1 Jahr bis weniger als 5 Jahre, 5 Jahre und mehr) zusammengefasst und ermöglicht so, die Verweildauer bzw. Stabilität des Status in die Auswahl aufzunehmen. Dahinter steht die Idee, dass Personen, die schon sehr lange in einem bestimmten Status verweilen, eine geringe Wahrscheinlichkeit haben, diesen Status zu verlassen. Findet man damit keinen geeigneten Spender, wird statt des detaillierten Erwerbsstatus aus Verwaltungsdaten im Referenzquartal der aggregierte Status (erwerbstätig, nicht erwerbstätig) verwendet. Findet man auch so keinen passenden Spender, genügt eine Übereinstimmung der Merkmale Geschlecht, Altersgruppe und aggregierter Erwerbsstatus aus Verwaltungsdaten im Referenzquartal. Ist auch das nicht erfolgreich und handelt es sich um einen Empfänger, der vom Lag- auf das Referenzquartal 15 Jahre alt wird, sind alle 15-Jährigen im Referenzquartal geeignete Spender.

¹⁸⁾ Die Ausprägungen „Nicht-Erwerbspersonen“ und „Präsenz- und Zivildienstler“ werden hier zu einer Gruppe zusammengefasst. Dieser aggregierte Erwerbsstatus hat die Ausprägungen „Erwerbstätig“, „Arbeitslose“, „Nicht-Erwerbspersonen, inkl. Präsenz- und Zivildienstler“ und „Person <15 Jahre“.

- Spender für alle erwerbs- und jobbezogenen Merkmale, die ergänzenden Merkmale zur Erwerbstätigkeit und das Bildungsmerkmal „in Ausbildung / nicht in Ausbildung“ müssen für Empfänger **ohne Erwerbsstatus aus Verwaltungsdaten** im Referenzquartal das gleiche Geschlecht, die gleiche Altersklasse, den gleichen aggregierten ILO-Erwerbsstatus und die gleiche gruppierte Dauer des Erwerbsstatus im Lag-Quartal haben. Findet man keinen Spender, der bezüglich dieser Merkmale die gleichen Eigenschaften wie der Empfänger hat, wird bei der Auswahl auf die Verweildauergruppe verzichtet.
- Spender für das Merkmal „höchste abgeschlossene Ausbildung“ können im Gegensatz zu allen anderen Imputationsfällen auch einen fehlenden Erwerbsstatus aus Verwaltungsdaten haben. Sie müssen im besten Fall das gleiche Geschlecht, die gleiche Altersklasse, die gleiche höchste abgeschlossene Ausbildung im Lag-Quartal und die gleiche berufliche Stellung im Lag-Quartal haben. Entspricht niemand diesen Vorgaben, wird auf Übereinstimmung der beruflichen Stellung im Lag-Quartal verzichtet.

Je nach Eigenschaft des besten Spenders wird bei der Übernahme der Merkmalsausprägungen für den Empfänger unterschiedlich vorgegangen. Wenn der beste gefundene Spender seit mehr als zwei Monaten seinen Job hat, wird angenommen, dass auch der Empfänger seinen Job schon länger hat, und es werden seine eigenen Merkmalsausprägungen des Lag-Quartals übernommen. Hat der beste Spender seinen Job erst seit höchstens zwei Monaten, oder er hat gar keinen Job, wird auch für den Empfänger eine hohe Wechselwahrscheinlichkeit angenommen. Die Werte des Lag-Quartals werden daher nicht übernommen, und stattdessen werden die Ausprägungen des Spenders übertragen.

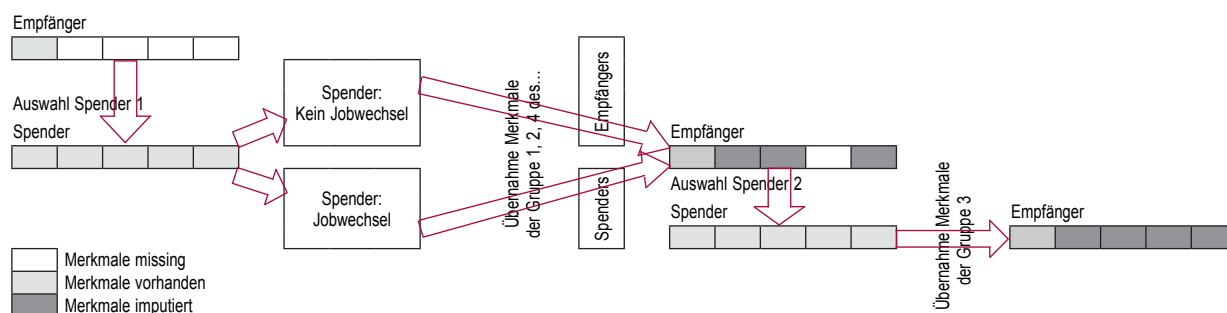
Um eine Vorstellung über die Größenordnungen der fehlenden, rekonstruierten und imputierten Werte zu erhalten, sind die entsprechenden Mengen als Durchschnitt des Jahres 2016 gegeben. Von insgesamt etwa 36.500 Personen im Flows-Bestand fehlt bei knapp 4% die Befragungsinformation aus dem Referenzquartal, bei knapp 3,5% fehlt sie für das Lag-Quartal. Von den Personen mit fehlender Befragungsinformation im Lag-Quartal können zirka 95% der fehlenden Information aus dem Referenzquartal rekonstruiert werden,¹⁹⁾ d.h. es wird allein die Information der Person selbst verwendet, es muss nicht auf andere Personen zurückgegriffen werden. Für die Personen und Merkmale, bei denen eine Rekonstruktion nicht möglich ist, wird der beste Spender ausgewählt, und seine Information wird übernommen. Dabei steht für ungefähr die Hälfte der Empfänger zumindest die Information über den eigenen Erwerbsstatus in Verwaltungsdaten zur Verfügung. Bei den Personen ohne Befragungsinformation im Referenzquartal können nur um die 14% der fehlenden Informationen²⁰⁾ rein mit Hilfe des

¹⁹⁾ Eine Ausnahme stellt nur das Merkmal „in Ausbildung / nicht in Ausbildung“ dar; hier können nur etwa 20% rekonstruiert werden.

²⁰⁾ Mit Ausnahme des Merkmals „höchste abgeschlossene Ausbildung“; hier sind es gut 90%.

Imputation fehlender Werte des Referenzquartals

Abbildung 5



Lag-Quartals rekonstruiert werden. Der Rest wird mit Hilfe geeigneter Spender befüllt. Auch hier können z.T. – wenn ein Statuswechsel unwahrscheinlich ist – die Merkmalsausprägungen des Empfängers selbst verwendet werden.

Gewichtung

Die Flows-Bestände werden mit demselben Verfahren wie die MZ-AKE-Quartalsbestände gewichtet, der iterativen proportionalen Anpassung.²¹ Die MZ-Flows-Stichprobenbestände werden an bekannte Informationen der Grundgesamtheit gebunden. Diese Vorgaben, auch Eckzahlen genannt, werden durch die so hochgerechneten Stichprobenwerte genau abgebildet, wobei eine vorher definierte Abweichung von im Normalfall höchstens 1% zulässig ist.²²

Hochrechnungsvorgaben

Im Gegensatz zu den MZ-AKE-Quartalsergebnissen, bei denen die Vorgaben aus vier Bereichen für jeweils einen Zeitpunkt (das Quartal) stammen (siehe z.B. *Meraner/Gumprecht/Kowarik, 2016*), sind die Vorgaben für die Flows-Bestände aus drei Bereichen, aber jeweils für zwei Zeitpunkte, das Lag- Quartal und das Referenzquartal:

- $N_{ga / lag}$... Geschlecht g ($= 1, 2$) x Altersgruppe a ($= 1, \dots, 6$) mit den Altersgruppen 0-14, 15-24, 25-34, 35-54, 55-64, 65+ im Lag-Quartal,
- $N_{ga / ref}$... Geschlecht g ($= 1, 2$) x Altersgruppe a ($= 1, \dots, 6$) mit den Altersgruppen 0-14, 15-24, 25-34, 35-54, 55-64, 65+ im Referenzquartal,
- $N_{ges / lag}$... Geschlecht g ($= 1, 2$) x Erwerbsstatus nach MZ-AKE e ($= 1, 2, \{3, 4\}, 9$) mit den Ausprägungen „Erwerbstätig“, „Arbeitslos“, „Nicht-Erwerbspersonen, inkl. Präsenz- und Zivildienstler“ und „Personen unter 15“ x Staatsbürgerschaft s ($= 1, 2$) mit den Gruppen „Österreich“ und „Nicht-Österreich“ im Lag-Quartal,
- $N_{ges / ref}$... Geschlecht g ($= 1, 2$) x Erwerbsstatus nach MZ-AKE e ($= 1, 2, \{3, 4\}, 9$) mit den Ausprägungen „Erwerbstätig“, „Arbeitslos“, „Nicht-Erwerbspersonen, inkl. Präsenz- und Zivildienstler“ und „Personen unter 15“ x Staatsbürgerschaft s ($= 1, 2$) mit den Gruppen „Österreich“ und „Nicht-Österreich“ im Referenzquartal und

²¹) Software: R, Paket simPop.

²²) In Einzelfällen war es notwendig, diese Toleranz zu erhöhen, um Konvergenz zu erreichen.

- Bevölkerungsveränderungen in drei Untergruppen. Die Vorgaben dafür basieren ebenfalls auf den Eckzahlen der zwei Quartale (Lag- und Referenz), allerdings werden diese gemittelt. Angepasst wird an den jeweiligen Mittelwert von Lag- und Referenzquartal:
 - Todesfälle (in Privathaushalten)
 - Geburten und
 - Wanderungen (Binnenwanderung und Ein- und Auswanderung in Privathaushalte).

Die **Todesfälle** in Privathaushalten in einem Quartal werden berechnet. Die Gesamtzahl der Gestorbenen im Quartal stammt aus den publizierten Ergebnissen von Statistik Austria.²³ Davon werden die Todesfälle von Personen, die vor ihrem Tod in einer Anstalt gelebt haben, abgezogen. Diese Gestorbenen der sogenannten Anstaltsbevölkerung stellen eine relativ inhomogene Gruppe dar, da die Anstaltsbevölkerung an sich inhomogen ist. Je nach Anstaltstyp (z.B. Pflegeheim, Pensionistenheim, Kinderheim, Obdachlosenheim, Kloster, Studentenheim, Gefängnis, etc.) gibt es ganz unterschiedliche Altersstrukturen und unterschiedliche Sterberaten. Die Anzahl der Gestorbenen in Pflege- und Pensionistenheimen ist für die Jahre 2005 bis 2015 bekannt; sie stellen mit Abstand die wichtigste Gruppe bei den Verstorbenen der Anstaltsbevölkerung dar. Für das Jahr 2015 können mit Hilfe des ZPR auch die Gestorbenen in anderen Anstalten (mit gewissen Unschärfen bzgl. der Anstaltsobjekte selbst) identifiziert werden. Damit können für die Quartale von 2015 die verstorbenen Personen in Privathaushalten einfach berechnet werden (insgesamt Verstorbenen minus Verstorbenen in Anstalten). Für das Jahr 2015 zeigt sich, dass die Verstorbenen in anderen Anstalten etwa 20% der Gestorbenen in Pflege- und Pensionistenheimen entsprechen. Diese Quote wird verwendet, um für die anderen Jahre die Verstorbenen in Privathaushalten zu berechnen. Für die Referenzquartale der Jahre 2008 bis 2014 werden von den insgesamt Verstorbenen je Quartal die Verstorbenen in Pflege- und Pensionistenheimen (diese Zahl ist bekannt) und die geschätzten Verstorbenen in anderen Anstalten (20% der Verstorbenen in Pflege- und Pensionistenheimen) abgezogen.

²³) Siehe Tabelle „Gestorbene seit 2004 nach Ereignismonat und Bundesländern“ auf der Website unter www.statistik.at > Bevölkerung > Gestorbene.

Für die Quartale ab dem Referenzjahr 2016 müssen alle Verstorbenen in Anstaltshaushalten geschätzt werden. Für die geschätzte Anzahl der Verstorbenen in Privathaushalten wird daher mit Hilfe eines ARIMA-Zeitreihenmodells ein Forecast der Quote „Verstorbene in Privathaushalten“ / „Verstorbene insgesamt“ erstellt, basierend auf den Quartalen der Jahre 2005 bis 2015. Die insgesamt Verstorbenen eines Quartals werden mit dieser Quote multipliziert und ergeben so die Anzahl der Verstorbenen in Privathaushalten.

Die Anzahl der **Geburten** je Quartal stammt ab dem Referenzjahr 2016 aus dem ZPR, jeweils zum Stand erster Dienstag des 2. Monats des Folgequartals. Für die Zeitpunkte davor entstammen sie den publizierten Ergebnissen von Statistik Austria.²⁴⁾ Da so gut wie alle Neugeborenen in Privathaushalten leben, wird hier nicht um die äußerst geringe Anzahl von Neugeborenen in Anstaltshaushalten korrigiert.

Unter dem Begriff **Wanderungen** sind alle Personen zusammengefasst, die ihren Wohnsitz ändern, also die sogenannte „mobile Bevölkerung“. Sie müssen in der Hochrechnung gesondert behandelt werden, da sie – genau wie die Verstorbenen und Neugeborenen – nur in einem Quartal im MZ-Flows-Stichprobenbestand enthalten sind. Sie unterscheiden sich strukturell von der immobilien Bevölkerung. Die Personen, die ihren Wohnsitz wechseln, bekommen zwar einen Erwerbsstatus und alle anderen Merkmale imputiert, gehören aber nur teilweise in beiden Quartalen zur österreichischen Bevölkerung in Privathaushalten – nämlich nur dann, wenn es sich um Binnenwanderungen handelt. Daher sind sie nur teilweise in den anderen Vorgaben (Geschlecht x Altersgruppe und Geschlecht x Erwerbsstatus x Staatsangehörigkeitsgruppe) enthalten. Die mobilen Personen der Stichprobe müssen hochgerechnet alle Wanderungen (Binnenwanderungen und Außenwanderungen; Letztere sind Zuzüge aus dem Ausland und Wegzüge ins Ausland) der österreichischen Bevölkerung in Privathaushalten ergeben (mit einer erlaubten Abweichung von 1%). Wie bei den Verstorbenen müssen von allen Wanderungsfällen jene der Anstaltspersonen entfernt werden, um auf die Gesamtzahl der mobilen Bevölkerung in Privathaushalten zu kommen. Eine dahingehende Analyse zeigt, dass 93% der Binnenwanderungen nur Wanderungen von und in Privathaushalte betreffen. Bei den Zu- und Wegzügen aus dem bzw. ins Ausland liegt die Quote bei 84%. Die Gesamtzahlen der Binnen- und Außenwanderungen werden mit diesen Quoten angepasst, um so auf die mobile Bevölkerung in Privathaushalten zu kommen. Diese (vorläufig-)endgültigen²⁵⁾ Wanderungszahlen eines Quartals sind 3,5 Monate nach Quartalsende intern verfügbar.²⁶⁾ Für die Jahre 2005 bis 2015 stehen die endgültigen Wanderungszahlen (Binnenwanderungen, Zuzug aus dem Ausland, Wegzug ins

²⁴⁾ Siehe Tabelle „Lebendgeborene seit 2004 nach Ereignismonat und Bundesländern“ auf der Website unter www.statistik.at > Bevölkerung > Geborene.

²⁵⁾ Für die endgültig-endgültigen Wanderungszahlen wird noch einmal jährlich eine Bereinigung um die Karteileichen durchgeführt.

²⁶⁾ Die Publikation erfolgt 4,5 Monate nach Quartalsende.

Ausland) für alle Quartale zur Verfügung. Für die Quartale des Jahres 2016 gibt es zusätzlich zu den endgültigen Zahlen auch (vorläufig-)vorläufige, welche bereits zwei Wochen nach Quartalsende intern verfügbar sind. Ab dem Berichtsjahr 2017 werden die MZ-Flows-Bestände gemeinsam mit den MZ-AKE-Quartalsbeständen erzeugt. Zum Zeitpunkt der Hochrechnung sind dabei nur die vorläufigen Wanderungszahlen verfügbar. Die Vorgaben für die Hochrechnung werden daher folgendermaßen bestimmt: Die vorläufigen Ergebnisse eines Quartals werden entsprechend dem Verhältnis von vorläufigen zu endgültigen Werten angepasst;²⁷⁾ hier wird für jedes Quartal eine eigene Quote verwendet, da die Wanderungen saisonal unterschiedlich sind.

Iterative proportionale Anpassung

Beim Verfahren der iterativen proportionalen Anpassung wird ausgehend von einem Basisgewicht, welches die Ziehungswahrscheinlichkeit der Samplingeinheit beinhaltet (Kehrwert), ein endgültiges Hochrechnungsgewicht bestimmt. Das Basisgewicht ist das Gewicht bei freier Hochrechnung; es ist für alle Personen und Wohnungen eines Bundeslandes b der Quotient M_b/m_b . Dabei bezeichnet M_b die Anzahl der bewohnten Wohnungen in der Grundgesamtheit im Bundesland b im Lag-Quartal, und m_b die Anzahl der Wohnungen der Netto-Stichprobe im Bundesland b im Lag-Quartal. Auch wenn die Bundeslandinformation nicht in den Vorgaben für die Hochrechnung der MZ-Flows-Bestände explizit vorgegeben wird, so ist sie zumindest indirekt über das Basisgewicht enthalten. Das Basisgewicht wird dann Schritt für Schritt abgeändert,²⁸⁾ um so den jeweiligen Vorgaben zu entsprechen. Jede Anpassung an eine neue Eckzahl führt dazu, dass eine bereits vorher erreichte Anpassung an eine andere Eckzahl wieder schlechter wird. Im Lauf des Prozesses werden die Veränderungen bei den Gewichten immer kleiner, bis sie schließlich konvergieren, d.h. alle Vorgaben bis auf eine tolerierte Abweichung von maximal 1% erfüllt sind. Diese Toleranz ist höher als bei der Hochrechnung der MZ-AKE-Quartalsbestände, da für die MZ-Flows-Bestände durch die Bindung an das Lag- und das Referenzquartal insgesamt mehr Vorgaben erfüllt werden müssen. In einigen Fällen kann es aufgrund der konkreten Stichprobe passieren, dass dieser Prozess nicht konvergiert. In diesen Fällen wurde dann die Toleranz bei der Anpassung an die Eckzahlen erhöht.²⁹⁾

²⁷⁾ Für 2017 stehen die Verhältnisse von vorläufigen und endgültigen Werten für nur ein Jahr zur Verfügung. Im Lauf der Zeit werden immer mehr Quartale mit vorläufigen und endgültigen Werten Eingang in die Berechnung finden.

²⁸⁾ Im Vergleich zur Hochrechnung der MZ-AKE-Quartalsbestände, bei der die Veränderung des Basisgewichts auf den Bereich $[\frac{1}{4} \text{ Basisgewicht}; 4 \times \text{Basisgewicht}]$ eingeschränkt ist, darf sich das Gewicht hier beliebig weit vom Basisgewicht entfernen, um auch in schwierigeren Fällen Konvergenz zu gewährleisten.

²⁹⁾ Das war in den Referenzjahren 2008 bis 2016 für insgesamt drei Flows-Bestände notwendig. Für 2010q2_lag1 und 2015q4_lag1 wurde die Toleranz auf 2% erhöht, für 2011q3_lag1 wurde die Toleranz bei der Vorgabe für die Personen unter 15 Jahren auf 20% erhöht (das ist die Gruppe, die für Analyse von geringster Bedeutung ist), dafür konnte sie bei allen anderen Vorgaben bei 1% bleiben.

Die Gewichte für die MZ-Flows-Bestände, auch Längsschnittgewichte genannt, führen also zu Verteilungen, die gleichzeitig zu den vorgegebenen MZ-AKE-Hauptergebnissen des Lag- und des Referenzquartals passen – passen im Sinne dessen, dass sie nicht mehr als 1% davon abweichen.

Genauigkeit und Fehlerrechnung

Die Fehlerrechnung für die MZ-Flows-Bestände wird, wie bei den MZ-AKE-Quartalsbeständen, mit Hilfe von Replikationsgewichten aus einem Bootstrapverfahren durchgeführt. Für jede Person des MZ-Flows-Bestandes werden zusätzlich zu den Längsschnittgewichten 500 Bootstrapgewichte berechnet. Das Verfahren ist für die MZ-AKE-Quartalsbestände in *Meraner/Gumprecht/Kowarik* (2016) beschrieben.

Die Berechnung des tatsächlichen Fehlers eines Schätzers erfolgt mit Hilfe des *mzR*-Pakets, einem von Statistik Austria entwickelten R-Paket für die schnelle und selbst durchführbare Fehlerrechnung (<https://github.com/statistik/mzR>). Dabei wird der Schätzer mit den 500 Bootstrapgewichten berechnet; die Standardabweichung dieser 500 Bootstrap-Schätzer entspricht dem geschätzten Stichprobenfehler.

Zusammenfassung und Ausblick

Die Ergänzung der MZ-AKE-Quartalsquerschnittsdatenbestände um Quartal-zu-Quartal-Längsschnittbestände stellt eine wichtige Erweiterung des Datenangebots der Statistik

Austria dar. Die Analysemöglichkeiten von Arbeitsmarktdynamiken werden damit deutlich vergrößert und erleichtert. Die Erstellung dieser Längsschnittdaten und der Aufbau einer Zeitreihe ab 2008 sind ein erster wichtiger Schritt. Weiterführend kann z.B. an die Erstellung von Längsschnittbeständen für Jahresveränderungen gedacht werden.

Eine Analyse der Ergebnisse und eine Präsentation wichtiger Indikatoren für die Beschreibung der Dynamik des österreichischen Arbeitsmarktes wird in einem eigenen Artikel in den Statistischen Nachrichten zu einem späteren Zeitpunkt publiziert werden.

Literatur

Bacher, J. / Tamesberger, D. / Leitgöb, H. / Lankmayer, T. (2013): „NEET-Jugendliche: Eine neue arbeitsmarktpolitische Zielgruppe in Österreich“. *WISO* 4/2013, 104-131.

Eurostat (2018): „Labour Market Flow Statistics in the EU“. Online in Statistics Explained: <http://ec.europa.eu> > Eurostat > Statistics Explained (Zugriff: 9.5.2018).

Meraner, A. / Gumprecht, D. / Kowarik, A. (2016): „Weighting Procedure of the Austrian Microcensus using Administrative Data“. *Austrian Journal of Statistics*, 45/9/2016, 3-14, Wien.

Schoiswohl, F. / Wüger, M. (2016): „A Representative Matched Cross-section Survey for Austria – Measuring Worker Flow Dynamics with the Austrian Labour Force Survey“. Department of Economics Working Paper Series, 223. WU Vienna University of Economics and Business, Wien.

Wanek-Zajic, B. (2016): „Registerbasierte Erwerbsverläufe“. Registerbasierte Statistiken 2016, Schnellbericht 10.28, Statistik Austria (Hg.), Wien.

Summary

The Austrian Labour Force Survey (LFS) is part of the microcensus, which is a rotating panel survey of private households. Once a household is selected, it stays in the sample for five consecutive quarters and all persons currently living in such a household are questioned in each of these five quarters. Each quarter one fifth of the total sample rotates in and one fifth rotates out. Persons leaving the household during these five quarters leave the microcensus sample, they are not followed up. Persons moving into a household are part of the sample for the remaining timespan.

Usually analyses and main results of the LFS focus on the cross dimension, i.e. quarters or years. Nevertheless it is worth to have a detailed look on the longitudinal dimension although the panel period is rather short. The analyses of quarter to quarter dynamics open various additional insights into the Austrian labour market. Instead of changes of aggregates only, e.g. the total numbers of employed and unemployed persons, changes on the individual levels, so called “flows”, allow to investigate the dynamics of the labour market. Flows from one status of employment to another come to the fore and allow studying the structure of changes on the aggregate level. Special datasets focusing on the longitudinal dimension of the LFS are needed for analyses of quarter-to-quarter changes. Creation and structure of longitudinal datasets are described in detail in this article. Results from the analysis of the labour market dynamics are described in a separate article.