

# Die neue Hochrechnung des Mikrozensus

---

## Methodenbeschreibung

**MERANER Angelika, GUMPRECHT Daniela, KOWARIK Alexander**

**23.09.2015**

## Einleitung

Der österreichischen Mikrozensus (MZ), mit dem auch die europäische Arbeitskräfteerhebung abgedeckt wird, ist eine wichtige Datenquelle für zentrale Arbeitsmarkt-Indikatoren und zugleich die größte regelmäßig durchgeführte Stichprobenerhebung Österreichs. Sie wird seit den 1970er Jahren durchgeführt und wurde im Laufe der Zeit immer wieder an die sich ändernden Gegebenheiten angepasst. Die letzte große Reform vor der aktuellen Änderung des Hochrechnungsverfahrens 2014 fand im Jahr 2004 statt. Eine der wichtigsten Entwicklungen seither war die zunehmende Verfügbarkeit und Verknüpfbarkeit von Verwaltungsdaten im Zuge der Vorbereitung und Durchführung der Registerzählung 2011. Diese verknüpfbaren Verwaltungsdaten eröffneten eine Reihe von Analyse- und Verbesserungsmöglichkeiten für den MZ. So zeigte eine mit Hilfe von Verwaltungsdaten durchgeführte Non-Response Analyse eine Übererfassung von Erwerbstätigen und eine Untererfassung von Nicht-Erwerbstätigen in der MZ Netto-Stichprobe. Die Verwaltungsdaten können verwendet werden, um die gefundenen Verzerrungen auszugleichen, indem bei der MZ Hochrechnung zusätzlich an einen Erwerbsstatus aus Administrativdaten gebunden wird. Durch diese zusätzliche Bindung an einen Erwerbsstatus aus Verwaltungsdaten wird der Stichprobenfehler verkleinert. Neben diesen methodischen Aspekten, die für eine Neugewichtung unter Einbeziehung von Verwaltungsdaten sprechen, bestehen noch weitere Gründe, warum beim Mikrozensus eine Revision der Gewichtung durchgeführt wird. Diese muss z.B. an die neuen Bevölkerungs- und Haushaltszahlen angepasst werden, die sich durch die Registerzählung 2011 ergeben. Die MZ Ergebnisse werden daher basierend auf dem neuem Hochrechnungsverfahren rückwirkend bis 2004 revidiert. Das verwendete Kalibrierungsverfahren sowie das zur Fehlerrechnung verwendete Bootstrapverfahren, es stellt ebenfalls eine Neuerung im Vergleich zum bisherigen Vorgehen dar, werden in diesem Artikel ebenfalls erläutert.

Der nachfolgende Artikel baut auf dem Methodenbericht aus dem Jahr 2006 (Haslinger & Kytir 2006) und dem Bericht in den Statistischen Nachrichten zum Thema Non-Response im Mikrozensus (Gumprecht & Oismüller 2013) auf, wobei Teile daraus textlich unverändert übernommen wurden, und stellt darüber hinaus die methodischen Änderungen des neuen Hochrechnungsverfahrens 2014 dar.

## Stichprobendesign, Stichprobenziehung und Erhebung

Der österreichische Mikrozensus ist eine Quartalerhebung bei privaten Haushalten in Österreich. Es handelt sich dabei um eine rotierende Stichprobe, d.h. jede Adresse bleibt in fünf aufeinanderfolgenden Quartalen in der Stichprobe, wobei in jedem Quartal ein Fünftel der Haushalte durch neue Haushalte ersetzt wird. Befragt werden alle Personen, die an einer Stichprobenadresse wohnhaft sind. Es besteht für alle Haushaltsmitglieder eine gesetzliche Auskunftspflicht, die in der Erwerbs- und Wohnungsstatistikverordnung (BGBl. II Nr. 111/2010) geregelt ist. Fremdauskünfte sind allerdings zulässig, eine andere erwachsene Person des Haushalts kann über die Zielperson Auskunft erteilen. Das Erstinterview erfolgt persönlich (Computer Assisted Personal Interviewing, CAPI), die Folgeinterviews werden in der Regel telefonisch (Computer Assisted Telephone Interviewing, CATI) abgewickelt. Jedem Haushalt wird eine bestimmte Woche im Quartal zugewiesen (Referenzwoche), auf die sich die wichtigsten Fragen des Frageprogramms beziehen. Die Haushalte werden dabei gleichmäßig auf die Referenzwochen aufgeteilt. Spätestens fünf Wochen nach Ende der

Referenzwoche sollte die Befragung des gesamten Haushalts abgeschlossen sein. Bei der Stichprobe handelt es sich um eine nach Bundesländern geschichtete Zufallsauswahl, wodurch die Ergebnisse auch auf Bundesländerebene repräsentativ sind (siehe Kytir & Stadler 2004, Haslinger & Kytir 2006). Die Stichprobenziehung selbst erfolgt auf Ebene der politischen Bezirke, wobei bei einigen Bezirken auf Grund der geringen Bevölkerungsdichte ein doppelter Auswahlsatz angewandt wird. Für alle anderen Bezirke wird der einfache Auswahlsatz des entsprechenden Bundeslandes verwendet.

Die gezogene Stichprobe des Mikrozensus umfasst pro Quartal etwa 20.000 Wohnungen. Als Auswahlrahmen für die Stichprobe dient das Zentrale Melderegister (ZMR). Vor der Stichprobenziehung werden Anstaltshaushalte und Adressen ausgeschlossen, die bereits in früheren Quartalen (10 Jahre rückwirkend) in der Stichprobe des Mikrozensus enthalten waren. Alle übrigen Privathaushalte mit mindestens einer hauptwohnsitzgemeldeten Person zu Beginn des Quartals, welches vor dem Referenzquartal liegt, dienen letztendlich als Auswahlrahmen für die Stichprobe (siehe Haslinger & Kytir 2006). Da zwischen Stichprobenziehung und Befragung etwa vier bis sechs Monate liegen, kann es auch vorkommen, dass ursprünglich an der Stichprobenadresse gemeldete Personen in der Zwischenzeit, d.h. bis zur Befragung ausgezogen sind. Haben bereits andere Personen die Wohnung bezogen, dann werden diese befragt. Steht die Wohnung zum Befragungszeitpunkt allerdings leer oder handelt es sich mittlerweile um ein Geschäftslokal, ist ein Ausfall zu verzeichnen. Daneben gibt es natürlich noch andere Gründe, warum Personen in einer Wohnung gemeldet sind, dort aber nicht oder nicht mehr leben, und umgekehrt. Auch bei Folgeerhebungen wird entsprechend vorgegangen: Es werden immer jene Personen befragt, die an der Stichprobenadresse wohnhaft sind, unabhängig davon, ob es sich um die Personen der Erstbefragung handelt und unabhängig davon, ob sie auch laut ZMR an dieser Adresse gemeldet sind.

## ILO-Erwerbsstatus und Erwerbsstatus aus Verwaltungsdaten

Das zentrale Merkmal des österreichischen Mikrozensus ist der Erwerbsstatus nach internationaler Definition (ILO - International Labour Organization). Neben diesem international vergleichbaren ILO-Erwerbsstatus gibt es eine Vielzahl anderer Möglichkeiten um einen nationalen Erwerbsstatus zu definieren und zu bestimmen. Eine Variante ist die Definition und Bestimmung eines Erwerbsstatus der nur auf Verwaltungsdaten basiert (Verwaltungsdatenerwerbsstatus („Verwerb“), mehr dazu in Kapitel [“Erwerbsstatus aus Verwaltungsdaten“](#)). Dieses Merkmal ist dann nicht nur für die Befragten, sondern für alle Personen der entsprechenden Grundgesamtheit verfügbar, welche in diesem Fall aus jenen Personen besteht, die in den jeweiligen Administrativdaten aufscheinen. Mit Hilfe eines geeigneten Verknüpfungsschlüssels können Verwaltungsdaten verschiedener Quellen untereinander und mit Befragungsdaten auf Personenebene verknüpft werden. Das bereichsspezifische Personenkennzeichen (bPK) stellt genauso einen Schlüssel dar. Es ist ein anonymisierter Personenschlüssel der dazu dient, unterschiedliche Datenquellen ohne Rückschlussmöglichkeit auf einzelne Personen miteinander zu verknüpfen. Der Verwaltungsdatenerwerbsstatus wird aus einer administrativen Datenquelle bezogen und den Personen der Stichprobe einzeln zugewiesen. Der ILO-Erwerbsstatus wird tatsächlich erfragt.

Der ILO-Erwerbsstatus ist international vergleichbar, aber nur für jene Personen bekannt, die im Rahmen des Mikrozensus befragt worden sind. Der Verwaltungsdatenerwerbstatus hingegen ist für die Grundgesamtheit, also die gesamte Wohnbevölkerung, bekannt, für einen internationalen Vergleich jedoch nicht verwendbar. Möchte man den ILO-Erwerbsstatus für die Wohnbevölkerung bestimmen, so müssen die Befragungsdaten auf die Bevölkerung hochgerechnet bzw. gewichtet werden. Ziel des Hochrechnungsverfahrens für den Mikrozensus (siehe Abschnitt [„Hochrechnung und Hochrechnungsvorgaben“](#)) ist es, die österreichische Wohnbevölkerung möglichst repräsentativ und unverzerrt abzubilden.

## Non-Response Analyse

Wie in Kapitel [„Stichprobendesign, Stichprobenziehung und Erhebung“](#) bemerkt, kommt es im Mikrozensus zu Antwortausfällen. Dieser Non-Response ist zwar sehr klein (2013: 5,7%), da es sich um eine auskunftspflichtige Erhebung handelt, trotzdem bestand die Vermutung, dass sich Nicht-Respondenten/-innen und Respondenten/-innen strukturell hinsichtlich ihres Erwerbsstatus unterscheiden. Die sich daraus ergebende Verzerrung (Bias) konnte durch das bis 2014 verwendete Hochrechnungsverfahren (siehe Haslinger & Kytir 2006) nicht (vollständig) ausgeglichen werden. Dies wurde in einer 2013 durchgeführten Non-Response Analyse bestätigt (siehe Gumprecht & Oismüller 2013). In der Mikrozensus-Arbeitskräfteerhebung, dem Teil des österreichischen Mikrozensus mit dem die europäische Arbeitskräfteerhebung abgedeckt wird, gibt es demnach überproportional viele Erwerbstätige unter den Respondenten/-innen und vergleichsweise zu wenige Nicht-Erwerbstätige, d.h. Arbeitslose und Nicht-Erwerbspersonen.

Für die Analyse des Non-Response hinsichtlich des Erwerbsstatus wurde auf Verwaltungsdaten des Hauptverbandes der Sozialversicherung (HV) und des Arbeitsmarktservice (AMS) zurückgegriffen. Sie erlauben Aussagen über den Erwerbsstatus aus Verwaltungsdaten einer Person, egal ob diese im MZ erfasst ist (Response) oder nicht (= Non-Response). Notwendig dafür ist ein Verknüpfungsschlüssel (bPK), damit kann die Verwaltungsdateninformation auf die gesamte Stichprobe (also inklusive der nicht befragten Personen) aufgebracht werden.

Bei der Non-Response Analyse wurde die Verteilung des Verwaltungsdatenerwerbstatus im Response und Non-Response mit jener der Gesamtstichprobe verglichen. Damit lässt sich ein struktureller Unterschied aufzeigen, der wiederum als Hinweis für eine mögliche Verzerrung der Ergebnisse dient. Zusätzlich dazu wurde die Verteilung des Erwerbsstatus aus Verwaltungsdaten im gewichteten Response (unter Verwendung des früheren Hochrechnungsverfahrens, das bis 2014 angewandt wurde) jener der Wohnbevölkerung gegenübergestellt. Damit lässt sich feststellen, ob die Ergebnisse des MZ einen Bias (zumindest hinsichtlich des Merkmals Verwerb) aufweisen. Die Vergleiche wurden für die österreichische Gesamtbevölkerung, sowie für die Bundesländer und getrennt nach einigen wichtigen soziodemographischen Merkmalen (Geschlecht, Alter, Staatsbürgerschaft) durchgeführt. Die Non-Response Studie ergab zusammenfassend für den MZ eine Übererfassung von unselbständig Beschäftigten mit Standardverträgen sowie Selbständigen bei gleichzeitiger Untererfassung aller anderen Gruppen, das sind Nicht-Standard-Unselbständige (z.B. Geringfügig Beschäftigte), Arbeitslose und Nicht-Erwerbspersonen. Diese Richtung von Über- und Untererfassung zeigt sich auch bei der Analyse der verschiedenen

soziodemographischen Gruppen, wobei das Ausmaß der Verzerrung je nach Merkmal variiert. Die größten Unterschiede ergaben sich in Wien sowie bei Personen mit nichtösterreichischer Staatsangehörigkeit. Die Gründe und Ursachen für die unterschiedliche Erreichbarkeit bzw. Bereitschaft zur Teilnahme an der Befragung von Erwerbstätigen und Nicht-Erwerbstätigen waren nicht Gegenstand dieser Non-Response-Studie, die Unterschiede hinsichtlich der Erwerbsbeteiligung von Respondenten/-innen und Nicht-Respondenten/-innen wurden jedoch deutlich.

Wie die Non-Response Analyse zeigt, sind im MZ die Erwerbstätigen tendenziell über- und Nicht-Erwerbstätige etwas untererfasst. Aus Abbildung 1 wird ersichtlich, dass das frühere Hochrechnungsverfahren (HR alt) diese Untererfassung der administrativen Arbeitslosen nicht bzw. nicht vollständig ausgleichen konnte. Es ist zu erkennen, dass sowohl die ungewichteten als auch die nach alter Methode gewichteten Daten des MZ die Eckzahl bzw. die Verteilung für die Gesamtbevölkerung, nicht korrekt zeigen. Das Muster im Quartalsverlauf wird zwar sehr gut wiedergegeben, allerdings auf einem zu niedrigen Niveau: Durch das alte Gewichtungsverfahren liegen die Ergebnisse näher an der Eckzahl als ohne Hochrechnung (ungewichtet), das tatsächliche Niveau in der Grundgesamtheit, welches für dieses Merkmal aus den Verwaltungsdaten bekannt ist (Eckzahlen), wird dennoch unterschätzt. Möchte man das Merkmal Erwerbsstatus aus Verwaltungsdaten korrekt darstellen, so kann man die aus der Grundgesamtheit bekannten Vorgaben in die Hochrechnung einbeziehen (siehe Kapitel „[Hochrechnung und Hochrechnungsvorgaben](#)“). Die Gewichte werden so gebildet, dass die hochgerechneten Werte des Verwaltungsdatenerwerbsstatus der Stichprobe exakt denen der Wohnbevölkerung entsprechen. Die Hochrechnung mit Bindung an den Erwerbsstatus aus Administrativdaten (HR neu) liefert per Definition das gleiche Ergebnis wie die Eckzahlen selbst, wodurch in Abbildung 1 die grüne Linie (HR neu) durch die schwarze Linie (Eckzahlen) überdeckt wird.

Für die Arbeitslosen nach ILO Definition zeigt sich das gleiche Bild. Bereits die alte Hochrechnung (HR alt ILO) führt im Vergleich zu den ungewichteten Werten (ungewichtet ILO) zu einer Erhöhung der Arbeitslosenquote. Die Bindung an den Erwerbsstatus aus Verwaltungsdaten (HR neu ILO) führt abermals zu einer stärkeren Verschiebung nach oben.

Diese Darstellung veranschaulicht ebenso wie die Non-Response-Analyse, dass das bis 2014 verwendete Hochrechnungsverfahren einen Ausgleich in die richtige Richtung aber keinen vollständigen Ausgleich der unterschiedlichen Erfassung von administrativ Erwerbstätigen und Nicht-Erwerbstätigen erzielt. Die Ergebnisse sind hinsichtlich des Merkmals Verwerb verzerrt. Dies beweist zwar nicht eine Verzerrung des ILO-Erwerbsstatus, durch die hohe Korrelation von Verwaltungsdaten- und ILO-Erwerbsstatus kann diese aber abgeleitet werden. Es kann also davon ausgegangen werden, dass auch nach ILO Definition Erwerbstätige überrepräsentiert und Arbeitslose sowie Nicht-Erwerbspersonen etwas untererfasst sind.

## Österreich Gesamt

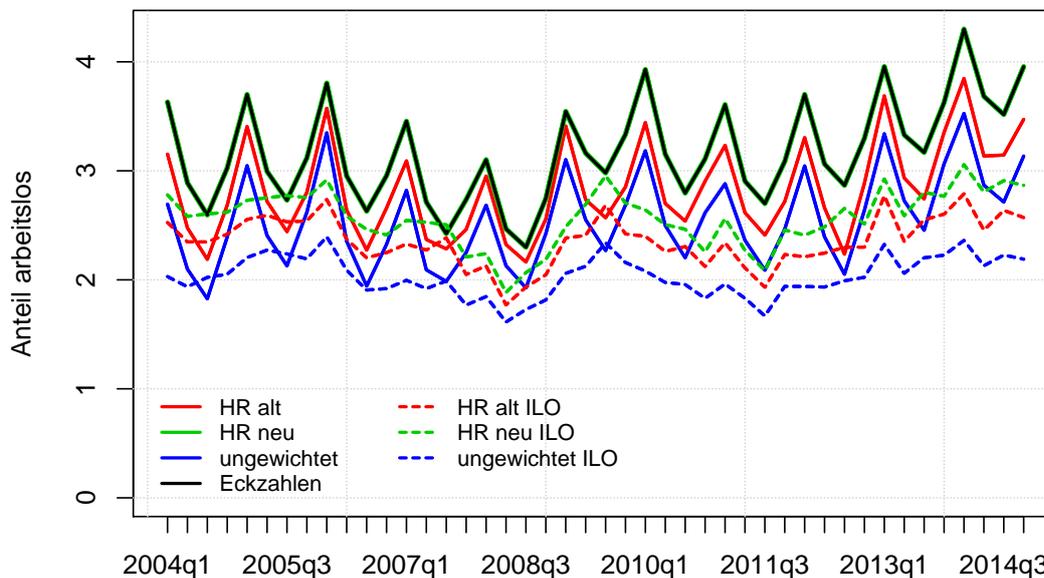


Abbildung 1: Anteil der Arbeitslosen an der Wohnbevölkerung in Privathaushalten laut Verwaltungsdaten und ILO Definition.

### Hochrechnung und Hochrechnungsvorgaben

Bei einer zufallsgesteuerten Stichprobenauswahl wird ein verkleinertes, aber repräsentatives Abbild der Grundgesamtheit geschaffen. Bei der Berechnung der statistischen Ergebnisse muss diese Verkleinerungsprozedur wieder rückgängig gemacht werden. Dieser Vorgang heißt Hochrechnung. Dabei werden die mit Hilfe der Stichprobe erhobenen Merkmalswerte zur Schätzung der unbekannt Parameter der Grundgesamtheit herangezogen. Dem Mikrozensus liegt eine geschichtete Stichprobe mit unterschiedlichen Auswahlsätzen pro Bundesland zugrunde. Um aus den Erhebungsdaten der Stichprobe wieder ein verzerrungsfreies Gesamtbild zu erzeugen, müssen die Erhebungswerte nach Schichten getrennt hochgerechnet werden, im einfachsten Fall durch Multiplikation mit dem Kehrwert des jeweiligen Auswahlsatzes. Diese Hochrechnungsgewichte oder -faktoren werden den entsprechenden Beobachtungen der Stichprobe zugewiesen. Bei der Tabellierung der Mikrozensusergebnisse wird jeder einzelne Merkmalswert vor der Addition mit dem Hochrechnungsgewicht multipliziert.

Zusätzlich wird bei der Hochrechnung an bekannte Informationen der Grundgesamtheit gebunden, um so die Stichprobe dahingehend zu gewichten, dass diese bekannten Eckzahlen exakt abgebildet werden. Merkmale, deren Verteilungen in der Grundgesamtheit aus anderen Quellen wie z.B. der Statistik des Bevölkerungsstandes o.ä. bekannt sind, sind beispielsweise Bundesland, Geschlecht, Alter und Staatsangehörigkeit. Durch die Bindung dieser Merkmale an ihre bekannte Verteilung der Grundgesamtheit sollen auch andere Merkmale, deren Verteilungen in der Grundgesamtheit nicht bekannt sind, unverzerrt oder zumindest unverzerrter

dargestellt werden. Theoretisch können alle Merkmale, die sowohl in der Stichprobe als auch in der Grundgesamtheit bekannt sind, für diese Hochrechnung verwendet werden. Praktisch ergeben sich Beschränkungen z.B. durch die Stichprobengröße. Die Personen der Stichprobe erhalten ein bestimmtes Hochrechnungsgewicht, so dass diese Eckzahlen erreicht werden. Das Gewicht wird mittels des Verfahrens der iterativen proportionalen Anpassung bestimmt.

## Gewichtungsvorgaben

Für die Hochrechnung werden folgende Gewichtungsvorgaben verwendet:

$N_{bga}$  ... Anzahl der Personen in Privathaushalten in Bundesland  $b$  ( $=1, \dots, 9$ )  $\times$  Geschlecht  $g$  ( $=1, 2$ )  $\times$  Altersklasse  $a$  ( $=1, \dots, 19$ ) mit den Altersgruppen 0-2, 3-5, 6-9, dann 5-jährige Altersgruppen (10-14, ..., 80-84) und 85+.

$N_{bs}$  ... Anzahl der Personen in Privathaushalten in Bundesland  $b$  ( $=1, \dots, 9$ )  $\times$  Staatsbürgerschaft  $s$  ( $=1, \dots, 6$ ) mit den Nationalitätengruppen "Österreich", "EU-15 (ohne Ö)", "EU ab 2004 (d.h. im Zeitraum 2004 bis 2014 der EU beigetreten)", "Europäische Nicht-EU-Staaten (ohne Türkei)", "Türkei", "Rest".

$N_{bgv}$  ... Anzahl der Personen in Privathaushalten in Bundesland  $b$  ( $=1, \dots, 9$ )  $\times$  Geschlecht  $g$  ( $=1, 2$ )  $\times$  Erwerbsstatus lt. Verwaltungsdaten „Verwerb“  $v$  ( $=1, \dots, 5$ ) mit den Ausprägungen "Erwerbstätig, unselbständig standard", "Erwerbstätig, unselbständig nicht-standard", "Erwerbstätig, selbständig", "Arbeitslos" und "Nicht-Erwerbsperson".

$M_{bh}$  ... Anzahl der Privathaushalte in Bundesland  $b$  ( $=1, \dots, 9$ )  $\times$  Haushaltsgröße  $h$  ( $=1, \dots, 5$ ) mit den Ausprägungen 1, 2, 3, 4, 5+.

Die Merkmale Geschlecht, Alter und Nationalität (Staatsbürgerschaft) stammen aus dem POPREG (bevölkerungsstatistische Datenbank), das Merkmal Verwerb aus HV- und AMS-Daten (mehr dazu in „[Erwerbsstatus aus Verwaltungsdaten](#)“). Die Haushaltsgrößen werden auf Basis der Registerzählung 2011 und in weiterer Folge der jährlich erstellten abgestimmten Erwerbsstatistik ermittelt (siehe Abschnitt „[Haushaltsgröße](#)“).

## Anstaltshaushalte

Da im POPREG alle Personen, die in Österreich wohnen, enthalten sind, der MZ aber nur die Personen in Privathaushalten abdecken soll, muss die Anstaltsbevölkerung (Pflegeheime, Gefängnisse, Internate usw.) abgezogen werden. Dies geschieht über Anstaltsquoten für alle Ausprägungskombinationen der Eckzahlen. Die Anstaltsquoten wurden für jene verfügbaren Daten, für welche eine konsistente Art und Weise der Berechnung gewährleistet werden kann, ermittelt, d.h. für die Volkszählung zum 15.05.2001, die Anstaltenerhebung zur Registerzählung zum 31.10.2011 und die Anstaltenerhebung zum 31.10.2012. Zwischen diesen Stichtagen werden die Quoten interpoliert, ab der Anstaltenerhebung zum 31.10.2012 werden sie konstant gehalten bis die Ergebnisse für den Stichtag 31.10.2013 zur Verfügung stehen (spätestens 2 Jahre nach dem Stichtag), die dann wieder konstant gehalten werden usw. Die Anstaltenerhebung wird mittlerweile jährlich unter denselben Voraussetzungen durchgeführt und erlaubt so eine zeitlich verzögerte, aber jährliche Anpassung der Anstaltsquoten.

## Nationalität

Mit dem EU-Beitritt Kroatiens war eine Anpassung der der Gewichtung zugrundeliegenden Staatsangehörigkeitsgruppen erforderlich, um im Mikrozensus weiterhin die Differenzierung in EU- und Nicht-EU-Staatsangehörige abbilden zu können. Bislang wurden fünf Nationalitätengruppen ("Österreich", "EU-15 (ohne Ö)", "Ehem. Jugoslawien", "Türkei", "Rest") verwendet, in Folge sind es sechs: "Österreich", "EU-15 (ohne Ö)", "EU ab 2004 (d.h. im Zeitraum 2004 bis 2014 der EU beigetreten)", "Europäische Nicht-EU-Staaten (ohne Türkei)", "Türkei" und "Rest". Die Gruppe "Ehem. Jugoslawien" wird folglich auf die Gruppen "EU ab 2004" und "Europäische Nicht-EU-Staaten (ohne Türkei)" aufgeteilt.

## Haushaltsgröße

Ab 2004 wurden bei der Hochrechnung zusätzlich zur Bindung an demographische Merkmale auch die sich aus dem Mikrozensus ergebenden Verteilungen der Wohnungen (=Haushalte) nach der Zahl der Bewohner an die Haushaltsprognose der Statistik Austria angepasst. Diese zusätzliche Restriktion wurde eingeführt, weil die Antwortausfälle im Mikrozensus umso höher sind, je weniger Personen in einer Wohnung leben. Als Basis für die Prognose der Verteilungen der Haushaltsgrößen je Bundesland dienten zu Beginn die Volkszählung 2001 und anschließend die Registerzählung 2011. Im ersten Schritt wurde die Anzahl der Haushalte pro Bundesland für das aktuelle Quartal ermittelt. Dies geschah durch Multiplikation der prognostizierten Anzahl der Personen pro Altersgruppe, Geschlecht und Bundesland mit den entsprechenden Haushaltsrepräsentantenquoten (das ist die Gesamtzahl aller privaten Haushalte dividiert durch die Gesamtbevölkerung in Privathaushalten.). Ein stufenweises Verfahren (siehe Ediev 2007) ermittelte anschließend die Verteilung der Haushalte nach Größe und Bundesland.

Bei der neuen Hochrechnung ab 2014 wird künftig ebenfalls an die Verteilung der Haushaltsgrößen gebunden. Im Unterschied zur Hochrechnung seit 2004 stehen mit der abgestimmten Erwerbstatistik jährliche Registerdaten für die Verteilungen der Haushalte nach Größe und Bundesland zur Verfügung. Die Veröffentlichung der Registerdaten geschieht mit einer Verzögerung von maximal zwei Jahren. Somit wird die bisherige, zweistufige Prognose der Verteilungen der Haushalte nach Größe und Bundesland beibehalten, sie basiert aber auf einer bedeutend aktuelleren und umfangreicheren, da jährlichen, Datenbasis. Darüber hinaus geht die über jährlich angepasste Anstaltsquoten ermittelte Bevölkerung in Privathaushalten in die Berechnungen ein.

## Erwerbsstatus aus Verwaltungsdaten

Die Verwaltungsdaten des Hauptverbandes der österreichischen Sozialversicherungsträger (HV) und des Arbeitsmarktservice Österreich (AMS), wie sie im Erwerbspersonenregister für die abgestimmte Erwerbsstatistik (siehe Statistik Austria 2010) gespeichert sind, können für die Bildung eines Erwerbsstatus aus Verwaltungsdaten und damit für die Bestimmung von Hochrechnungsvorgaben für den Erwerbsstatus verwendet werden. Sie stellen die Verteilung der Grundgesamtheit, an die die Stichprobenergebnisse zusätzlich gebunden werden können, dar. Wichtig ist hier, dass das gleiche Merkmal Verwerb, an das gebunden werden soll, in der Stichprobe und der Grundgesamtheit vorhanden ist. Vorsicht ist beim Begriff Grundgesamtheit geboten. Die Grundgesamtheit für den MZ ist die österreichische Wohnbevölkerung in

Privathaushalten, diese ist (gemeinsam mit der Anstaltsbevölkerung, welche für die Bestimmung der Hochrechnungsvorgaben entfernt werden muss, siehe Kapitel „[Anstaltshaushalte](#)“) im POPREG enthalten. In den Daten von HV und AMS stellen die in Österreich versicherten bzw. beim HV und/oder AMS gemeldeten Personen die Grundgesamtheit dar, das können durchaus auch im Ausland lebende Personen sein. Andererseits sind Personen, welche in Österreich leben, aber im Ausland beschäftigt sind, nicht notwendigerweise in den Daten der österreichischen Sozialversicherung enthalten. Um also eine korrekte Vorgabe für die Hochrechnung zu erhalten, muss aus den Daten von HV und AMS die österreichische Wohnbevölkerung selektiert werden. Über einen Verknüpfungsschlüssel, das bereichsspezifische Personenkennzeichen (bPK), können die Personen aus den HV- und AMS-Daten ausgewählt werden, die der österreichischen Wohnbevölkerung angehören. Die Personen der Wohnbevölkerung, die laut Administrativdaten weder erwerbstätig noch arbeitslos sind, stellen die Menge der Nicht-Erwerbspersonen dar. Über das bPK kann jeder MZ Person ihre HV- und AMS-Information (sofern vorhanden), welche die Basis für das Merkmal Verwerb ist, zugewiesen werden. Somit ist die Bestimmung des Verwaltungsdatenerwerbsstatus auch auf der Stichprobenebene möglich. Ist für eine Person kein bPK vorhanden (rund 6% der Stichprobe) wird das Merkmal Verwerb mittels Hot-Deck Verfahren imputiert. Für die Verknüpfung mit den MZ Personen über das bPK werden die HV- und AMS-Informationen der jeweiligen MZ-Referenzwoche verwendet.

Die HV- und AMS-Daten der Grundgesamtheit beziehen sich immer auf das Monatsende, sie stellen Monatsbestände dar. Das POPREG hingegen hat als Referenztag immer den Quartalsbeginn. Zum Zeitpunkt der Erstellung der Eckzahlen für die Hochrechnung liegen aus dem POPREG keine endgültigen Informationen vor, es wird ein vorläufiger POPREG-Stand verwendet. Dieser ist etwa 1,5 Monate nach Quartalsbeginn verfügbar, berücksichtigt jedoch weder eine Mindestdauer für den Aufenthalt in Österreich (bei z.B. 90 Tagen Mindestaufenthaltsdauer sind es ca. 3.000 Personen weniger) noch Bereinigungen von in der jährlichen Einwohnerzahl für den Finanzausgleich identifizierten Karteileichen (ca. 8.000 Personen weniger). Die Abweichungen bewegen sich also in Summe etwa bei 1,3 Promille der Gesamtbevölkerung.

Es herrscht somit eine gewisse zeitliche Diskrepanz zwischen den soziodemographischen und den Erwerbsmerkmalen. Um aus den HV- und AMS-Daten das Merkmal Verwerb für die Hochrechnung bilden zu können, müssen die HV- und AMS-Monatsbestände an die Wohnbevölkerung laut POPREG angepasst werden. Dabei werden einerseits die Personen, die nicht zur Wohnbevölkerung gehören, entfernt, und andererseits werden die Personen der Wohnbevölkerung, die weder bei HV noch AMS gemeldet sind, hinzugefügt. Dadurch ergibt sich ein Problem, welches im Laufe eines Quartals größer wird: Personen, die während eines Quartals ein- oder auswandern, sind in den Versicherungsdaten von HV und AMS möglicherweise schneller enthalten oder entfernt als dies im POPREG der Fall ist. Auch Binnenwanderungen, sofern eine Bundeslandgrenze überschritten wird, können bei den Bundesländerergebnissen zu Verzerrungen führen, die im Laufe eines Quartals von Monat zu Monat größer werden.

Schließlich wird aus den Monatsendständen ein Mittelwert gebildet, um auf die Monats- bzw. Quartalsvorgaben für den Erwerbsstatus zu kommen. Für die Ermittlung der Monatseckzahlen werden der Referenzmonatsendstand und der Endstand des Vormonats herangezogen (Monats-Wohnbevölkerungsfiles bzw. „Monats-WB-Files“, siehe Abbildung 2) um den Mittelwert zu bilden. Wenn der Vormonat im vorherigen Quartal liegt, dann werden die HV- und AMS-Vormonatsdaten an den aktuellen POPREG

Quartalsstand angepasst (a-File), siehe Abbildung 2. Die anderen Hochrechnungsvorgaben (Bundesland, Geschlecht, Nationalität) werden direkt aus dem POPREG Quartalsanfangsbestand gebildet.

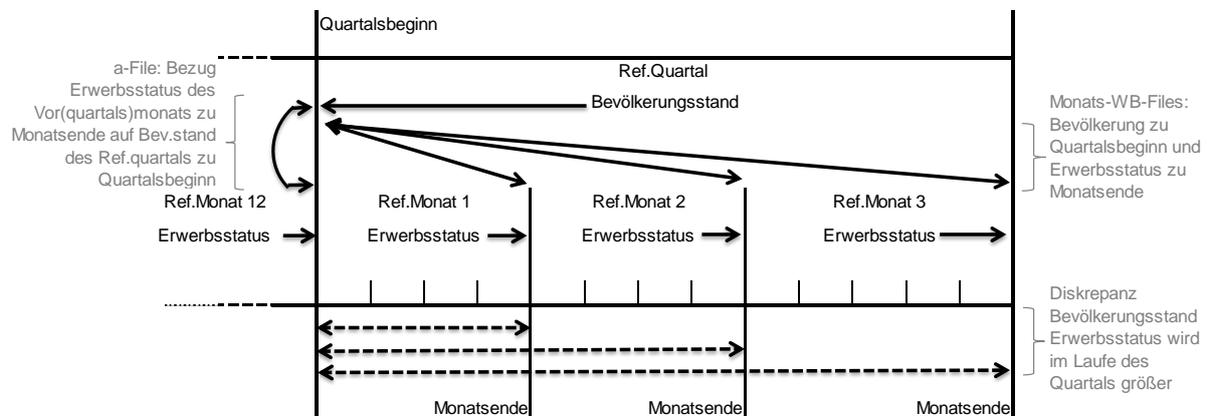


Abbildung 2: Bezüge von Bevölkerungs- und Erwerbsdaten auf Quartals- und Monatsstände.

### Bildung des Merkmals Verwerb

Das Merkmal Verwerb wird rein aus Administrativdaten von HV (Qualifikationen) und AMS (Vormerkungen) gebildet. Es hat fünf Ausprägungen:

1. Erwerbstätig, unselbständig standard (aus HV-Qualifikationen gebildet: Beamte, Angestellte, Arbeiter, Zivildienst, Präsenzdienst)
2. Erwerbstätig, unselbständig nicht-standard (aus HV-Qualifikationen gebildet: Freie Dienstnehmer, Geringfügige Angestellte, Geringfügige Arbeiter, sonstiges Geringfügig, Teilpension, Gleitpension, Karenz, Krankengeld, Zeitsoldat, Karenz und sonstige Abwesenheit)
3. Erwerbstätig, selbständig (aus HV-Qualifikationen gebildet: Selbständige in der Landwirtschaft, Mithelfende in der Landwirtschaft, sonstige Selbständige, Freiberufler)
4. Arbeitslos (aus AMS Vormerkungen gebildet: Arbeitslose Personen, Lehrstellensuchende Personen, Teilnehmerinnen und Teilnehmer an Schulungen, Arbeitsuchende Personen)
5. Nicht-Erwerbsperson (Restkategorie): Personen der österreichischen Wohnbevölkerung, die keiner anderen Erwerbskategorie angehören, dazu zählen u.a. Kinder, Pensionisten, ausschließlich haushaltsführende Personen etc.

Jede Person kann nur eine dieser fünf Ausprägungen besitzen weshalb bei der Bildung des Merkmals Dominanzkriterien eingehalten werden, gemäß denen eine Erwerbstätigkeit immer vor der Arbeitslosigkeit gereiht wird.

### Hochrechnung

#### Basisgewichte

Die Ermittlung der endgültigen Hochrechnungsgewichte erfolgt in einem iterativen Prozess ausgehend von den Gewichten bei freier Hochrechnung. Letztere berechnen sich in der Regel für alle Personen und Wohnungen eines Bundeslandes  $b$  als der Quotient

$$\frac{M_b}{m_b} \quad (1)$$

Dabei bezeichnet  $M_b$  die Anzahl der bewohnten Wohnungen in der Grundgesamtheit im Bundesland  $b$  und  $m_b$  die Anzahl der Wohnungen der Nettostichprobe im Bundesland  $b$ .

In einigen wenigen Bezirken wurde der Auswahlsatz im Vergleich zu den restlichen Bezirken des jeweiligen Bundeslandes verdoppelt, um auch in diesen relativ dünn besiedelten Bezirken eine genügend hohe Auslastung der Interviewer/-innen zu erzielen. Dadurch ergibt sich in diesen Bezirken das halbe Gewicht. Ist ein Bundesland von dieser Verdoppelung des Auswahlsatzes betroffen so werden die Hochrechnungsgewichte für Personen und Wohnungen aus Bezirken mit normalem Auswahlsatz wie folgt berechnet:

$$\frac{M_b}{m_{b1} + m_{b2}/2} \quad (2)$$

$m_{b1}$  bezeichnet dabei die Anzahl der Wohnungen der Nettostichprobe in den Bezirken des Bundeslandes  $b$  mit normalem Auswahlsatz und  $m_{b2}$  bezeichnet die Anzahl der Wohnungen der Nettostichprobe in den Bezirken des Bundeslandes  $b$  mit doppeltem Auswahlsatz. Für Personen und Wohnungen aus Bezirken mit doppeltem Auswahlsatz berechnet sich das Hochrechnungsgewicht als die Hälfte dieses Quotienten (2).

### Anpassung an vorgegebene Bevölkerungsstrukturen

Die mittels der Gewichte (1) - (2) errechneten Bevölkerungszahlen weichen hinsichtlich der Verteilung nach Alter, Geschlecht, Nationalität und Verwaltungsdatenerwerbsstatus in der Regel von den entsprechenden Zahlen des mit HV- und AMS-Daten verknüpften Bevölkerungsstands aus dem POPREG ab.

Wie eingangs bemerkt, haben viele Auswertungen des Mikrozensus die Berechnung von Umfängen von speziellen Personengesamtheiten zum Ziel, so etwa die Anzahl an unselbständig Erwerbstätigen im Gesundheits- und Sozialwesen, um nur ein Beispiel zu nennen. Diese Schätzwerte korrelieren sehr eng mit der Bevölkerungszahl. Es lässt sich daher die Genauigkeit dieser Schätzwerte erhöhen, indem man die aus der Stichprobe errechneten Bevölkerungszahlen an die aus dem POPREG und den entsprechenden HV- und AMS-Daten bekannten Eckzahlen anpasst. Dies wird erreicht, indem man die Hochrechnungsgewichte aus den Formeln (1) - (2) in einem iterativen Verfahren so abändert, dass die darauf berechneten Mikrozensusergebnisse mit den vorgegebenen Bevölkerungsstrukturen übereinstimmen.

### Iterative Proportionale Anpassung

Für die Zwecke der Iteration werden für jede Person  $i$  der Stichprobe die Merkmalsausprägungen für das Bundesland  $b$  ( $=1, \dots, 9$ ), das Geschlecht  $g$  ( $=1, 2$ ), das Alter  $a$  ( $=1, \dots, 19$ ), die Staatsbürgerschaft  $s$  ( $=1, \dots, 6$ ), den Verwaltungsdatenerwerbsstatus  $v$  ( $=1, \dots, 5$ ), die fortlaufende Wohnungsnummer  $j$  (Kombination aus Pol. Bezirk und Wohnungskennzahl) und das vorläufige Hochrechnungsgewicht aus (1) - (2) benötigt (siehe Kapitel „[Gewichtungsvorgaben](#)“).

Wir bezeichnen dieses Gewicht mit  $w_i^0$ . Die Iteration besteht aus fünf sich wiederholenden Schritten. Zum Start wird der Laufindex  $k=0$  gesetzt. Ist die Bedingung in der Abfrage nach Schritt 5 erfüllt, wird  $k$  um 1 erhöht und das Verfahren beginnt wieder mit Schritt 1.

### Schritt 1-3

Das nach Formel (1) - (2) errechnete Gewicht jeder Person wird durch Multiplikation mit einem Faktor so abgeändert, dass die mit den abgeänderten Gewichten hochgerechnete Verteilung der Bevölkerung in jedem Iterationsschritt mit der jeweiligen Gewichtungsvorgabe  $N_e \in \{N_{bga}, N_{bs}, N_{bgv}\}$  (siehe „[Gewichtungsvorgaben](#)“) übereinstimmt. Die Gewichtungsvorgabe  $N_e$  mit  $e \in \{bga, bs, bgv\}$  variiert demnach je nach Iterationsschritt  $t$ :

Falls  $t=1$ :  $N_e = N_{bga}$

Falls  $t=2$ :  $N_e = N_{bs}$

Falls  $t=3$ :  $N_e = N_{bgv}$

Das an die jeweilige Vorgabe  $N_e$  gebundene Gewicht wird in jedem Schritt  $t$  wie folgt berechnet:

$$w_i^{5k+t} = w_i^{5k+t-1} \frac{N_{e_i}}{\sum_l w_l^{5k}}$$

für alle  $i$ .

Die Summierung erstreckt sich dabei für  $e_i = b_i g_i a_i$  über alle Beobachtungen  $l$  mit denselben Ausprägungen der Merkmale  $b$ ,  $g$  und  $a$  wie die Beobachtung  $i$ , für  $e_i = b_i s_i$  über alle Beobachtungen  $l$  mit denselben Ausprägungen der Merkmale  $b$  und  $s$  wie die Beobachtung  $i$  und für  $e_i = b_i g_i v_i$  über alle Beobachtungen  $l$  mit denselben Ausprägungen der Merkmale  $b$ ,  $g$  und  $v$  wie die Beobachtung  $i$ . Falls der Quotient  $\frac{w_i^{5k+t}}{w_i^0} > 4$  ist, wird  $w_i^{5k+t} = 4w_i^0$  gesetzt, und falls  $\frac{w_i^{5k+t}}{w_i^0} < \frac{1}{4}$  ist, wird  $w_i^{5k+t} = \frac{w_i^0}{4}$  gesetzt.

Die Wahl dieser Schranken hat sich in der Vergangenheit im Rahmen des Mikrozensus bewährt. Eine Einschränkung der Veränderung der Basisgewichte führt im Allgemeinen zu einer Verringerung der Streuung der Gewichte und hat somit einen positiven Effekt auf die Stichprobenfehler.

### Schritt 4

Durch die in den Schritten 1-3 vorgenommenen Anpassungen kommt es dazu, dass das ursprünglich für alle Personen einer Wohnung einheitliche Gewicht nun von Person zu Person differiert, sofern nicht zufällig alle Personen einer Wohnung dem gleichen Geschlecht, der gleichen Altersklasse, der gleichen Staatsbürgerschaft und dem gleichen Verwaltungsdatenerwerbsstatus angehören. Solche Gewichte haben den Nachteil, dass damit hochgerechnete Tabellen mit der Darstellungseinheit Wohnung mit entsprechenden Personentabellen inkonsistent werden. Daher besteht das Bestreben, alle Personen einer Wohnung mit dem gleichen Gewicht zu versehen. Dies wird erreicht, indem im 4. Schritt alle Personen einer Wohnung als neues Gewicht den Durchschnitt ihrer Gewichte nach Abschluss des 3. Schrittes erhalten.

$$w_i^{5k+4} = \frac{\sum_l w_l^{5k+3}}{N_j}$$

für alle  $i$ .

Die Summierung erstreckt sich über alle Personen  $l$  der Wohnung  $j$ , in der sich die Person  $i$  befindet.  $N_j$  bezeichnet die Anzahl dieser Personen. Dadurch geht möglicherweise die Anpassung an die Bevölkerungsstruktur wieder verloren.

### Schritt 5

Die im Schritt 4 erzielten Gewichte werden multiplikativ so abgeändert, dass die mit den entstehenden Gewichten hochgerechnete Verteilung der Wohnungen nach Bundesland und Zahl der Bewohner mit der Vorgabe  $M_{bh}$  annähernd übereinstimmt. Eine exakte Übereinstimmung wird bezüglich der Wohnungszahlen nicht angestrebt, weil die Vorgaben selbst mit einer gewissen Unschärfe behaftet sind. Die Unschärfe wird für die Zahl der Wohnungen je Bundesland und Haushaltsgröße  $h$  mit 5‰ ( $p_h=0,005$  für  $h=1, \dots, 4$ ) angenommen. Die Zahl der Wohnungen mit 5 und mehr Bewohnern wird je Bundesland mit einer Unschärfe von 20% ( $p_5=0,2$ ) angenommen. Die Veränderung der ursprünglichen Gewichte  $w_i^0$  soll wieder das Vierfache nicht überschreiten bzw. ein Viertel nicht unterschreiten. Es gilt

$$w_i^{5k+5} = \begin{cases} w_i^{5k+4} \frac{M_{b_i h_i}}{\sum_l w_l^{5k+4}} & \text{falls } \sum_l w_l^{5k+4} \notin \left( (1-p_h)M_{b_i h_i}, (1+p_h)M_{b_i h_i} \right) \\ w_i^{5k+4} & \text{sonst} \end{cases}$$

Die Summierung erstreckt sich dabei über alle Wohnungen  $l$  mit denselben Ausprägungen hinsichtlich der Merkmale Bundesland und Haushaltsgröße wie die Beobachtung  $i$ . Falls der Quotient  $\frac{w_i^{5k+5}}{w_i^0} > 4$  ist, wird  $w_i^{5k+5} = 4w_i^0$  gesetzt, und falls  $\frac{w_i^{5k+5}}{w_i^0} < \frac{1}{4}$  ist, wird  $w_i^{5k+5} = \frac{w_i^0}{4}$  gesetzt.

### Abfrage

Es erfolgt nun eine Abfrage, ob die mit den neuen Gewichten errechnete Bevölkerungsstruktur noch in irgendeiner Zelle von der vorgegebenen Struktur um mehr als einen festgelegten Wert (0,01%) abweicht:

$$\max_e \left| \frac{\sum_l w_l^{5k+5} - N_e}{N_e} \right| > 0,0001$$

für  $e \in \{bga, bs, bgv\}$ .

Die Summierung erstreckt sich dabei für  $e = bga$  über alle Personen  $l$  mit denselben Ausprägungen der Merkmale  $b$ ,  $g$  und  $a$ , für  $e = bs$  über alle Personen mit denselben Ausprägungen der Merkmale  $b$  und  $s$  und für  $e = bgv$  über jene Personen mit denselben Ausprägungen der Merkmale  $b$ ,  $g$  und  $v$ . Falls die Bedingung erfüllt ist, wird  $k$  um 1 erhöht und das Verfahren beginnt wieder mit Schritt 1. Dies wird bei den ersten Durchläufen der Iteration noch der Fall sein, weil ja in den Schritten 2-4 etwa die Anpassung an die Altersverteilung wieder zerstört wird. Das Verfahren beginnt also wieder mit Schritt 1, wobei anstatt des Gewichtes aus Formel (1) - (2) nun das nach

Durchlaufen des 5. Schrittes erzeugte Gewicht verwendet wird. Ist die Bedingung nicht erfüllt, so ist das definitive Hochrechnungsgewicht mit  $w_i^{5k+5}$  gefunden und damit das Ziel der Iteration erreicht.

Die in die Iteration eingebaute Sicherheitsschranke bewirkt, dass die Gewichte nach Formel (1) - (2) durch die Iteration um nicht mehr als das Vierfache erhöht bzw. verringert werden und somit die Varianz beschränkt wird. Nach Abschluss der Iteration wird jeder Beobachtung ihr definitives Hochrechnungsgewicht zugewiesen, wie es sich für die Privathaushalte aufgrund der Iteration ergibt.

### Verwendung der Hochrechnungsgewichte bei Quartals- und Jahresauswertungen

Die quartalsweise Berechnung des Umfangs von bestimmten Haushalts- oder Personenmassen in der Grundgesamtheit geschieht durch Addition der endgültigen Gewichte aller Elemente der entsprechenden Stichprobenmasse. Die Totalwerte von quantitativen Merkmalen ergeben sich durch Multiplikation mit den Hochrechnungsgewichten und anschließende Addition, Quotienten durch Division von Totalwerten bzw. Summen. Die vierteljährlich stattfindenden Befragungen im Rahmen des Mikrozensus werden für manche Themenkreise zur Erstellung von Jahresreihen benützt. Die Jahresfiles ergeben sich als Vereinigungsmenge der vier Quartalsfiles, wobei die Hochrechnungsgewichte der Quartalsfiles geviertelt werden. Die Berechnung von Jahreswerten geschieht analog zur Berechnung von Quartalswerten, allerdings mit der vierfachen Datenmasse, in der jeder Datensatz ein auf ein Viertel reduziertes Gewicht aufweist.

### Monatliche Hochrechnung

Neben den Hauptergebnissen für die Quartale und Jahre werden auch die monatlichen Arbeitslosenzahlen nach internationaler Definition aus dem MZ berechnet und an Eurostat übermittelt. Die Vorgehensweise ist im Prinzip gleich wie bei der Erstellung der Quartalshochrechnung. Kleine Unterschiede ergeben sich aufgrund von datentechnischen Einschränkungen wie z.B. der Stichprobengröße in den Monaten. Für die Eckzahlen wird, wie bei den Quartalen, ein Mittelwert aus dem Monatsendstand des Vormonats und des Referenzmonats gebildet. Es wird jedoch bei der Hochrechnung nicht an die Bundesland-Aggregate  $N_{bs}$  des Merkmals Staatsbürgerschaft gebunden sondern an die Österreich-Aggregate  $N_s$ :

$N_s$  ... Anzahl der Personen in Privathaushalten mit Staatsbürgerschaft  $s$  ( $=1, \dots, 6$ ) mit den Nationalitätengruppen "Österreich", "EU-15 (ohne Ö)", "EU ab 2004 (d.h. im Zeitraum 2004 bis 2014 der EU beigetreten)", "Europäische Nicht-EU-Staaten (ohne Türkei)", "Türkei", "Rest".

Abgesehen von der zusätzlichen Bindung an den Erwerbsstatus aus Verwaltungsdaten (mehr dazu in Kapitel „[Erwerbsstatus aus Verwaltungsdaten](#)“) und der Änderung der Gruppeneinteilung der Nationalitäten (siehe Abschnitt „[Nationalität](#)“) wird die Produktion und Übermittlung der Monatsergebnisse wie bisher durchgeführt, siehe dazu Gumprecht et al. 2011. Die Ergebnisse zeigen wie bei den Quartalswerten, im Vergleich zu den bisherigen Hochrechnungsergebnissen, eine Niveauverschiebung. Die Trends bleiben größtenteils unverändert. Sollten die Verwaltungsdaten und/oder der Verknüpfungsschlüssel nicht zeitgerecht zur Verfügung stehen, wird mittels Zeitreihenmodell und Imputation ein vorläufiges Gewicht erstellt. Mit Abschluss eines

Quartals werden die Monatsgewichte im Rahmen der Quartalshochrechnung in jedem Fall revidiert.

## Genauigkeit und Fehlerrechnung

Die Fehlerrechnung wird mit Hilfe von Replikationsgewichten aus einem Bootstrapverfahren durchgeführt. Für jede Person bzw. jeden Haushalt in der Stichprobe gibt es zusätzlich zu den Stichprobengewichten 500 Bootstrapgewichte.

Die Berechnung der Bootstrapgewichte erfolgt durch die Ziehung von Bootstrapstichproben. Dabei wird eine Stichprobe der Größe  $m$  durch Ziehung mit Zurücklegen aus der ursprünglichen Haushaltsstichprobe derselben Größe ermittelt, sodass jedes Stichprobenelement 0- bis  $m$ -mal vorkommen kann. Die Häufigkeit des Vorkommens  $r_i^j$  der Beobachtung  $i$  in der  $j$ -ten Bootstrapstichprobe multipliziert mit den kalibrierten Stichprobengewichten  $w_i$  der ursprünglichen Stichprobe ergibt die unkalibrierten Bootstrapgewichte  $bw_i^{0,j} = r_i^j w_i$ . Als nächster Schritt in der Berechnung werden die kalibrierten Bootstrapgewichte  $bw_i^j$  berechnet, indem das oben beschriebene Iterationsverfahren zur Kalibrierung mit Startwert  $bw_i^{0,j}$  durchlaufen wird.

Da es sich bei der Stichprobenziehung des MZ um eine Ziehung ohne Zurücklegen aus einer endlichen Grundgesamtheit handelt, kann nicht einfach das oben beschriebene "naive" Bootstrap-Verfahren (Efron & Tibshirani 1993) verwendet werden. Für diese Situation eignet sich das „Rescaled Bootstrap“ Verfahren (siehe Preston 2009 und Rao & Wu 1993), wobei der Hauptunterschied zum klassischen Bootstrap darin besteht, dass die Häufigkeit des Vorkommens  $r_i^j$  in einer Bootstrapstichprobe nicht mehr ganzzahlig ist.

Zur tatsächlichen Fehlerrechnung wird mit den 500 Bootstrapgewichten  $bw_i^j$  jeweils der Schätzer  $\hat{\theta}$  der zu schätzenden Populationsgröße  $\theta$  berechnet. Die Standardabweichung dieser 500 Schätzungen  $\hat{\theta}^j$

$$SE = \sqrt{\frac{\sum_j (\hat{\theta}^j - \bar{\hat{\theta}})^2}{500 - 1}}$$

mit  $\bar{\hat{\theta}}$ , dem arithmetischen Mittel der Bootstrap Replicates  $\hat{\theta}^j$ , entspricht dem geschätzten Stichprobenfehler. Die untere bzw. obere Grenze des Konfidenzintervalls ist definiert durch das 2.5% bzw. 97.5% Quantil der erhaltenen 500  $\hat{\theta}^j$  Werte.

## Clustering auf Wohnungsebene

Auch die Bootstrapstichproben sind auf Wohnungsebene geclustert, d.h. die Ziehung der Bootstrapstichproben wird auf Wohnungsebene durchgeführt. Dadurch erhalten alle Personen in einer Wohnung dasselbe Bootstrapgewicht.

## Rotation

In jedem Quartal wird nur rund ein Fünftel der Quartalstichprobe durch eine neue Zufallsstichprobe ersetzt, für diesen Teil der Stichprobe werden neue Bootstrapstichproben gezogen und damit  $r_i^j$  berechnet. Bei den restlichen 4/5 werden

diese  $r_i^j$  von dem vorhergehenden Quartal beibehalten, sodass auch die Überschneidung der Stichproben bei Schätzung von Veränderungen im Bootstrapverfahren abgebildet ist.

### Approximative Fehlerrechnung

Der relative Stichprobenfehler für Häufigkeitstabellen lässt sich sowohl für Wohnungsmassen als auch für Personenmassen unter der Annahme einer einfachen Zufallsauswahl von Wohnungen bzw. Personen mit Hilfe des Binomialansatzes abschätzen. Streng genommen sind die absoluten Häufigkeiten bei einem Ziehungsvorgang ohne Zurücklegen hypergeometrisch verteilt, da aber der Auswahlsatz sehr klein ist, ist die Verwendung der Binomialverteilung gerechtfertigt. Bei diesem Näherungsverfahren werden also die Effekte der komplexen Hochrechnung (iterative Anpassung an vorgegebene Verteilungen der Personen und Wohnungen) und der geklumpten Auswahl von Personen ausgeklammert. Berücksichtigt werden hingegen die pro Bundesland unterschiedlichen Auswahlsätze. Nach diesem Ansatz ergibt sich der relative Stichprobenfehler in Prozent bei 95-prozentiger statistischer Sicherheit für eine hochgerechnete Fallzahl  $x$  (Wohnungen oder Personen) nach der Formel

$$2 * 100 \sqrt{\frac{(N_b - n_b)(N_b - x)}{N_b n_b x}}$$

für Bundeslandergebnisse bzw.

$$2 * 100 \sqrt{\sum_b \frac{(N_b - n_b)(N - x)N_b}{n_b N^2 x}}$$

für österreichbezogene Häufigkeiten  $x$ . Dabei bezeichnen  $N_b$  die Grundgesamtheit im Bundesland  $b$ ,  $n_b$  den Stichprobenumfang im Bundesland  $b$  und  $N$  die Grundgesamtheit in Österreich. Der Faktor 2 wird als Approximation des entsprechenden Quantils der Normalverteilung verwendet.

Im Falle der Wohnungen sind in den obigen Formeln die Größen  $N$ ,  $N_b$  und  $n_b$  durch  $M$ ,  $M_b$  und  $m_b$  zu ersetzen.

Bei der Herleitung der zweiten Formel wird vorausgesetzt, dass die Beiträge der einzelnen Bundesländer zum Wert von  $x$  proportional zur Größe der einzelnen Bundesländer sind.

Beiden Formeln liegt die Varianzschätzung gemäß dem Binomialansatz zugrunde. Dabei wird die Varianz von  $x = Np$  mit dem entsprechenden Anteil  $p$  folgendermaßen geschätzt:

$$v(x) = \frac{N(N - n)}{n} p(1 - p)$$

Der relative Stichprobenfehler berechnet sich daraus wie folgt:

$$\frac{1}{Np} \sqrt{v(x)}$$

Die mittels obiger Formel berechneten relativen Stichprobenfehler bei 95% Sicherheit sind in Tabelle 1 für Personen und in Tabelle 2 für Wohnungen dargestellt. Für eine

Personenzahl von 50.000 in der hochgerechneten Stichprobe liegt der tatsächliche Wert in Österreich mit 95%iger Sicherheit zwischen 43.200 (50.000-13.6%) und 56.800 (50.000+13.6%). Es wird nochmals darauf hingewiesen, dass die Tabellen nur einen Hinweis auf die Größenordnung des Fehlers und nicht den exakten Wert angeben.

Personen in 1.000	Burgenland	Kärnten	Niederösterreich	Oberösterreich	Salzburg	Steiermark	Tirol	Vorarlberg	Wien	Österreich
1	56.2	67.8	107.8	101.6	67.2	97.2	75.9	54.2	113.3	96.3
2	39.6	47.9	76.2	71.8	47.4	68.7	53.7	38.3	80.1	68.1
3	32.3	39.1	62.2	58.6	38.7	56.1	43.8	31.2	65.4	55.6
4	27.9	33.8	53.9	50.7	33.5	48.6	37.9	27	56.6	48.2
5	24.9	30.2	48.2	45.4	29.9	43.4	33.9	24.1	50.6	43.1
6	22.7	27.6	44	41.4	27.3	39.6	30.9	22	46.2	39.3
7	21	25.5	40.7	38.3	25.2	36.7	28.6	20.3	42.8	36.4
8	19.6	23.8	38	35.8	23.6	34.3	26.7	19	40	34
9	18.5	22.4	35.9	33.8	22.2	32.3	25.2	17.9	37.7	32.1
10	17.5	21.3	34	32	21.1	30.6	23.9	16.9	35.7	30.5
20	12.1	14.9	24	22.6	14.7	21.6	16.8	11.8	25.2	21.5
30	9.7	12.1	19.5	18.3	11.9	17.5	13.6	9.5	20.5	17.6
40	8.2	10.3	16.8	15.8	10.2	15.1	11.7	8.1	17.7	15.2
50	7.2	9.2	15	14.1	9	13.5	10.4	7.1	15.8	13.6
60	6.5	8.3	13.7	12.8	8.2	12.2	9.4	6.4	14.4	12.4
70	5.8	7.6	12.6	11.8	7.5	11.3	8.6	5.8	13.3	11.5
80	5.3	7	11.8	11	6.9	10.5	8	5.4	12.4	10.7
90	4.9	6.5	11	10.4	6.5	9.9	7.5	5	11.6	10.1
100	4.5	6.1	10.4	9.8	6.1	9.3	7	4.6	11	9.6
200	2.2	3.8	7.1	6.7	3.7	6.3	4.6	2.6	7.5	6.7
300		2.6	5.6	5.2	2.5	4.9	3.3	1.4	6	5.5
400		1.8	4.7	4.3	1.7	4	2.5		5	4.7
500		0.9	4	3.6	0.7	3.3	1.9		4.3	4.2
600			3.5	3.1		2.8	1.2		3.8	3.8
700			3.1	2.7		2.4			3.3	3.5
800			2.7	2.4		2			3	3.2
900			2.4	2		1.6			2.6	3
1000			2.1	1.7		1.3			2.3	2.9
2000										1.9
3000										1.4
4000										1.1
5000										0.9
6000										0.7
7000										0.5
8000										0.2

Tabelle 1: Relativer Stichprobenfehler hochgerechneter Personenzahlen (in Prozent) bei 95% statistischer Sicherheit.

Wohnungen in 1.000	Burgen- land	Kärnten	Nieder- österreich	Ober- österreich	Salzburg	Steier- mark	Tirol	Vorarl- berg	Wien	Österreich
1	53.7	66.2	105.9	100.9	66	94.6	73.4	53.3	114.1	96
2	37.8	46.7	74.9	71.3	46.6	66.8	51.8	37.5	80.6	67.9
3	30.7	38.1	61.1	58.2	38	54.5	42.2	30.5	65.8	55.4
4	26.5	32.9	52.9	50.3	32.8	47.2	36.5	26.4	57	48
5	23.6	29.4	47.2	45	29.3	42.1	32.6	23.5	50.9	42.9
6	21.4	26.7	43.1	41	26.7	38.4	29.7	21.4	46.5	39.2
7	19.8	24.7	39.9	37.9	24.6	35.5	27.5	19.7	43	36.3
8	18.4	23.1	37.3	35.5	23	33.2	25.6	18.4	40.2	33.9
9	17.3	21.7	35.1	33.4	21.6	31.3	24.1	17.3	37.9	32
10	16.3	20.5	33.3	31.7	20.5	29.6	22.9	16.3	35.9	30.3
20	11	14.2	23.4	22.2	14.1	20.8	15.9	11.2	25.2	21.4
30	8.5	11.3	18.9	18	11.3	16.8	12.7	8.8	20.5	17.5
40	6.9	9.6	16.3	15.4	9.5	14.4	10.8	7.3	17.6	15.1
50	5.7	8.4	14.4	13.7	8.3	12.7	9.5	6.2	15.7	13.5
60	4.8	7.4	13.1	12.4	7.3	11.5	8.5	5.4	14.2	12.3
70	4.1	6.7	12	11.3	6.6	10.5	7.7	4.8	13.1	11.4
80	3.4	6.1	11.1	10.5	6	9.7	7	4.2	12.2	10.6
90	2.7	5.6	10.4	9.8	5.4	9.1	6.5	3.7	11.4	10
100	2	5.1	9.8	9.2	5	8.5	6	3.2	10.7	9.5
200		2	6.3	5.8	1.7	5.2	3		7.1	6.6
300			4.6	4.1		3.5			5.4	5.3
400			3.4	2.9		2.2			4.2	4.5
500			2.5	1.9					3.4	4
600			1.5						2.6	3.6
700									1.9	3.3
800										3
900										2.8
1000										2.6
2000										1.5
3000										0.8

Tabelle 2: Relativer Stichprobenfehler hochgerechneter Wohnungszahlen (in Prozent) bei 95% statistischer Sicherheit.

## Literatur

- Ediev, D. M. (2007): "On Projecting the Distribution of Private Households by Size". Vienna Institute of Demography of Austrian Academy of Sciences. Working Paper 04/2007.*
- Efron, B. / Tibshirani, R. (1993): "An Introduction to the Bootstrap". Chapman & Hall.*
- Gumprecht, D. / Haslinger, A. / Kowarik A. (2011): „Austrian LFS Monthly Unemployment Rates“, Austrian Journal of Statistics 40/4/2011, 297-313, Wien.*
- Gumprecht, D. / Oismüller, A. (2013): „Non-Response im Mikrozensus“, Statistische Nachrichten 11/2013, 1046-1061, Wien.*
- Haslinger, A. / Kytir, J. (2006): „Stichprobendesign, Stichprobenziehung und Hochrechnung des Mikrozensus ab 2004“, Statistische Nachrichten 6/2006, 510-519, Wien.*
- Kytir, J. / Stadler, B. (2004): „Die kontinuierliche Arbeitskräfteerhebung im Rahmen des neuen Mikrozensus“, Statistische Nachrichten 6/2004, 511-520, Wien.*
- Preston J. (2009): Rescaled bootstrap for stratified multistage sampling. Survey Methodology 35(2) 227-234*
- Rao JNK, Wu CFJ. (1993): "Bootstrap inference for sample surveys". Proc Section on Survey Research Methodology (866-871)*
- STATISTIK AUSTRIA (2010): „Methodenhandbuch zur Abgestimmten Erwerbsstatistik“, [http://www.statistik.at/web\\_de/static/methodenhandbuch\\_zur\\_abgestimmten\\_erwerb\\_sstatistik\\_053160.pdf](http://www.statistik.at/web_de/static/methodenhandbuch_zur_abgestimmten_erwerb_sstatistik_053160.pdf).*