

Stichprobendesign, Stichprobenziehung und Hochrechnung des Mikrozensus ab 2004

ALOIS HASLINGER
JOSEF KYTIR

Mit Anfang des Jahres 2004 wurde der Mikrozensus grundlegend umgestaltet. Der vorliegende Artikel gibt einen Überblick über methodische und praktische Aspekte des Stichprobendesigns, der Stichprobenziehung und Hochrechnung des „neuen“ Mikrozensus. Dem Mikrozensus liegt eine einfach geschichtete (Bundesländer) Zufallsstichprobe zugrunde. Als Auswahlrahmen für die Ziehung der Stichprobeneinheiten (private Haushalte) dient ein quartalsweise aktualisierter Datenabzug aller hauptwohnsitzgemeldeten Personen aus dem Zentralen Melderegister. Die Ausschöpfung der Stichprobe lag 2005 bei rund 90%. Die Hochrechnung wird an die Bevölkerungsstände des Bevölkerungsregisters der STATISTIK AUSTRIA und an die Haushaltsprognose gebunden und erfolgt technisch mittels des Verfahrens der iterativen proportionalen Anpassung. Zur Abschätzung des relativen Stichprobenfehlers von absoluten Personen- und Wohnungshäufigkeiten werden Näherungsverfahren vorgestellt. Für einige zentrale Schätzwerte des Mikrozensus (z.B. Zahl der Arbeitslosen) wurde eine genauere Fehlerrechnung durchgeführt, die zeigt, dass das Design des Mikrozensus die von der EU geforderten Genauigkeitsvorgaben erfüllt.

Aus formalen (rechtlichen) Gründen, aber auch aus Qualitätsüberlegungen heraus, war es notwendig geworden, Österreichs größte Stichprobenerhebung, den Mikrozensus, mit Jahresbeginn 2004 vollständig neu zu gestalten.¹⁾ Die Umstellungen betrafen auch das Stichprobendesign und die Stichprobenziehung sowie die Hochrechnung der erhobenen Daten. Der vorliegende Beitrag dokumentiert diese Veränderungen.

Generell dient der Mikrozensus ab dem 1. Quartal 2004 im Wesentlichen zwei Zwecken: zum einen zur Durchführung der Europäischen Arbeitskräfteerhebung in kontinuierlicher Form, also mit einer auf alle Wochen eines Jahres gleichmäßig verteilten Stichprobe, und zum anderen als Grundlage für die Abschätzung der Entwicklung der Wohnungsausgaben im Rahmen der monatlichen Verbraucherpreisindizes.

Rechtliche Rahmenbedingungen

Nationale Rechtsgrundlage für den „neuen“ Mikrozensus ist die Erwerbs- und Wohnungsstatistik-Verordnung (EWStV; www.statistik.at/mikrozensus/verordnung.pdf) des Bundesministeriums für Wirtschaft und Arbeit vom 28. November 2003. Hier finden sich auch die entsprechenden Verweise

¹⁾ Siehe dazu die Beiträge „Die kontinuierliche Arbeitskräfteerhebung im Rahmen des neuen Mikrozensus“, Heft 6/2004, S. 511 ff. und „Das Datenmanagement im neuen Mikrozensus. Eine Prozessbeschreibung“, Heft 3/2006, S. 156 ff.

auf die relevanten Rechtsgrundlagen auf europäischer Ebene, insbesondere die Verordnung (EG) Nr. 577/98 zur Durchführung einer Stichprobenerhebung über Arbeitskräfte in der Gemeinschaft sowie die Verordnung (EG) Nr. 2494/95 über harmonisierte Verbraucherpreisindizes.

Der Begriff des Mikrozensus findet sich in der EWStV im §6 genannt. Die Bundesanstalt Statistik Österreich wird hier unter dem Titel „Auswahl der Mikrozensus-Stichprobe“ verpflichtet, eine Stichprobenerhebung entsprechend der Verordnung (EG) Nr. 577/98 über Arbeitskräfte in der Gemeinschaft durchzuführen. Im Art. 2 dieser Verordnung heißt es wiederum: „Die Erhebung wird in jedem Mitgliedstaat bei einer Stichprobe von Haushalten oder Einzelpersonen, die zum Zeitpunkt der Erhebung ihren Wohnsitz im Wirtschaftsgebiet des jeweiligen Staates haben, durchgeführt.“, sowie in weiterer Folge: „Die Grundgesamtheit der Erhebung besteht in erster Linie aus den Personen in Privathaushalten im Wirtschaftsgebiet jedes Mitgliedstaates.“. Art. 3 regelt dann die Repräsentativität der Stichprobe: „Für eine Gruppe von Arbeitslosen, die 5% der Bevölkerung im erwerbsfähigen Alter ausmacht, darf der relative Standardfehler der Schätzungen von Jahresdurchschnittswerten ... auf der Ebene NUTS II (Bundesländer) höchstens 8% der betreffenden Bevölkerungsgruppe betragen.“, wobei Regionen mit weniger als 300.000 Einwohnern von dieser Bestimmung ausgenommen werden. Als zweites Kriterium der Repräsentativität wird festgelegt, dass für Merkmale, die 5% der Be-

völkerung im erwerbsfähigen Alter betreffen, „der ... Standardfehler für die Schätzung von Veränderungen dieser Merkmale zwischen zwei aufeinander folgenden Quartalen auf nationaler Ebene höchstens 3% der betreffenden Bevölkerungsgruppe betragen darf.“

Erhebungsmethoden und Stichprobendesign

Die organisatorischen Änderungen des Mikrozensus an der Jahreswende 2003/2004 betrafen insbesondere auch die Erhebungsmethode. Im „alten“ Mikrozensus vor 2004 gab es eine Stichprobe mit Achtel-Rotation, d.h., jeder Haushalt verblieb insgesamt acht Quartale lang in der Stichprobe, und jede der acht Befragungen sollte vom Interviewerstab Face-to-Face durchgeführt werden. Für den „neuen“ Mikrozensus wurde u.a. aus Gründen der Respondentenentlastung dem Modell einer Fünftel-Rotation der Vorzug gegeben, und es wurde eine Erhebungsinfrastruktur für einen gemischten Erhebungsmodus aufgebaut. Befragungen in Form von Face-to-Face-Interviews sind nun im Prinzip nur für sog. „Erstbefragungen“ vorgesehen, also bei jenem Fünftel der Haushalte, die jedes Quartal neu in die Stichprobe gezogen werden. Die vier „Folgebefragungen“ erfolgen telefonisch, in Form des sog. „Computer Assisted Telephone Interviewing“ (CATI).²⁾ Zu diesem Zweck wurde in der STATISTIK AUSTRIA mit Jahresbeginn 2004 ein Telefonstudio eingerichtet.

Dieser gemischte Erhebungsmodus ermöglichte gegenüber der Situation vor 2004 eine starke Verkleinerung des Interviewerstabs im Face-to-Face-Bereich von rund 800 auf 145 Personen, die im Gegensatz zur dezentralen Betreuung durch die Landesstatistiken nun zentral von der STATISTIK AUSTRIA geführt werden. In weiterer Folge fiel damit

²⁾ Tatsächlich werden rund 30% der (erfolgreichen) Interviews eines Quartals Face-to-Face und 70% telefonisch durchgeführt, da eine Reihe von Haushalten über kein Telefon verfügt oder keine Telefonnummer angeben möchte. Zudem gibt es die Möglichkeit des Methodenwechsels während der Erhebungsphase (von Face-to-Face-Interviews zu telefonischer Befragung oder umgekehrt), was optimale Voraussetzungen hinsichtlich hoher Ausschöpfung bei guter Qualität bietet.

auch die bis dahin gegebene Klumpung auf rund 1.100 Gemeinden, also die Zweistufigkeit der Stichprobe, weg. Um den voranstehend zitierten Vorgaben der EU-Verordnung für die Repräsentativität der Ergebnisse gerecht zu werden, ist allerdings weiterhin eine Schichtung der Stichprobe nach NUTS-II-Regionen (Bundesländern) notwendig. Die Mikrozensus-Stichprobe ist damit eine geschichtete uneingeschränkte Zufallsauswahl von Wohnungen bzw. eine geschichtete Klumpenauswahl von Personen. Einen Überblick über die Brutto-Stichproben sowie die Auswahlsätze nach Bundesländern bietet *Tabelle 1*.

Die Brutto-Stichprobengröße beträgt österreichweit 22.740 Haushalte pro Quartal (4.548 Erst- und 18.192 Folgebefragungshaushalte). Bezogen auf die Zahl der Privathaushalte zum Zeitpunkt der Volkszählung 2001 entspricht das einem Auswahlatz von 0,68%. Diese Österreich-Werte ergeben sich dabei als Summe der Stichproben für die Bundesländer (Schichten), wobei in den Bundesländern Kärnten, Oberösterreich, Salzburg, Tirol und Vorarlberg die Bruttostichprobe bei jeweils 2.450 Haushalten liegt. Für das Burgenland wurde aufgrund der unter dem Wert von 300.000 liegenden Einwohnerzahl eine etwas kleinere Stichprobe (1.750 Haushalte), für Wien wegen der erfahrungsgemäß geringeren Ausschöpfung und der großen Zahl von Einpersonen-Haushalten eine etwas größere Bruttostichprobe (3.530 Haushalte) festgelegt. In Niederösterreich und in der Steiermark wurde der Auswahlatz in einzelnen sehr dünn besiedelten Bezirken verdoppelt, um eine für die Aufrechterhaltung eines Interviewerstabs notwendige Zahl an Stichprobenhaushalten zu erzielen. Die Stichprobengrößen betragen damit für Niederösterreich 2.645 und für die Steiermark 2.560 Haushalte. Die Auswahlätze bewegen sich damit zwischen 0,42% in Niederösterreich und 1,82% in Vorarlberg.

Das ZMR als Auswahlrahmen

Das seit dem Jahr 2002 aktive Zentrale Melderegister (ZMR) schafft nicht nur für die Bevölkerungsstatistik neue Voraussetzungen und Möglichkeiten,³⁾ es kann gleichzeitig auch als Auswahlrahmen für Haushalts- und Personenstichproben genutzt werden. Die grundsätzliche Möglichkeit der Verwendung der Daten des ZMR durch die amtliche Statistik ist dabei im § 16b Abs. 8 Meldegesetz rechtlich verankert. („Die im Zentralen Melderegister gespeicherten Daten dürfen für statistische Zwecke nach dem Bundesstatistikgesetz 2000 an Organe der Bundesstatistik ... übermittelt werden.“)

Die Vorteile der Nutzung des ZMR als Auswahlrahmen für Stichprobenerhebungen liegen dabei vor allem in der Vollständigkeit (alle mit Hauptwohnsitz in Österreich gemeldeten Personen) und Aktualität der Daten. Dem stehen bei Haushaltsstichproben allerdings als Nachteil die Qualitätsmängel des ZMR im Bereich der korrekten Zuordnung der

Brutto-Quartalsstichprobe des Mikrozensus 2005

Tabelle 1



Bundesland	Privathaushalte			
	Erstbefragung	Folgebefragung	insgesamt	
	absolut		absolut	Auswahlatz ¹⁾ in %
Burgenland	351	1.404	1.755	1,65
Kärnten	490	1.960	2.450	1,09
Niederösterreich	529	2.116	2.645	0,42
Oberösterreich	490	1.960	2.450	0,45
Salzburg	490	1.960	2.450	1,18
Steiermark	512	2.048	2.560	0,55
Tirol	490	1.960	2.450	0,94
Vorarlberg	490	1.960	2.450	1,82
Wien	706	2.824	3.530	0,46
Österreich	4.548	18.192	22.740	0,68

¹⁾ Auswahlatz bezogen auf die Zahl der Privathaushalte bei der Volkszählung 2001.

³⁾ Siehe Heft 3/2005, S. 203 ff.

gemeldeten Personen zu den einzelnen Wohnungen (= Haushalten) eines Gebäudes gegenüber. Die rund 2.400 Meldebehörden (= Gemeinden) sind von diesem Problem in quantitativ sehr unterschiedlicher Weise betroffen; erhebliche Qualitätsmängel weisen in dieser Hinsicht etwa die Meldedaten der Städte Klagenfurt und Salzburg auf. (Zu den Konsequenzen der Unschärfe der Wohnungsabgrenzung für die Hochrechnung siehe die nachstehenden Ausführungen.)

Die EWStV präzisiert die Verwendung des ZMR als Auswahlrahmen für den Mikrozensus, indem auf die im Meldegesetz § 16b Abs. 1 für die Zwecke der Wanderungsstatistik ohnehin laufend vorgesehenen Datenlieferungen an die STATISTIK AUSTRIA verwiesen wird. Diese Datenlieferungen erfolgen quartalsweise und umfassen u.a. den indirekt personenbezogenen Meldedatenbestand der hauptwohnsitzgemeldeten Personen zum Stichzeitpunkt Quartalsbeginn, 0 Uhr. Vor der Stichprobenziehung werden Adressen ausgeschieden, die entweder als Anstaltshaushalte identifiziert wurden oder die bereits in früheren Quartalen in die Stichprobe gewählt wurden. Auswahlrahmen der Mikrozensus-Stichprobe bilden damit alle übrigen Wohnungen (= Privathaushalte) mit zumindest einer hauptwohnsitzgemeldeten Person, bezogen auf den Beginn jenes Quartals, das unmittelbar vor dem jeweiligen Referenzquartal liegt.

Auf Ebene der Personen orientiert sich der Mikrozensus allerdings nicht an den an einer Adresse im ZMR hauptwohnsitzgemeldeten Personen („Melderealität“). Den Interviewern wird diese Information bewusst nicht kommuniziert. Sie sollen vielmehr ohne Kenntnis der „Melderealität“ in den Haushalten direkt erfragen und dokumentieren, welche Personen im Haushalt leben („Befragungsrealität“).

Stichprobenziehung

Die Stichprobenziehung selbst erfolgt pro Politischen Bezirk durch das Aufbringen einer gleichverteilten Zufallszahl auf alle Wohnungen der Grundgesamtheit. Anschließend werden die Wohnungen bezirksweise nach dieser Zufallszahl sortiert. Dem vorgegebenen Auswahlatz des Bundeslands entsprechend, in dem der jeweilige Bezirk liegt, ergeben dann die n ersten Wohnungen eines Bezirks jene Adressen, die neu in die Stichprobe rotieren.

Den in die Stichprobe gewählten Haushalten muss anschließend ein Interviewersprengel und eine Referenzwoche (jene Woche, auf die sich die erfragten Angaben in der Regel beziehen) zugewiesen werden. Auf Basis der rund 22.000 (bewohnten) Ortschaften Österreichs wurden 116 Interviewersprengel gebildet, deren Abgrenzungen im Wesentlichen unverändert bleiben. Die Zuweisung der Referenzwochen erfolgt jedes Quartal automatisiert unter Verwendung einer eigens dafür entwickelten Applikation im Rahmen eines Geographischen Informationssystems (GIS). Der verwendete Algorithmus sorgt dafür, dass - unter Beachtung der Prämisse einer gleichmäßigen Verteilung der Stichprobenhaushalte über alle Wochen eines Quartals - die Zuteilung der Referenzwochen in räumlicher Hinsicht für die Interviewer optimiert (räumliche Nähe von Haushalten mit gleicher Referenzwoche innerhalb eines Interviewersprengels, insbesondere in jenen mit mehr als einem Interviewer) vorgenommen wird.

Da die Ausgangsdaten für die Stichprobenziehung nur einen indirekten Personenbezug aufweisen, ist parallel dazu ein weiterer Schritt notwendig. Wie im § 7 Abs. 1 EWStV vorgesehen, werden die ZMR-Zahlen der an den gezogenen Adressen gemeldeten Personen elektronisch an das Zentrale Melderegister übermittelt. Das Zentrale Melderegister gibt daraufhin die Namen der diesen ZMR-Zahlen zugeordneten Personen bekannt. Die Namen dienen in weiterer Folge dazu, den ausgewählten Haushalten ca. eine Woche vor dem vorgesehenen Befragungstermin (die der Referenzwoche unmittelbar folgende Kalenderwoche) ein personalisiertes Ankündigungsschreiben einschließlich eines Informationsfolers über den Mikrozensus zuschicken zu können.

Ausschöpfung

Durch die optimalen organisatorischen Rahmenbedingungen und die in der EWStV verankerte Auskunftspflicht erreicht der Mikrozensus Ausschöpfungsquoten, die weit über jenen liegen, die in sozialwissenschaftlichen Stichprobenerhebungen üblicherweise erzielbar sind. Für das Jahr 2005 konnten von den insgesamt 90.949 in der Stichprobe befindlichen Haushalten insgesamt 81.122 komplett, also jeweils alle Haushaltsmitglieder, befragt werden. Die Ausschöpfungsquote (ohne neutrale Ausfälle wie z.B. zum Interviewzeitpunkt nicht (mehr) bewohnte Wohnungen) lag damit bei 96,7% (Tabelle 2).

Brutto- und Nettostichprobe (Haushalte) des Mikrozensus 2005 nach Bundesländern

Tabelle 2



	Österreich	Burgenland	Kärnten	Nieder- österreich	Ober- österreich	Salzburg	Steiermark	Tirol	Vorarlberg	Wien
Brutto/Bruttostichprobe ¹⁾	90.949	7.020	9.940	11.110	9.800	9.820	10.669	9.780	9.800	13.010
Neutrale Ausfälle	6.932	307	546	824	591	879	858	625	567	1.735
Bruttostichprobe ²⁾	84.017	6.713	9.394	10.286	9.209	8.941	9.811	9.155	9.233	11.275
Verweigerungen	102	3	12	11	17	6	20	6	7	20
Keine Information verfügbar	2.693	104	191	324	214	283	338	167	204	868
Nettostichprobe ³⁾	81.222	6.606	9.191	9.951	8.978	8.652	9.453	8.982	9.022	10.387
Ausschöpfungsquote (%)	96,7	98,4	97,8	96,7	97,5	96,8	96,4	98,1	97,7	92,1

¹⁾ Inklusive neutraler Ausfälle. - ²⁾ Ohne neutrale Ausfälle. - ³⁾ Erfolgreiche Befragung aller Haushaltsmitglieder.

Erwartungsgemäß liegt Wien hinsichtlich der Ausschöpfung mit 92% deutlich unter den Flächenbundesländern. Eine Ausschöpfung von mehr als 97% erreichte der Mikrozensus im Jahr 2005 im Burgenland, in Kärnten, in Oberösterreich sowie in Tirol und Vorarlberg.

Hochrechnung und Hochrechnungsvorgaben

Bei einer zufallsgesteuerten Stichprobenauswahl wird ein verkleinertes, aber sonst im Schnitt wirklichkeitsgetreues Abbild des Merkmalskörpers 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 interessierenden, aber unbekanntenen Parameter der Grundgesamtheit herangezogen.

Dem Mikrozensus liegt eine geschichtete Stichprobe mit unterschiedlichen Auswahlsätzen pro Bundesland zugrunde. Hinzu kommt, dass Stichprobenadressen schichtspezifisch verschieden stark ausfallen. Das Abbild der Gesamtheit in der Stichprobe ist also in einem unterschiedlichen Ausmaß verkleinert. 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 auf den Datensätzen gespeichert. Bei der Tabellierung der Mikrozensusergebnisse wird jeder einzelne Merkmalswert vor der Addition mit dem Hochrechnungsfaktor multipliziert.

Tatsächlich wird die Hochrechnung an die Bevölkerungsstände des Bevölkerungsregisters der STATISTIK AUSTRIA gebunden, d.h., die Hochrechnungsgewichte werden so bestimmt, dass die aus dem Mikrozensus resultierenden Verteilungen von Personenmassen nach Bundesland, Alter und Geschlecht sowie nach Bundesland und Nationalität mit den entsprechenden Werten des Bevölkerungsregisters übereinstimmen.

Wirklich neu an der Hochrechnung des Mikrozensus ab 2004 ist nur, dass die Hochrechnungsgewichte so bestimmt werden, dass auch die sich aus dem Mikrozensus ergebenden Verteilungen der Wohnungen (= Haushalte) nach der Zahl der Bewohner mit den entsprechenden Verteilungen der Haushaltsprognose der STATISTIK AUSTRIA übereinstimmen. Diese zusätzliche Restriktion wurde eingeführt, weil einerseits die Wohnungsabgrenzung im Zentralen Melderegister mit einer gewissen Unschärfe behaftet ist und weil andererseits die Antwortausfälle im Mikrozensus umso höher sind, je weniger Personen in einer Wohnung leben.

Folgende Vorgaben der STATISTIK AUSTRIA werden für die Hochrechnung herangezogen:

N_{bag} ... Gesamtzahl der Personen in Privathaushalten (laut Bevölkerungsregister) im Bundesland b ($= 1 \dots 9$), Alter a (in Jahren; $1= 0-2$, $2= 3-5$, $3= 6-9$, $4= 10-14$, ... $18= 80-84$, $19= 85$ und älter) und Geschlecht g ($= 1, 2$) im jeweiligen Quartal.

N_{bs} ... Gesamtzahl der Personen in Privathaushalten (laut Bevölkerungsregister) im Bundesland b ($= 1 \dots 9$) und Staatsbürgerschaft s ($1= \text{Österreich}$, $2= \text{EU-15 ohne Österreich}$, $3= \text{Ex-Jugoslawien}$, $4= \text{Türkei}$, $5= \text{sonstiges Ausland}$) im jeweiligen Quartal.

M_{bh} ... Gesamtzahl an bewohnten Wohnungen (= Privathaushalte laut Haushaltsprognose der STATISTIK AUSTRIA) im Bundesland b ($= 1 \dots 9$) mit h Bewohnern (Haushaltsgröße $h= 1, 2, 3, 4, 5, 6+$) im jeweiligen Quartal.

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 Bundeslands 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 .

Bei der Stichprobenauswahl für das Jahr 2004 wurde in 13 Bezirken der Auswahlsatz im Vergleich zu den übrigen Bezirken des jeweiligen Bundeslands verdoppelt, um auch in diesen relativ dünn besiedelten Bezirken eine genügend hohe Auslastung der Interviewer zu erzielen. Ab dem Jahr 2005 wird diese Erhöhung des Auswahlsatzes nur mehr in fünf Bezirken durchgeführt. In den von dieser Verdoppelung des Auswahlsatzes betroffenen Bundesländern ergibt sich das Hochrechnungsgewicht bei freier Hochrechnung als

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

für Personen und Wohnungen aus Bezirken mit normalem Auswahlsatz und als die Hälfte dieses Quotienten für Bezirke mit doppeltem Auswahlsatz. m_{b1} bezeichnet die Anzahl der Wohnungen der Nettostichprobe in den Bezirken des Bundeslands b mit normalem Auswahlsatz und m_{b2} bezeichnet die Anzahl der Wohnungen der Nettostichprobe in den Bezirken des Bundeslands b mit doppeltem Auswahlsatz.

Anpassung an vorgegebene Bevölkerungsstrukturen: Die mittels der Gewichte (1) und (2) errechneten Bevölkerungszahlen weichen hinsichtlich der Verteilung nach Alter, Geschlecht

und Nationalität in der Regel von den entsprechenden Zahlen des Bevölkerungsregisters ab.

Viele Auswertungen des Mikrozensus haben die Berechnung von Umfängen von speziellen Personengesamtheiten zum Ziel. 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 Bevölkerungsregister bekannten Eckzahlen anpasst. Dies wird erreicht, indem man die Hochrechnungsgewichte aus den Formeln (1) und (2) in einem iterativen Verfahren so abändert, dass die hernach errechneten Mikrozensusergebnisse mit den vorgegebenen Bevölkerungsstrukturen übereinstimmen.

Für die Zwecke der **Iteration** werden für jede Person i die Merkmalsausprägungen für das Bundesland b ($= 1, \dots, 9$), das Alter a ($= 1, \dots, 19$), das Geschlecht g ($= 1, 2$), die fortlaufende Wohnungsnummer j (Kombination aus Pol. Bezirk und Wohnungskennzahl), die Staatsbürgerschaft s ($= 1, \dots, 5$) und das vorläufige Hochrechnungsgewicht aus (1) und (2) benötigt. Wir bezeichnen dieses Gewicht mit w_i^0 . Die Iteration besteht aus vier sich ständig wiederholenden Schritten. Zum Start wird der Laufindex $k=0$ gesetzt.

Schritt 1: Das nach Formel (1) und (2) errechnete Gewicht jeder Person wird durch Multiplikation mit einem Faktor so abgeändert, dass die mit den abgeänderten Gewichten hochgerechnete Verteilung der Bevölkerung nach Bundesland, Alter und Geschlecht mit den Vorgaben N_{bag} übereinstimmt.

$$w_i^{4k+1} = w_i^{4k} \frac{N_{b_i a_i g_i}}{\sum_l w_l^{4k}} \text{ für alle } i$$

Die Summierung erstreckt sich dabei über alle Datensätze l von Personen mit denselben Ausprägungen hinsichtlich der Merkmale Bundesland, Alter und Geschlecht wie der Satz i . Falls der Quotient $w_i^{4k+1}/w_i^0 > 4$ ist, wird $w_i^{4k+1} = 4w_i^0$ gesetzt, und falls $w_i^{4k+1}/w_i^0 < 1/4$ ist, wird $w_i^{4k+1} = w_i^0/4$ gesetzt.⁴⁾

Schritt 2: Die im Schritt 1 erzielten Gewichte werden multiplikativ so abgeändert, dass die mit den entstehenden Gewichten hochgerechnete Verteilung der Bevölkerung nach Bundesland und Staatsbürgerschaft mit der Vorgabe N_{bs} übereinstimmt. Die Veränderung der ursprünglichen Ge-

wichte w_i^0 soll wieder das Vierfache nicht überschreiten bzw. ein Viertel nicht unterschreiten. Es gilt

$$w_i^{4k+2} = w_i^{4k+1} \frac{N_{b_i s_i}}{\sum_l w_l^{4k+1}} \text{ für alle } i$$

Die Summierung erstreckt sich dabei über alle Datensätze l von Personen mit denselben Ausprägungen hinsichtlich der Merkmale Bundesland und Staatsbürgerschaft wie der Satz i . Falls der Quotient $w_i^{4k+2}/w_i^0 > 4$ ist, wird $w_i^{4k+2} = 4w_i^0$ gesetzt, und falls $w_i^{4k+2}/w_i^0 < 1/4$ ist, wird $w_i^{4k+2} = w_i^0/4$ gesetzt. Durch diese Anpassung geht diejenige hinsichtlich Alter und Geschlecht aus Schritt 1 im Allgemeinen wieder verloren.

Schritt 3: Durch die in den Schritten 1 und 2 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 und der gleichen Staatsbürgerschaft 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 3. Schritt alle Personen einer Wohnung als neues Gewicht den Durchschnitt ihrer Gewichte nach Abschluss des 2. Schrittes erhalten.

$$w_i^{4k+3} = \frac{\sum_l w_l^{4k+2}}{N_j} \text{ für alle } i$$

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 4: Die im Schritt 3 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 fünf Bewohnern wird je Bundesland mit einer Unschärfe von 5% ($p_5=0,05$) angenommen, die Zahl der Wohnungen mit sechs und mehr Bewohnern wird je Bundesland mit einer Unschärfe von 15% ($p_6=0,15$) angenommen. Die Veränderung der ursprünglichen Gewichte w_i^0 soll wieder

⁴⁾ Die Wahl dieser Schranken hat sich in der Vergangenheit im Rahmen des Mikrozensus bewährt. Eine Einschränkung der Veränderung der Basisgewichte ist auch in anderen Ländern üblich. Vgl. hiezu den Artikel von A. Afentakis u. W. Bihler „Das Hochrechnungsverfahren beim unterjährigen Mikrozensus ab 2005“ in *Wirtschaft und Statistik*, Heft 10/2005, S. 1039 ff.

das Vierfache nicht überschreiten bzw. ein Viertel nicht unterschreiten. Es gilt

$$w_i^{4k+4} = \begin{cases} w_i^{4k+3} \frac{M_{b_i h_i}}{\sum_l w_l^{4k+3}}, & \text{falls } \sum_l w_l^{4k+3} \notin ((1 - p_h) M_{b_i h_i}, (1 + p_h) M_{b_i h_i}) \\ w_i^{4k+3}, & \text{sonst} \end{cases}$$

Die Summierung erstreckt sich dabei über alle Datensätze l von Wohnungen mit denselben Ausprägungen hinsichtlich der Merkmale Bundesland und Bewohneranzahl wie der Datensatz i . Falls der Quotient $w_i^{4k+4} / w_i^0 > 4$ ist, wird $w_i^{4k+4} = 4 w_i^0$ gesetzt, und falls $w_i^{4k+4} / w_i^0 < 1/4$ ist, wird $w_i^{4k+4} = w_i^0 / 4$ gesetzt.

Schritt 5: 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.

$$\max_{\text{bag}} \left| \frac{\sum_l w_l^{4k+4} - N_{\text{bag}}}{N_{\text{bag}}} \right| > 0,0001$$

Die Summierung erstreckt sich dabei in der ersten Ungleichung über alle Personen mit denselben Ausprägungen der Merkmale b , a und g und in der zweiten Ungleichung über alle Personen mit denselben Ausprägungen der Merkmale b und s .

Falls wenigstens eine der beiden Bedingungen 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 und 3 etwa die Anpassung an die Altersverteilung wieder zerstört wird. Das Verfahren beginnt also wieder mit Schritt 1, wobei anstatt des Gewichts aus Formel (1) und (2) nun das nach Durchlaufen des 4. Schrittes erzeugte Gewicht verwendet wird. Ist keine der beiden Bedingungen erfüllt, so ist das definitive Hochrechnungsgewicht mit w_i^{4k+4} gefunden und damit das Ziel der Iteration erreicht.

Obiges Verfahren wird nach den bisherigen Erfahrungen etwa 50- bis 100-mal durchlaufen, bis die so definierte Konvergenz erreicht ist. Nach 100 Iterationen wird auf jeden Fall abgebrochen. Die in die Iteration eingebaute Sicherheitsschranke bewirkt, dass die Gewichte nach Formel (1) und (2) durch die Iteration um nicht mehr als das Vierfache erhöht werden. Nach Abschluss der Iteration wird jeder Datensatz mit dem definitiven Hochrechnungsgewicht ange-

reichert, wie es sich für die Privathaushalte aufgrund der Iteration ergibt.

Verwendung der Hochrechnungsgewichte bei Quartals- und Jahresauswertungen:

Die quartalsweise Berechnung des Umfangs bestimmter Haushalts- oder Personenmassen in der Grundgesamtheit geschieht durch Addition der endgültigen Gewichte aller Elemente der entsprechenden Stichprobenmasse. Die Totalwerte quantitativer Merkmale ergeben sich durch Multiplikation mit den Hochrechnungsgewichten und anschließende Addition, die Quotienten durch Division von Totalwerten bzw. Summen.

Die vierteljährlich stattfindenden Befragungen im Rahmen des Mikrozensus werden für manche Themenkreise zur Er-

$$\max_{\text{bs}} \left| \frac{\sum_l w_l^{4k+4} - N_{\text{bs}}}{N_{\text{bs}}} \right| > 0,0001$$

stellung 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.

Genauigkeit und Fehlerrechnung

Die bei der Arbeitskräfteerhebung erhobenen Merkmale sind größtenteils qualitativer bzw. diskreter Natur. Die Auswertung dieser Merkmale erfolgt vorwiegend durch Berechnung von absoluten bzw. relativen Häufigkeitstabellen, die für jede Ausprägungskombination der tabellenerzeugenden Merkmale die darauf entfallende hochgerechnete Fallzahl angeben. Diese hochgerechneten Werte sind mit einer gewissen Unschärfe behaftet, da ihre Berechnung nicht auf Basis einer Vollerhebung, sondern einer Stichprobe erfolgt. Diese Unschärfe wird hier mit dem relativen Stichprobenfehler bei 95-prozentiger Sicherheit (das 1,96-fache des Quotienten aus Standardfehler dividiert durch die geschätzte Häufigkeit) bzw. mit dem Variationskoeffizienten (= relativer Stichprobenfehler bei 68-prozentiger Sicherheit und entspricht dem Quotienten aus Standardfehler dividiert durch die geschätzte Häufigkeit) gemessen.

Der relative Stichprobenfehler für solche Häufigkeitstabellen lässt sich sowohl für Wohnungsmassen als auch für Personenmassen unter der Annahme einer einfachen Zufallsauswahl von Wohnungen bzw. Personen mit Hilfe des Binomialansatzes⁵⁾ abschätzen. Bei diesem Verfahren werden also die Effekte der komplexen Hochrechnung (iterative Anpassung an vorgegebene Verteilungen der Personen und Wohnungen) und der geklumpten Auswahl von Personen ausgeklammert (bei der Schätzung des Stichprobenfehlers für Personenhäufigkeiten beide Effekte, bei Wohnungen nur der Effekt der komplexen Hochrechnung).

Berücksichtigt werden hingegen die pro Bundesland unterschiedlichen Auswahlätze. Nach diesem Ansatz ergibt sich näherungsweise der geschätzte relative Stichprobenfehler in Prozent bei 95-prozentiger statistischer Sicherheit für eine hochgerechnete Fallzahl x (Wohnungen oder Personen) nach der Formel

$$200 \sqrt{\frac{(N_b - n_b) (N_b - x)}{N_b n_b x}} \quad \text{für Bundeslandergebnisse und}$$

⁵⁾ Streng genommen sind die absoluten Häufigkeiten bei einem Ziehungsverfahren ohne Zurücklegen hypergeometrisch verteilt; da aber der Auswahlatz sehr klein ist, ist die Verwendung der Binomialverteilung gerechtfertigt.

$$200 \sqrt{\sum_b \frac{(N_b - n_b) (N - x) N_b}{n_b N^2 x}} \quad \text{für österreichbezogene Häufigkeiten x}$$

Dabei bezeichnen N_b die Grundgesamtheit im Bundesland b, n_b den Stichprobenumfang im Bundesland b und N die Grundgesamtheit in Österreich.

Im Falle der Wohnungen sind in den obigen Formeln die Größen N, N_b , n und n_b durch M, M_b , m 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.

Die mittels obiger Formeln errechneten relativen Stichprobenfehler für hochgerechnete Häufigkeiten von Personen bzw. Wohnungen sind in den *Tabellen 3 und 4* dargestellt. Daraus lässt sich der Zusammenhang zwischen Merkmalshäufigkeit und Stichprobenfehler für Bundesland- und Österreicherergebnisse näherungsweise abschätzen. Hierbei bezieht sich die *Tabelle 3* auf die Grundgesamtheit der bewohnten Wohnungen und die *Tabelle 4* auf 8 Mio. Personen. Für ein Merkmal, das in der österreichischen Bevölkerung auf 20.000 Personen zutrifft, ergibt sich also

Relativer Stichprobenfehler hochgerechneter Wohnungszahlen bei 95% statistischer Sicherheit

Tabelle 3



Wohnungen in 1.000	Österreich	Burgenland	Kärnten	Nieder-österreich	Ober-österreich	Salzburg	Steiermark	Tirol	Vorarlberg	Wien
	Relativer Stichprobenfehler in Prozent									
1	52,3	65,5	100,0	100,0	64,1	91,1	71,6	50,0	117,9	94,9
2	36,8	46,2	70,7	70,7	45,3	64,4	50,5	35,3	83,3	67,1
3	29,9	37,7	57,7	57,7	36,9	52,5	41,2	28,7	68,0	54,8
4	25,8	32,6	49,9	49,9	31,8	45,4	35,6	24,7	58,9	47,4
5	22,9	29,1	44,6	44,6	28,4	40,6	31,8	22,1	52,6	42,4
6	20,8	26,5	40,7	40,7	25,9	37,0	28,9	20,1	48,0	38,7
7	19,2	24,4	37,6	37,6	23,9	34,2	26,7	18,5	44,4	35,9
8	17,9	22,8	35,2	35,1	22,3	32,0	25,0	17,2	41,5	33,5
9	16,8	21,5	33,1	33,1	21,0	30,1	23,5	16,2	39,1	31,6
10	15,8	20,3	31,4	31,4	19,9	28,5	22,2	15,3	37,1	30,0
20	10,6	14,0	22,0	22,0	13,7	20,0	15,4	10,4	26,1	21,2
30	8,2	11,2	17,8	17,8	10,9	16,1	12,3	8,1	21,1	17,3
40	6,6	9,4	15,3	15,3	9,2	13,8	10,5	6,7	18,2	14,9
50	5,4	8,2	13,6	13,5	8,0	12,2	9,2	5,7	16,2	13,3
60	4,5	7,3	12,3	12,2	7,0	11,0	8,2	4,9	14,6	12,2
70	3,7	6,5	11,3	11,2	6,3	10,1	7,4	4,2	13,5	11,2
80	3,0	5,9	10,5	10,4	5,7	9,3	6,7	3,7	12,5	10,5
90	2,3	5,4	9,8	9,7	5,2	8,7	6,2	3,2	11,7	9,9
100	1,4	4,9	9,2	9,1	4,7	8,1	5,7	2,7	11,0	9,4
200		1,7	5,9	5,7	1,2	4,9	2,6		7,2	6,5
300			4,2	3,9		3,2			5,4	5,2
400			3,0	2,7		1,9			4,2	4,5
500			2,1	1,4					3,2	3,9
600			0,9						2,4	3,5
700									1,5	3,2
800										2,9
900										2,7
1.000										2,5
2.000										1,4
3.000										0,6

Relativer Stichprobenfehler hochgerechneter Personenzahlen bei 95% statistischer Sicherheit

Tabelle 4



Personen in 1.000	Österreich	Burgenland	Kärnten	Nieder- österreich	Ober- österreich	Salzburg	Steiermark	Tirol	Vorarlberg	Wien
	Relativer Stichprobenfehler in Prozent									
2	36,2	45,0	67,8	68,6	44,0	62,2	49,2	34,3	82,3	64,5
3	29,5	36,7	55,4	56,0	35,9	50,8	40,1	28,0	67,2	52,6
4	25,5	31,8	47,9	48,5	31,0	44,0	34,7	24,2	58,1	45,6
5	22,8	28,4	42,9	43,4	27,7	39,3	31,0	21,6	52,0	40,8
6	20,7	25,9	39,1	39,6	25,3	35,9	28,3	19,7	47,4	37,2
7	19,2	23,9	36,2	36,6	23,4	33,2	26,2	18,2	43,9	34,4
8	17,9	22,4	33,8	34,2	21,9	31,0	24,5	17,0	41,1	32,2
9	16,8	21,1	31,9	32,3	20,6	29,2	23,1	16,0	38,7	30,4
10	15,9	20,0	30,3	30,6	19,5	27,7	21,9	15,2	36,7	28,8
20	11,1	14,0	21,3	21,6	13,7	19,5	15,3	10,6	25,9	20,4
30	8,9	11,3	17,4	17,5	11,0	15,9	12,4	8,5	21,1	16,6
40	7,5	9,7	15,0	15,1	9,5	13,7	10,7	7,2	18,2	14,4
50	6,6	8,6	13,4	13,5	8,4	12,2	9,5	6,4	16,2	12,9
60	5,9	7,8	12,1	12,3	7,6	11,1	8,6	5,7	14,7	11,7
70	5,3	7,1	11,2	11,3	6,9	10,2	7,9	5,2	13,6	10,9
80	4,8	6,6	10,4	10,5	6,4	9,5	7,3	4,8	12,7	10,1
90	4,4	6,1	9,8	9,9	6,0	8,9	6,8	4,4	11,9	9,6
100	4,1	5,8	9,3	9,4	5,6	8,4	6,4	4,1	11,3	9,1
200	1,9	3,6	6,3	6,3	3,4	5,7	4,1	2,3	7,7	6,4
300		2,5	5,0	5,0	2,3	4,4	3,0	1,1	6,1	5,2
400		1,7	4,1	4,1	1,5	3,6	2,2		5,0	4,4
500		0,9	3,5	3,5	0,5	3,0	1,6		4,3	3,9
600			3,1	3,0		2,5	1,0		3,8	3,6
700			2,7	2,6		2,1			3,3	3,3
800			2,4	2,2		1,8			2,9	3,1
900			2,1	1,9		1,4			2,6	2,9
1.000			1,8	1,6		1,1			2,3	2,7
2.000										1,8
3.000										1,3
4.000										1,0
5.000										0,8
6.000										0,6
7.000										0,4
8.000										0,1

unter den genannten Bedingungen ein Unsicherheitsbereich von etwa $\pm 20\%$. Für kleinere Personengruppen wird der Stichprobenfehler entsprechend höher. Die Vertrauensbereiche sind derart angegeben, dass der Stichprobenfehler mit 95% Sicherheit innerhalb der angegebenen Grenzen liegt, sofern die Effekte der geklumpten Auswahl von Personen und der gebundenen Hochrechnung vernachlässigbar klein sind.

Zwei Beispiele sollen den **Gebrauch dieser Tabellen** demonstrieren: Bei einer Mikrozensushebung wurden in Österreich 57.000 Eigentumswohnungen mit einem Baujahr zwischen 1961 und 1970 festgestellt. Durch Interpolation ergibt sich aus *Tabelle 3*, dass der relative Stichprobenfehler etwa 12,5% oder $57.000 \cdot 0,125 = 7.125$ Wohnungen ausmacht. Mit einer Wahrscheinlichkeit von 95% würde sich also die bei einer Vollerhebung ermittelte Zahl von Wohnungen dieses Typs um höchstens 7.125 Wohnungen von den 57.000 Wohnungen unterscheiden.

Bei der gleichen Mikrozensushebung wurden im Burgenland 14.200 Selbständige gezählt. Aus *Tabelle 4* folgt, dass der relative Stichprobenfehler etwa 13% oder $14.200 \cdot 0,13 =$

1.846 Personen ausmacht. Die wahre Anzahl der Selbständigen im Burgenland liegt demnach mit einer Sicherheit von (höchstens)⁶⁾ 95% zwischen 12.354 und 16.046 Personen.

Auch für die Berechnung des Stichprobenfehlers der hochgerechneten Häufigkeiten anderer Untersuchungsobjekte wie etwa von Haushalten oder Familien ist näherungsweise die *Tabelle 3* heranzuziehen.

Es wird nochmals darauf hingewiesen, dass die Tabellen nur einen Hinweis auf die **Größenordnung des Fehlers** und nicht den exakten Wert angeben. Bei der Berechnung der Tabellen wurde beispielsweise die Anpassung an die fortgeschriebene Wohnbevölkerung nicht berücksichtigt. Häufigkeiten, die Umfänge von Massen repräsentieren, die dieser Anpassung zugrunde liegen, weisen natürlich keinen Zufallsfehler auf (z.B. die Bevölkerungszahl gewisser Altersklassen). Auch sind die Fehler vergrößernden Effekte der Klumpung nicht berücksichtigt, weil dazu kaum generelle Aussagen möglich sind. Diese Elemente des Stichprobenplans

⁶⁾ Höchstens 95%, weil der Stichprobenfehler für Personenhäufigkeiten mit diesem Verfahren tendenziell unterschätzt wird.

wirken sich auf jedes Merkmal und jede Untergliederung einer Tabelle unterschiedlich aus.

Will man den Stichprobenfehler genauer abschätzen, indem man alle Aspekte der Stichprobenziehung und Hochrechnung berücksichtigt und/oder nichtlineare Schätzfunktionen z.B. für Quotienten verwendet, muss man aufgrund der Komplexität des Designs auf andere Methoden (Taylorlinearisierung, Bootstrapverfahren) zugreifen.

Dabei ergab sich für einige wichtige **Kennziffern der Arbeitskräfteerhebung im 4. Quartal 2004** folgende Schätzung der relativen Stichprobenfehler in Prozent (bei 95-prozentiger statistischer Sicherheit):⁷⁾

Übersicht 1		
Kennziffer	Anzahl/Anteil	Relativer Stichprobenfehler in %
Erwerbstätige	3.771.400	1,2
Teilzeit-Erwerbstätige	805.700	3
Arbeitslose	195.800	7,7
Arbeitslosenquote in Prozent	4,9	7,6
Durchschnittliche Arbeitszeit in Stunden	34,8	0,8

Für die hochgerechnete Zahl an **Arbeitslosen** betrug der relative Stichprobenfehler im Jahr **2004** je Bundesland:

Übersicht 2				
Bundesland	Arbeitslose im 4. Quartal 2004	Relativer Stichprobenfehler in % bei 95% statistischer Sicherheit	Arbeitslose im Jahresdurchschnitt 2004	Relativer Stichprobenfehler in % bei 68% statistischer Sicherheit
Burgenland	7.800	19,6	7.300	6,9
Kärnten	11.200	19,1	11.900	6,2
Niederösterreich	29.700	21,4	32.000	6,8
Oberösterreich	22.300	25,6	25.000	7,2
Salzburg	9.500	24,6	9.600	7,7
Steiermark	22.000	21,2	20.900	7,9
Tirol	12.100	21,8	11.100	7,8
Vorarlberg	7.700	20,1	7.400	7,3
Wien	73.400	14,2	69.300	5,3

Durch den Vergleich des relativen Stichprobenfehlers von hochgerechneten Häufigkeiten von Personen aus den obigen Übersichten mit den grob geschätzten Werten gemäß Tabelle 4 ergibt sich, dass bei einer genaueren Fehlerrechnung von Quartalswerten der relative Fehler bei den betrachteten Merkmalen um 10% bis 20% höher ausfällt als bei einer groben Schätzung mit Hilfe der Tabelle 4.

⁷⁾ Für die Schätzung wurde sowohl ein in der STATISTIK AUSTRIA erstelltes SAS-Programm verwendet, das methodisch auf V. Verma: The Estimation and Presentation of Sampling Errors, UN World Fertility Survey, Technical Bulletins No. 11, New York 1982 beruht als auch das von Statistics Sweden entwickelte Programm CLAN 97.

Für **Jahresdurchschnittswerte** errechnet sich der relative Stichprobenfehler $\epsilon(x_j)$ aus dem relativen Stichprobenfehler eines Quartalswerts $\epsilon(x_Q)$ näherungsweise nach der Formel:

$$\epsilon(x_j) = \epsilon(x_Q) \sqrt{\frac{4 + 2 \sum_{i>j} P_{ij} R_{ij}}{16}}$$

Dabei bezeichnet P_{ij} die Überschneidung der Stichproben der Quartale i und j und R_{ij} die Korrelation der Merkmalswerte zwischen den Quartalen i und j . Unter der Annahme, dass die R_{ij} nicht allzu unterschiedlich sind, und da die P_{ij} bei einer Fünftelrotation von Quartal zu Quartal feststehen, vereinfacht sich die Formel zu:

$$\epsilon(x_j) = \epsilon(x_Q) \sqrt{\frac{4 + 8\bar{R}}{16}}$$

Für $\bar{R} = 1$ gilt daher $\epsilon(x_j) = \epsilon(x_Q) \sqrt{0,75}$

(siehe hierzu L. Kish (1965), Survey Sampling, S. 466 ff.).

Den relativen Stichprobenfehler von Jahresdurchschnittswerten hochgerechneter Wohnungs- und Personenzahlen kann man also näherungsweise bestimmen, indem man die Werte der Tabellen 3 und 4 mit dem Faktor $\sqrt{0,75} \approx 0,85$ multipliziert. Bei diesem Vorgehen liegt man zumindest bei Wohnungsmerkmalen auf der sicheren Seite, d.h. die mit diesen Werten berechneten Vertrauensbereiche schließen den wahren Wert mit mindestens 95% statistischer Sicherheit ein.

In der Übersicht 2 sind die aus dem Mikrozensus hochgerechneten Arbeitslosen je Bundesland im 4. Quartal und im Jahresdurchschnitt 2004 mit ihrem Stichprobenfehler ausgewiesen. Der Stichprobenfehler ist für die Quartalswerte für eine statistische Sicherheit von 95% ausgewiesen (wie in den Publikationen der STATISTIK AUSTRIA üblich) und für die Jahresdurchschnitte für eine statistische Sicherheit von 68% (wie meist von EUROSTAT publiziert). Der relative Stichprobenfehler der Arbeitslosenzahlen (bei 68% statistischer Sicherheit) liegt im Jahresdurchschnitt in jedem Bundesland unter den in der Verordnung (EG) Nr. 577/98 verlangten 8%.

Das derzeitige Design des österreichischen Mikrozensus erfüllt auch die zweite von der EU festgelegte Genauigkeitsanforderung, dass nämlich die Ungenauigkeit der Schätzung einer Veränderung zwischen zwei Quartalen auf nationaler Ebene höchstens 3% betragen darf, sofern es sich um Bevölkerungsgruppen handelt, die 5% der Bevölkerung im erwerbsfähigen Alter entsprechen (für Österreich sind dies ca. 275.000 Personen). Eine Bevölkerungsgruppe dieses Umfangs ist etwa die Masse der selbständig Erwerbstätigen ohne Arbeitnehmer. Die Veränderung dieser Gruppe zwischen zwei benachbarten Quartalen wird für Österreich insgesamt mit einer Fehlermarge von 2,4% geschätzt.

Summary

Starting with 2004 the Austrian micro-census was completely reorganized. The present article deals with methodological and practical aspects of the sample design, of sample selection and of grossing up and variance estimation of the “new” micro-census. The sample design is regionally stratified and within each stratum a given number of private households is randomly selected. Each quarter the most recent copy of the Austrian central registration register is used as frame of selection (each quarter rotates one fifth of the total sample size in and out). The response rate was about 90% in 2005. For the estimation of the results weights are used which are calibrated with the population counts of the register of residents and with household forecasts of STATISTICS AUSTRIA. Calibration is done by iterative proportional fitting. For variance estimation for totals of persons and households some simple methods are presented. For some main results of the micro-census (e.g. number of unemployed) a more exact variance estimation was done, which proves that the Austrian micro-census fulfills the requirements for the relative standard error stipulated in the EU regulation Nr. 577/98.