

Gewichtungsverfahren zur Hochrechnung von EU-SILC-Querschnittergebnissen

THOMAS GLASER
MATTHIAS TILL

In Österreich wurde die Gemeinschaftsstatistik über Einkommen und Lebensbedingungen (EU-SILC - „Community Statistics on Income and Living Conditions“) erstmals im Jahr 2003 als einmalige Querschnitterhebung durchgeführt. Seither dient EU-SILC als zentrale Datenquelle für die Berichterstattung über Einkommen, Armut und Lebensbedingungen und als Basis für Indikatoren zur sozialen Eingliederung wie beispielsweise die Armutsgefährdungsquote. Im Jahr 2004 wurde EU-SILC nach einer für alle Staaten verpflichtenden EU-Verordnung durchgeführt und der Aufbau einer rotierenden Panelstichprobe begonnen, um die verpflichtende Quer- und Längsschnitterhebung in einem integrierten Design durchführen zu können. Mit der EU-SILC-Erhebung 2007 wurde das rotierende Panel mit vier Teilstichproben vollständig implementiert. Seither wird jedes Jahr die über einen Zeitraum von vier Jahren beobachtete Panelstichprobe durch eine neue (Querschnitt-) Teilstichprobe von Privathaushalten ersetzt. Dieses Design impliziert ein komplexes Gewichtungsverfahren, bei dem unterschiedliche Auswahlwahrscheinlichkeiten sowie selektive Ausfälle kompensiert werden müssen. Schlussendlich wurde die Hochrechnung an den Mikrozensus sowie an Daten des Hauptverbands der österreichischen Sozialversicherungsträger gebunden, um möglichst unverzerrte Schätzungen für die Bevölkerung in Privathaushalten zu gewährleisten. Im vorliegenden Artikel werden die Grundzüge des von Eurostat vorgegebenen Verfahrens und die notwendigen Schritte zur Ermittlung der Gewichte eines Erhebungsjahres erläutert sowie die Eigenschaften der Gewichte für EU-SILC 2008 in jedem Schritt erklärt.

Einleitung

EU-SILC (engl. „Community Statistics on Income and Living Conditions“) wurde in Österreich bereits im Jahr 2003 erstmals durchgeführt.¹⁾ Ab dem Jahr 2004 verpflichtet eine gemeinsame Rahmenverordnung²⁾ des Europäischen Rates und des Europäischen Parlaments (1177/2003) alle Mitgliedstaaten, solche Daten zu Einkommen und Lebensbedingungen zu erheben, Indikatoren zu berechnen und gemeinsam mit detaillierten Qualitätsberichten an Eurostat zu übermitteln. Die Aufgabe dieser Erhebung ist es, ein umfassendes Bild der sozialen Eingliederung in Österreich zu geben. Der bekannteste und auch einer der wichtigsten durch EU-SILC erstellten Indikatoren ist die Armutsgefährdungsquote.³⁾

Die Stichprobe von EU-SILC besteht aus privaten Haushalten. Im Jahr 2008 umfasste die Bruttostichprobe in Österreich insgesamt 8.099 Haushalte, das entspricht einem Auswahlatz von etwa 1/435 bzw. rund 2,3 Promille. Die Er-

hebung wird in Form einer Haushaltsbefragung durchgeführt. Die Teilnahme ist für die Auskunftspersonen freiwillig. 2008 schlossen insgesamt 5.711 Haushalte die erforderlichen Interviews ab und machten für die Analyse verwertbare Angaben. Die Interviews erfolgen für alle Haushaltsmitglieder ab 16 Jahren mittels computerunterstützten persönlichen Interviews (CAPI) bzw. computerunterstützten telefonischen Interviews (CATI).⁴⁾

Die EU-Verordnung verlangt von allen Staaten neben verlässlichen Querschnittergebnissen auch Paneldaten.⁵⁾ Dazu ist es notwendig, dass für die Personen der Ersterhebungsstichprobe für mindestens vier aufeinanderfolgende Jahre Informationen zur Verfügung stehen. Wie in den meisten EU-Staaten wird in Österreich eine integrierte Quer- und Längsschnitterhebung durchgeführt. Demnach kann EU-SILC in Österreich als eine rotierende Panelerhebung beschrieben werden, bei der Personen einer Rotationsstichprobe für jeweils vier aufeinanderfolgende Jahre in der Stich-

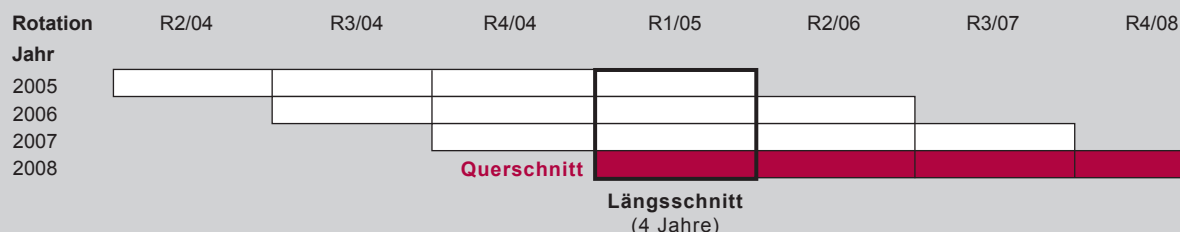
¹⁾ Für eine ausführliche Beschreibung zum Start der EU-SILC-Erhebung vgl. *Statistische Nachrichten*, Heft 3/2005, S. 224-231.

²⁾ Die Rahmenverordnung wird durch zahlreiche von der Europäischen Kommission festgelegte Durchführungsverordnungen ergänzt.

³⁾ Vgl. *STATISTIK AUSTRIA 2009b*, S. 50 ff.

⁴⁾ Details zur Methodik finden sich in *STATISTIK AUSTRIA 2010*.

⁵⁾ Panelinformationen werden insbesondere benötigt, um den Anteil der im Zeitraum von vier aufeinanderfolgenden Jahren dauerhaft armutsgefährdeten Personen berechnen zu können, vgl. *Statistische Nachrichten*, Heft 1/2010, S. 31-40.



Q: EU-SILC 2005-2008

probe verbleiben. Somit wird jährlich etwa ein Viertel der Stichprobe durch eine neue Stichprobe ersetzt.⁶⁾ Dieses Design ermöglicht zwei analytische Perspektiven: Einerseits können die erhobenen Daten jahresweise im Querschnitt ausgewertet werden, andererseits können Merkmale auch über unterschiedlich lange Zeitverläufe beobachtet werden. In jedem Erhebungsjahr gibt es eine Querschnittstichprobe, welche aus vier Rotationen besteht. Darunter befinden sich drei jeweils bereits zum zweiten, dritten, bzw. vierten Mal befragte Rotationen sowie eine neu gezogene Teilstichprobe. *Grafik 1* zeigt dieses Design anhand der Jahre 2005 bis 2008.

Nach einer einmalig durchgeführten Querschnitterhebung im Jahr 2003 startete das integrierte Quer- und Längsschnittdesign von EU-SILC in Österreich im Jahr 2004. Im Erhebungsjahr 2007 gab es somit zum ersten Mal eine über vier Jahre befragte Teilstichprobe. Diese Panelrotation (R4/04) wurde beendet und im Jahr 2008 durch eine neue Rotation (R4/08) ersetzt. Der Querschnitt (*in Grafik 1 rot dargestellt*) bestand 2008 aus den folgenden vier Rotationen: Rotation 1 (Ersterhebung 2005), Rotation 2 (Ersterhebung 2006), Rotation 3 (Ersterhebung 2007) und Rotation 4 (Ersterhebung 2008). Aufgrund dieses komplexen Designs muss die Gewichtung der Daten in mehreren Schritten durchgeführt werden. In jedem dieser Schritte wird eine spezifische Phase der Erhebung in der Gewichtung berücksichtigt. Der vorliegende Artikel beschränkt sich auf eine überblicksartige Darstellung des derzeit gültigen Hochrechnungsverfahrens für Querschnittergebnisse. Auf eine detailliertere Beschreibung der zwischen den Rotationen und Erhebungsjahren jeweils etwas unterschiedlichen Ausfallprozesse sowie der für Längsschnittanalysen modifizierten Gewichte muss hier verzichtet werden.

Gewichtungsverfahren

Das einem Stichprobenelement zugewiesene Gewicht entspricht jener Anzahl der Erhebungseinheiten der Grundge-

⁶⁾ Aufgrund von Ausfallsprozessen und notwendigen Anpassungen des Stichprobenumfangs sind die Teilstichproben in der Regel unterschiedlich groß und die jeweils erstmalig erhobene Teilstichprobe meist deutlich größer als die bereits über mehrere Jahre befragten Panelrotationen.

samtheit, die durch das Stichprobenelement repräsentiert werden. Die drei Hauptmotive für die Gewichtung von Erhebungsdaten sind das Stichprobendesign, die selektive Beteiligung an der Erhebung und die Kohärenz mit externen Informationen. Bei einer einfachen Zufallsauswahl (*engl. „simple random sample“ bzw. SRS*) hat jedes Element der Grundgesamtheit dieselbe Wahrscheinlichkeit, in die Stichprobe zu gelangen. In diesem Fall erfolgt die Hochrechnung von einer Stichprobe n auf die Grundgesamtheit N durch den inversen Auswahlssatz, also den für alle Stichprobeneinheiten konstanten Faktor N/n . Bei einer geschichteten Auswahl und insbesondere, wenn die Aufteilung der Adressen nicht gemäß der Verteilung der Haushalte erfolgt (disproportionale Allokation), sind die Auswahlwahrscheinlichkeiten der Stichprobeneinheiten in der Regel unterschiedlich. Nur wenn diese Unterschiede bei der Hochrechnung berücksichtigt werden, ist es möglich, unverzerrte Schätzungen für die Grundgesamtheit auf Basis der Stichprobe zu erlangen.⁷⁾

Insbesondere bei freiwillig durchgeführten Erhebungen wie EU-SILC sind nicht alle für die Befragung ausgewählten Haushalte erreichbar oder können motiviert werden, an der Erhebung teilzunehmen. Eine entsprechende Gewichtung wirkt daher Verzerrungen aufgrund von selektiven Ausfällen und Antwort-Verweigerungen (*engl. „non-response“*)⁸⁾ entgegen.

Schließlich kann durch gebundene Hochrechnung (bzw. Kalibrierung) sichergestellt werden, dass die Verteilung bestimmter soziodemographischer Merkmale mit anderen, besonders zuverlässigen Quellen übereinstimmt und somit

⁷⁾ Zur Notwendigkeit von Gewichtung *vgl. Kish 1990, S. 123 ff.*

⁸⁾ Genaugenommen handelt es sich dabei um die sog. „unit non-response“, bei denen eine Haushalt bzw. eine Person nicht an der Befragung teilnimmt. Diese „Gesamtausfälle“ dürfen nicht mit der „item non-response“ verwechselt werden, welche die Antwortverweigerung bestimmter Fragen beschreibt. Immer dann, wenn im vorliegenden Artikel von Nonresponse die Rede ist, sind damit Gesamtausfälle, also *unit non-response* gemeint. Als weitere Nonresponse-Art ist der Ausfall einzelner Personen eines Mehrpersonen-Haushalts zu nennen. Diese Ausfälle betreffen nur wenige Personen und werden ebenso wie Item-Nonresponse mittels eines Imputationsverfahrens ausgeglichen.

zufallsbedingte Abweichungen ausgeschlossen werden. Sofern die verwendeten Verteilungen in Zusammenhang mit den spezifischen Erhebungsmerkmalen stehen (z.B. Arbeitslosengeldbezug und Armutsgefährdung), kann die gebundene Hochrechnung auch dazu beitragen, nicht anders zu behebende systematische Verzerrungen zu minimieren. Wenn die Bindung für mehrere Merkmale erfolgen soll, müssen diese Anpassungen meist iterativ wiederholt werden (engl. „*raking*“), bis eine zufriedenstellende Konvergenz aller Merkmalsverteilungen der Stichprobe mit externen Quellen erzielt wird. Eine direkte Bindung von Zelhäufigkeiten (Poststratifikation) ist nur möglich, wenn die Zahl der Merkmale begrenzt ist und verlässliche Informationen über die gemeinsame Verteilung der Merkmalsausprägungen vorliegen. In den meisten Umfragen von Marktforschungsinstituten und selbst beim österreichischen Mikrozensus wird auf Design-Gewichtung und Nonresponse-Gewichtung verzichtet und die gebundene Hochrechnung direkt vorgenommen.⁹⁾ Diese Vereinfachung ist allerdings nur dann gerechtfertigt, wenn die Selektivität von Ausfällen als sehr gering angenommen werden kann und die zur gebundenen Hochrechnung verwendeten Eckzahlen weitestgehend der für die Ziehung verwendeten Auswahlgrundlage entsprechen. Besonders kritisch sind Abweichungen jedoch bei Panel-Erhebungen, da sich hier Stichprobenausfälle (engl. „*attrition*“) im Zeitverlauf kumulieren und die ursprüngliche Auswahlgrundlage immer stärker von der aktuellen Bevölkerungsstruktur abweicht (engl. „*coverage problem*“).

Die für EU-SILC in Österreich angewendete Gewichtungszprozedur baut auf der Erfahrung des Europäischen Haushaltspanels,¹⁰⁾ der Vorgängererhebung von EU-SILC, auf und implementiert das von Eurostat empfohlene Verfahren.¹¹⁾ Es handelt sich dabei um eine „Fallgewichtung“ (Rösch 1994, S. 7). Jedes Element der Stichprobe erhält eine reelle Zahl, die dann als Hochrechnungsfaktor für Auswertungen angewendet wird. Die Berechnung der Gewichte erfolgt bei EU-SILC in mehreren aufeinanderfolgenden Schritten. Das endgültige Gewicht ist das Produkt aus mehreren Faktoren. Die wichtigsten Gewichtungsschritte sind in der Verordnung 1982/2003 der Europäischen Kommission beschrieben: (EC No1982/2003, § 7.4): „*Weighting factors shall be calculated as required to take into account the units' probability of selection, non-response and, as appropriate, to adjust the sample to external data relating to the distribution of households and persons in the target population, (...)*“ (vgl. Eurostat 2010: S. 32).

Das Verfahren zur Gewichtung der EU-SILC-Daten gliedert sich somit im Wesentlichen in drei Schritte:

⁹⁾ Für eine Beschreibung des Gewichtungsverfahrens des Mikrozensus vgl. *Statistische Nachrichten*, Heft 6/2004, S. 510-519.

¹⁰⁾ Vgl. Till 2001.

¹¹⁾ Vgl. Eurostat 2010 bzw. Osier et al 2006.

- 1) Design-Gewicht; Inverse der Ziehungswahrscheinlichkeit: Mit Hilfe dieses Gewichts können unterschiedliche Auswahlwahrscheinlichkeiten im Zuge der Stichprobenziehung ausgeglichen werden.
- 2) Nonresponse-Gewicht; Inverse der Antwortwahrscheinlichkeit: Um eine Verzerrung aufgrund selektiver Ausfallmuster zu kompensieren, werden Haushalte/Personen mit unterschiedlich hoher/niedriger Antwortwahrscheinlichkeit verschieden niedrig/hoch gewichtet.
- 3) Anpassungs-Gewicht bzw. Kalibrierung: Die Verteilungen einiger wichtiger Variablen werden auf Haushalts- und Personenebene an externe Quellen angepasst, um Inkonsistenzen zu kompensieren und Verzerrungen zu verringern.

Das endgültige Querschnitt-Gewicht für ein bestimmtes Jahr der Erhebung ergibt sich schließlich als Produkt aus Design-Gewicht x Nonresponse-Gewicht x Anpassungs-Gewicht.

Gewichtungsschritte für die Erstbefragung

Schritt 1: Design-Gewichte für die ausgewählten Adressen der Erstbefragung

Die primären Stichprobeneinheiten (engl. „*primary sampling units - PSU*“) bestehen für die österreichische EU-SILC-Erhebung aus privaten Haushalten. Es handelt sich also um eine einstufige, nicht geklumpte Auswahl. Die Stichproben für die Erstbefragungen der Jahre 2003-2006 wurden als einfache Zufallsauswahl von Privathaushalten aus dem Zentralen Melderegister realisiert, und die Befragung wurde von externen Erhebungsinstituten durchgeführt. Im Jahr 2007 wurde die Erstbefragung zum ersten Mal von der hauseigenen Erhebungsinfrastruktur der STATISTIK AUSTRIA abgewickelt. Die Ziehung der Adressen erfolgte für die Erstbefragung 2007 pro Interviewer-Sprengel¹²⁾ (177 Schichten). Dies sollte eine möglichst gleichmäßige Auslastung der Erhebungspersonen gewährleisten. Die für jede Schicht zu erhebende Anzahl an Haushalten wurde im direkten Verhältnis zur Zahl der Haushalte im Auswahlrahmen festgelegt (proportionale Allokation). Für die Teilstichprobe der Ersterhebung 2008 wurde hingegen ein stratifiziertes Design mit $K=70$ Schichten angewendet. Diese Schichtung erfolgte nach mehreren Kriterien:

- (i) Bundesland
- (ii) Kleine/große „Interviewer-Sprengel“

¹²⁾ Interviewer-Sprengel werden als Einheiten im österreichischen Mikrozensus verwendet, um gezogene Adressen an die Erhebungspersonen zu verteilen. Die Grenzziehung der Sprengel richtet sich nach dem jeweils verfügbaren Interviewerpool und wird laufend adaptiert. Die Stichprobenziehung nach Sprengeln soll eine verhältnismäßig ausgewogene Anzahl an zugeordneten Adressen ermöglichen und liefert eine gleichmäßige geographische Verteilung der gezogenen Adressen.

- (iii) Ausgewählte Zielgruppenmerkmale innerhalb der großen Sprengel (Überrepräsentation für Haushalte mit Zugewanderten und Haushalte mit erhöhtem Risiko armutsgefährdet zu sein).¹³⁾

Die Anzahl der zu ziehenden Haushalte in den Schichten der Kriterien (i) und (ii) wurde mit Hilfe einer Zuteilung proportional zur Verteilung im Auswahlrahmen festgelegt. Aufgrund von Rundungsfehlern unterscheiden sich die tatsächlichen Auswahlätze der pro Schicht zu ziehenden Haushalte geringfügig.¹⁴⁾ Zur Überrepräsentation (engl. „oversample“) der Schichten nach Kriterium (iii) wurde ein höherer Auswahlatz f_s angewendet. Dieser ergibt sich aus der relativen Häufigkeit der pro Schicht zu ziehenden Haushalte n_s im Verhältnis zur Anzahl der pro Schicht vorhandenen Haushalte N_s .

$$f_s = \frac{n_s}{N_s} \quad s \in \{1, 2, \dots, K\} \quad (1)$$

Die Design-Gewichte d_s werden als Inverse des Auswahlatzes berechnet, um die n Stichprobenelemente auf die N Haushalte in der Grundgesamtheit hochzurechnen.¹⁵⁾ Die Auswahlätze liegen zwischen 1/511 und 1/1.719. Entsprechend diesen unterschiedlichen Auswahlätzen variieren auch die Gewichte von Schicht zu Schicht (die Eurostat-Bezeichnung der Design-Gewichte lautet DB080).

$$d_s = DB080_s = \frac{N_s}{n_s} \quad (2)$$

Schritt 2: Nonresponse-Gewichte für erstmalig befragte Haushalte

Antwortverweigerung ist bei Befragungen ein spezielles Problem. Die Teilnahme der Befragten ist bei EU-SILC (im Gegensatz zum Mikrozensus) freiwillig. Es kommt daher auch vor, dass Haushalte ihre Mitarbeit verweigern oder (vorübergehend) nicht erreichbar sind. Wenn ein Haushalt nicht an einer Befragungswelle teilnimmt, wird er auch von allen folgenden EU-SILC-Erhebungswellen ausgeschlossen.¹⁶⁾ Mit Fortdauer einer Panel-Erhebung kumulieren sich

¹³⁾ Da neu zugewanderte Personen nur in den jeweiligen Erstbefragungstichproben erfasst werden können, würden diese Personen bei der später durchgeführten Kombination der Teilstichproben höhere Gewichtungsfaktoren erhalten. Durch das Oversample wird versucht, die Streuung der letztlich für die Hochrechnung verwendeten Querschnitt-Gewichte etwas zu verringern. Nach den Prinzipien der sogenannten „Neyman allocation“ (Cochran 1977, 98 f.) verbessert der erhöhte Auswahlatz für Risikogruppen auch die Genauigkeit der Schätzergebnisse in Bezug auf die Armutsgefährdungsquote. Eine detailliertere Dokumentation des Stichprobendesigns findet sich in STATISTIK AUSTRIA 2010, S. 6 ff.

¹⁴⁾ In einzelnen, besonders kleinen Sprengeln wurden außerdem etwas höhere Auswahlätze angewendet, um für jede Erhebungsperson eine gewisse Mindestanzahl an Adressen zu gewährleisten.

¹⁵⁾ Vgl. Groves et al. 2004, S. 98.

¹⁶⁾ Bei EU-SILC 2007 wurde versucht, auch Haushalte zu befragen, die einmalig nicht an der Erhebung teilgenommen haben (z.B. 2005 oder 2006). Die Gewichte sogenannter „returnee“ Haushalte wurden aber im Zuge der Nonresponse-Gewichtung bereits auf die im Panel verbleibenden Haushalte aufgeteilt. Deshalb musste die Gewichtung des Panelausfalls in

diese Ausfälle. Verzerrungen (engl. „bias“) aufgrund von Nonresponse treten dann auf, wenn bestimmte, für die Erhebung wichtige Eigenschaften in den ausfallenden Haushalten über- bzw. unterrepräsentiert sind. Dies wird dann zu einem Problem, wenn solche Eigenschaften wesentlich für die Analyse der Daten sind. Zum Beispiel war die Teilnahmebereitschaft (bei Folgebefragungen) von Haushalten mit einer Haupteinkommensquelle aus selbständiger Beschäftigung geringer als bei jenen mit einer Haupteinkommensquelle aus unselbständiger Beschäftigung. Erstere enthalten auch viele Landwirte und sogenannte atypische Beschäftigungsformen (Werkverträge, freie Dienstverträge) und sind beinahe doppelt so häufig von Armutsgefährdung betroffen wie Personen, deren Einkommen aus hauptsächlich unselbständiger Tätigkeit stammt. Der Wegfall von Stichprobeneinheiten aufgrund von Antwortausfall würde so zu einer verzerrten Schätzung der Armutsgefährdungsquote führen.

Um Verzerrungen aufgrund von Antwortausfällen zu kompensieren, werden die Gewichte der aus der Erhebung fallenden Einheiten auf die verbleibenden Haushalte mit möglichst ähnlichen Merkmalen aufgeteilt. Im Fall der Erstbefragung werden die Design-Gewichte der ausgewählten Haushalte auf die erfolgreich befragten Haushalte aufgeteilt. Die dafür notwendigen Nonresponse-Gewichte ergeben sich als Inverse der Antwortwahrscheinlichkeiten r_b , der in der Erhebung verbleibenden Haushalte. Da die exakten Antwortwahrscheinlichkeiten für jeden Haushalt nicht bekannt sind, müssen sie geschätzt werden. In Übereinstimmung mit den Eurostat-Richtlinien wurden die Antwortwahrscheinlichkeiten, abhängig von bestimmten Eigenschaften der Befragten, mit Hilfe eines logistischen Regressionsmodells bestimmt. Abhängige Variable in diesem Modell ist das Antwortverhalten *Resp*, welches die Werte 0 (Antwortverweigerung) bzw. 1 (Teilnahme) trägt. Unabhängige Variablen X_j sind Merkmale, die in Zusammenhang mit dem Antwortverhalten stehen.

Für die Erstbefragungsrotation 2008 (R4/08) betrug die Ausschöpfung 64,4%.¹⁷⁾ Um die Selektivität der Erhebungsteilnahme zu untersuchen, stehen bei der Erstbefragung nur Informationen aus der Stichprobengrundlage (ZMR) zur Verfügung, wobei Unterschiede vor allem nach regionalen Merkmalen wie Bundesland und Urbanisierungsgrad beobachtet werden können.

$$\hat{r}_b = P(Resp = 1 | X_j) = \frac{\exp(\hat{\beta}_0 + \hat{\beta}_1 X_1 + \dots + \hat{\beta}_J X_J)}{1 + \exp(\hat{\beta}_0 + \hat{\beta}_1 X_1 + \dots + \hat{\beta}_J X_J)} \quad (3)$$

Bezug auf das Erstbefragungsjahr der jeweiligen Rotation durchgeführt werden. Die Weiterverfolgung dieser Haushalte ist sehr aufwendig und die erhobenen Daten werden nicht für die Berechnung des in der EU-Verordnung vorgesehenen Längsschnittindikators „Dauerhafte Armutsgefährdung“ benötigt.

¹⁷⁾ Unauffindbare Adressen wurden wie Antwortverweigerungen behandelt.

Im logistischen Regressionsmodell zur Schätzung der Antwortwahrscheinlichkeiten in der Erstbefragungsrotation R4/08 wurden nur Haushaltseigenschaften im Modell belassen, die einen signifikant von Null verschiedenen Einfluss auf die Antwortwahrscheinlichkeit haben. Die Modellspezifikation erfolgt automatisch mit Hilfe eines schrittweisen rückwärts gerichteten Ausschlusses nicht signifikanter erklärender Variablenausprägungen ($\alpha=0,1$).¹⁸⁾

Tabelle 1 zeigt die Koeffizienten des logistischen Regressionsmodells, das für die Schätzung der Antwortwahrscheinlichkeit verwendet wurde. Bei der Interpretation der Koeffizienten ist allerdings Vorsicht geboten: Aufgrund des stufenweisen Einschusses von erklärenden Variablenausprägungen ist die Referenzkategorie nicht durch inhaltliche, sondern durch technische Kriterien (Signifikanz) festgelegt. Die ursprünglich kategorialen erklärenden Variablen wurden in binär kodierte Variablen umgewandelt. Dabei mussten für eine Variable mit k Kategorien ($k-1$) Dummy-Variablen erstellt werden. In der Erstbefragung können die verwendeten unabhängigen Variablen unterschiedliches Antwortverhalten nur in sehr beschränktem Ausmaß erklären. Die Anpassungsgüte (Pseudo R^2) beträgt knapp 2%. Der geringe Anteil der durch das Modell erklärten Varianz deutet darauf hin, dass Ausfälle weitgehend zufällig auftreten oder, dass die in Zusammenhang mit dem Ausfall stehenden Merkmale nicht beobachtet wurden.¹⁹⁾ Wichtig sind hier vor allem die geschätzten Antwortwahrscheinlichkeiten $\hat{\pi}_j$.

¹⁸⁾ Bei Variablen mit vielen Ausprägungen (z.B. Bundesland), stehen meist nur einzelne Kategorien besonders hervor, während sich die Ausfallwahrscheinlichkeiten der anderen Kategorien kaum voneinander unterscheiden. Die automatisierte Auswahl von Variablen würde somit von der Zahl der Ausprägungen einer Variablen beeinflusst. Für die trennscharfe Abgrenzung von Gruppen mit auffällig erhöhter bzw. verringerter Teilnahmebereitschaft werden daher die einzelnen Ausprägungen separat auf Signifikanz geprüft. Dieses Vorgehen erschwert jedoch die inhaltliche Interpretierbarkeit der resultierenden Koeffizienten.

¹⁹⁾ Frühere Studien (z.B. *Rendtel 1995*) zeigten, dass gerade bei Panelerhebungen Merkmale, die in Zusammenhang mit der Erhebungsperson bzw. der Feldsituation stehen, entscheidend für die Antwortbereitschaft sein können. Beispielsweise können z.B. die Fluktuation oder individuelle Kapazitätsengpässe der Erhebungspersonen eine wichtige Rolle spielen. Derartige Phänomene treten in allen Erhebungen auf und verringern die Ausschöpfung und wirtschaftliche Effizienz einer Erhebung. Mit der Übernahme der Erhebung durch die Erhebungsinfrastruktur der STATISTIK AUSTRIA können Auslastung und die Motivation durch die laufende Betreuung und Schulung der Erhebungspersonen und die Ausschöpfungsqualität aber besser kontrolliert werden als bei externen Erhebungsinstituten.

Insgesamt wurden 64,4% des Bruttosamples (d.h. alle gezogenen Haushalte bereinigt um neutrale Ausfälle), erfolgreich befragt. Die erste Spalte in *Tabelle 1* zeigt die (bivariate) Antwortrate für bestimmte Gruppen. Besonders gering war die Teilnahmebereitschaft zum Beispiel in der Steiermark mit unter 57%. Besonders hoch war die Antwortrate hingegen bei Haushalten in Zweifamilien- oder Reihenhäusern, wo fast 70% aller ausgewählten Haushalte auch erfolgreich befragt werden konnten. Die Zahl der Beobachtungen ist aber im Hinblick auf die Zahl der möglichen Merkmalskombinationen (z.B. Zweifamilienhäuser in der Steiermark) zu klein, um die Selektivität verlässlich direkt aus den Daten ablesen zu können. Durch die multivariate Modellierung im logistischen Regressionsmodell kann die Teilnahmebereitschaft für jede Kombination von Merkmalen geschätzt werden. Koeffizienten mit einem negativen Vorzeichen zeigen, dass das jeweilige Merkmal - ceteris paribus - die Antwortbereitschaft verringert, während positive Koeffizienten auf eine erhöhte Antwortwahrscheinlichkeit schließen lassen. Anschaulicher ist der Effektkoeffizient. Er zeigt beispielsweise für die Variable Steiermark, dass die verhältnismäßige Chance (*engl. „odds“*) an der Befragung teilzunehmen, in der Steiermark im Mittel um den Faktor 0,642 geringer ist als in den übrigen Bundesländern.²⁰⁾

Die Konstante des Modells repräsentiert die Antwortwahrscheinlichkeit einer Referenzkategorie. Sie ist definiert als jene Gruppe, auf die keine der im Modell verwendeten Erklärungsvariablen zutrifft. Für die Referenzgruppe sind die Werte der binären Variablen X_i bis X_j ($j=7$) aus Formel (3) daher alle gleich Null.²¹⁾ Die Antwortwahrscheinlichkeit für diese Gruppe kann nun unter Verwendung von Formel (3) geschätzt werden. Es gilt $X_1, \dots, X_7=0$.

²⁰⁾ Die Odds-Ratio für die Steiermark ergibt sich aus der Wahrscheinlichkeit für eine Teilnahme an der Befragung, dividiert durch die Wahrscheinlichkeit gegen die Teilnahme an der Befragung, jeweils unter der Bedingung, dass die dafür verwendeten Haushalte aus der Steiermark stammen: $Odds(Resp_{1,0}) = \frac{P(Resp = 1 | Strmk. (X_1 = 1))}{P(Resp = 0 | Strmk. (X_1 = 1))}$

²¹⁾ Beim Modell in *Tabelle 1* sind das jene Haushalte die sich nicht in der Steiermark befinden, zu einem Gebiet mit geringer Besiedlungsdichte gehören, deren Gebäudeart kein Zweifamilien- oder Reihenhäuser ist, worin Kinder wohnen, die jüngste Person im Haushalt nicht zwischen 55 und 64 Jahren alt ist und die älteste Person jünger als 25 bzw. älter als 34 Jahre ist.

Signifikante Variablen ($\alpha=0,1$) zur Schätzung der Antwortwahrscheinlichkeiten (Erstbefragung 2008)

Tabelle 1



Parameter	Antwortrate in % (empirisch)	Koeffizient β	Standardfehler	Chi ² Wald	Effektkoeffizient Exp(β)
Konstante	(64,4)	0,894	0,106	706,2 **	-
X_1 Steiermark	56,6	-0,443	0,111	159,4 **	0,642
X_2 Dicht besiedeltes Gebiet	61,7	-0,296	0,093	100,7 **	0,744
X_3 Gebiet mit mittlerer Besiedlungsdichte	64,4	-0,193	0,107	32,8 *	0,824
X_4 Gebäudeart: Zweifamilienhaus oder Reihenhäuser	69,7	0,219	0,124	31,4 *	1,245
X_5 Anzahl der Kinder: 0	63,3	-0,186	0,098	36,0 *	0,830
X_6 Jüngste Person im Haushalt: 55-64 Jahre	67,4	0,221	0,133	27,5 *	1,247
X_7 Älteste Person im Haushalt: 25-34 Jahre	69,2	0,284	0,126	50,5 **	1,328

Q: EU-SILC 2008. - * Signifikanzniveau $\alpha=0,1$. - ** Signifikanzniveau $\alpha=0,05$.

$$P(\text{Resp} = 1 | X_1 = 0; \dots; X_7 = 0) = \frac{\exp(\widehat{\beta}_0 + \widehat{\beta}_1 \cdot 0 + \dots + \widehat{\beta}_7 \cdot 0)}{1 + \exp(\widehat{\beta}_0 + \widehat{\beta}_1 \cdot 0 + \dots + \widehat{\beta}_7 \cdot 0)} = \frac{\exp(0,894)}{1 + \exp(0,894)} = 0,710 \quad (4)$$

Die geschätzte Antwortwahrscheinlichkeit beträgt nun für jene Gruppe, bei der alle erklärenden Dummy-Variablen gleich Null sind, 71,0%. Zum Vergleich lässt sich zum Beispiel die geschätzte Antwortquote in der Steiermark analog berechnen. In diesem Fall nimmt X_7 den Wert 1 an.

$$P(\text{Resp} = 1 | X_1 = 1; X_2 = 0; \dots; X_7 = 0) = \frac{\exp(\widehat{\beta}_0 + \widehat{\beta}_1 \cdot 1 + \widehat{\beta}_2 \cdot 0 + \dots + \widehat{\beta}_7 \cdot 0)}{1 + \exp(\widehat{\beta}_0 + \widehat{\beta}_1 \cdot 1 + \widehat{\beta}_2 \cdot 0 + \dots + \widehat{\beta}_7 \cdot 0)} = \frac{\exp(0,894 - 0,443)}{1 + \exp(0,894 - 0,443)} = 0,611 \quad (5)$$

Mit 61,1% ist die geschätzte Antwortwahrscheinlichkeit für Haushalte mit denselben Merkmalen wie in der Referenzgruppe (nur eben in der Steiermark) aufgrund des negativen Koeffizienten $\widehat{\beta}_1$ tatsächlich deutlich geringer. Zur Schätzung der für die Gewichtung verwendeten Antwortwahrscheinlichkeiten werden die Werte 0 bzw. 1 für alle möglichen Kombinationen von Merkmalen in Formel (3) eingesetzt. Schlussendlich erhält jeder Haushalt eine entsprechende Schätzung, mit welcher Wahrscheinlichkeit \widehat{r}_b dieser aufgrund der Zugehörigkeit zu einer bestimmten Gruppe an der Befragung teilnimmt. Bei der Erstbefragung für EU-SILC 2008 gibt es insgesamt 56 Gruppen, bei denen \widehat{r}_b Werte zwischen 0,492 und 0,802 annimmt.

Die Nonresponse-Gewichte werden als Inverse der geschätzten Antwortwahrscheinlichkeit berechnet. Die um die Nonresponse angepassten Design-Gewichte b_b sind das Produkt aus Design-Gewicht und Nonresponse-Gewicht. Dabei erhält jeder der $n^{(i)}$ an der Befragung teilnehmenden Haushalte ein Nonresponse-Gewicht.

$$b_b = d_s \cdot \frac{1}{\widehat{r}_b} \quad b \in \{1, \dots, n^{(i)}\}, b \in s \quad (6)$$

Schritt 3: Anpassungs-Gewichte für erstmalig befragte Haushalte

Um die Genauigkeit der Schätzer, welche mit Hilfe der EU-SILC-Erhebung berechnet werden, zu erhöhen, ist es erforderlich, die Verteilungen wichtiger soziodemographischer Variablen wie Alter oder Geschlecht an verlässliche externe Datenquellen anzupassen. Praktisch geschieht die Kalibrierung der Gewichte durch iterative Anpassung an diese Randverteilungen. EU-SILC in Österreich verwendet dafür „CALMAR“,²²⁾ ein vom französischen Statistikamt INSEE entwickeltes Makro für das Programm SAS.

²²⁾ „CALage sur MARges“, vgl. http://www.insee.fr/fr/methodes/default.asp?page=outils/calmar/accueil_calmar.htm.

Die Haupt-Datenquelle für die Kalibrierung ist der Mikrozensus.²³⁾ Diese Erhebung wird in jedem Quartal bei mehr als 22.000 zufällig ausgewählten Haushalten durchgeführt, beinhaltet die Arbeitskräfteerhebung und ist so die wichtigste Basis für arbeitsmarktspezifische Daten.²⁴⁾ Aufgrund der relativ großen Stichprobe und der Auskunftspflicht für die Befragten liefert sie auch verlässliche Daten für allgemeine demographische Variablen wie Alter, Geschlecht, Nationalität oder Zugehörigkeit zu Haushalten bestimmter Größe. Der Schwerpunkt der EU-SILC-Erhebung liegt auf der Erfassung des Einkommens und seiner Komponenten. Deshalb werden zusätzlich zum Mikrozensus auch Daten vom Hauptverband der österreichischen Sozialversicherungsträger für die Kalibrierung verwendet. So soll sichergestellt werden, dass die Anzahl und der Entwicklungstrend bei Leistungsbeziehenden von Arbeitslosengeld bzw. Notstandshilfe durch die Gewichtung korrekt wiedergegeben wird.

Die Kalibrierung kann grundsätzlich für Haushalts- und Personenmerkmale durchgeführt werden. Bei EU-SILC 2008 wurde das Verfahren auf Haushaltsmerkmale und auf Haushaltsebene aggregierte Personenmerkmale angewendet. Folglich haben alle Personen im Haushalt dasselbe Gewicht und es besteht kein Unterschied zwischen Haushalts- und Personengewichten. Folgende Variablen wurden an externe Quellen angepasst:

Haushaltsebene

- Haushaltsgröße (vier Kategorien: Haushalte mit ein, zwei, drei, oder vier und mehr Haushaltsmitgliedern)
- Rechtsverhältnis an der Wohnung (zwei Kategorien: Mietwohnung/-haus oder Eigentum)
- Region (neun Kategorien: Bundesländer)

Personenebene

- Anzahl der Männer und Frauen in vier Altersgruppen
- Anzahl ausländischer Staatsbürger/-innen (ab 16 Jahren)
- Anzahl der Beziehenden von Arbeitslosengeld bzw. Notstandshilfe für eine Dauer von mehr als einem Monat

Mit Hilfe des SAS-Makros CALMAR wird der Anpassungsfaktor g_b berechnet. In Verbindung mit dem für die Nonresponse angepassten Design-Gewicht ergibt dieser Faktor das Haushalts-Gewicht $W_b^{(1)}$ der Erstbefragung:²⁵⁾

$$W_b^{(1)} = g_b \times b_b \quad (7)$$

Mit Hilfe dieser Anpassung der Gewichte kann die Kohärenz der EU-SILC-Daten mit anderen Bevölkerungsstatistiken gewährleistet werden.²⁶⁾

²³⁾ Bei EU-SILC 2008 wurde der Jahresdurchschnitt des Mikrozensus 2008 als Quelle für die in der Anpassung benötigten Randverteilungen verwendet.

²⁴⁾ Vgl. *STATISTIK AUSTRIA 2009a*, S. 31.

²⁵⁾ Als Distanzfunktion wurde die Logit-Methode gewählt. Die obere Grenze für Extremwerte wurde mit 0,5, die untere Grenze mit 2 festgelegt.

²⁶⁾ Detaillierte Kohärenzprüfungen sind z.B. in *STATISTIK AUSTRIA 2010*, S. 21 ff. dargestellt.

Gewichtungsschritte bei Folgebefragungen

Die EU-SILC-Erhebung erstreckt sich für jede Panel-Rotation über vier Jahre. Im ersten Jahr werden die Gewichtungsschritte für Design-Gewicht, Nonresponse-Gewicht und Anpassungs-Gewicht der Reihe nach durchgeführt. Ab dem zweiten Jahr der Erhebung werden Gewichte aus dem Vorjahr weiterverwendet und auf Personenebene adaptiert.

Schritt 1: Basis-Gewichte für mehrfach befragte Stichprobenpersonen

Das „Rückgrat“ des EU-SILC-Gewichtungsprozesses sind die sogenannten „Basis-Gewichte“ mit der Eurostat-Bezeichnung *RB060*. Diese berücksichtigen die Veränderung der Panel-Rotationen über die Zeit. Für das Ersterhebungsjahr sind sie identisch mit den kalibrierten Haushaltsgewichten. Jede in einem privaten Haushalt *h* lebende Person *i* hat im ersten Jahr der Erhebung das gleiche Basis-Gewicht.

$$RB060_{h_i}^{(1)} = W_b^{(1)} \quad h \in \{1, \dots, n^{(1)}\}^{27}, i \in \{1, \dots, m\}^{28} \quad (8)$$

Ab dem zweiten Erhebungsjahr werden natürliche Veränderungen der Bevölkerung sowie Antwortverweigerungen im Zuge dieser Folgebefragungen in den Gewichtungsprozess einbezogen.

Schritt 2: Nonresponse-Gewichte für mehrfach befragte Stichprobenpersonen

Die Anpassung an die Nonresponse wird dabei wie bei der Erstbefragung mit Hilfe der Schätzung der Antwortwahrscheinlichkeiten durch ein logistisches Regressionsmodell für jede Folgebefragungsrotation erstellt. Die Ausschöpfung im Jahr 2008 betrug bei den Folgebefragungsrotationen 80,3% (R1/05), 78,8% (R2/06) und 78,2% (R3/07). Im Unterschied zur Erstbefragung stehen bei den Folgebefragungen aufgrund des Längsschnitts ungleich mehr Informationen zum Antwortverhalten zur Verfügung. Das logistische Regressionsmodell aus Formel (3) lieferte eine Schätzung der Antwortwahrscheinlichkeit für jeden Haushalt. In der Folgebefragung können Stichprobenpersonen sogenannte Split-Haushalte gründen, oder in andere Haushalte einziehen. Daher müssen diesmal die Bleibewahrscheinlichkeiten für Personen berechnet werden. Formel (9) zeigt, wie die Basis-Gewichte der nicht mehr an der Erhebung teilnehmenden Personen auf die im Panel bleibenden Personen *p* aufgeteilt werden. Dabei werden zum Zeitpunkt *t* die Vorjahresgewichte $RB060_p^{(t-1)}$ mit der inversen geschätzten Antwortwahrscheinlichkeit $1/\hat{r}_p^{(t)}$ multipliziert.

²⁷⁾ Stichprobengröße $n^{(t)}$... Anzahl der Haushalte in der Erstbefragung ($t=1$) und den Folgebefragungen ($t>1$).

²⁸⁾ m ... Anzahl der Personen im Haushalt *h*.

$$RB060_p^{(t)} = \frac{RB060_p^{(t-1)}}{\hat{r}_p^{(t)}} \quad t \in \{2,3,4\}, p \in \{1, \dots, P^{(t)}\}^{29} \quad (9)$$

Für die erste Folgebefragung ($t=2$) gilt:

$$RB060_p^{(t-1)} = RB060_{h_i}^{(1)} \quad (10)$$

Ab dem zweiten Erhebungsjahr variiert das Basis-Gewicht von Person zu Person. Die Erklärungskraft der Modelle ist bei den Folgebefragungen etwas höher als bei der Erstbefragung. Das (angepasste) Bestimmtheitsmaß R^2 beträgt 8,6% (R1/05), 13,8% (R2/06) und 16,8% (R3/07).

Übersicht 1 zeigt am Beispiel von Rotation R3/07 die im logistischen Regressionsmodell aufgenommenen erklärenden Variablen. Aus diesen Variablen wurden insgesamt 43 Prädiktoren gebildet (wobei jede Variablenkategorie als Dummy-Variable im Modell verwendet wurde).

Signifikante Variablen ($\alpha=0,1$) zur Schätzung der Antwortwahrscheinlichkeiten

Übersicht 1

Interviewmerkmale

- Anzahl der Kontaktversuche
- Dauer des Haushaltsinterviews
- Person zum Interviewzeitpunkt anwesend

Ortsbezogene Variablen

- Besiedlungsdichte
- Bundesland
- Gebäudeart

Soziodemographische Personenmerkmale

- Altersgruppe
- Derzeitige Hauptaktivität
- Staatsbürgerschaft

Haushaltsbezogene Personenmerkmale

- Einzugsjahr der am längsten im Haushalt lebenden Person
- Rechtsverhältnis an der Wohnung

Einkommen und Lebensbedingungen

- Einkommensdezil (Äquivalisiertes Haushaltseinkommen)
- Auskommen mit Netto-Haushaltseinkommen
- Bei Zahlungen im Rückstand
- Armutslagen
- Unerwartete Ausgaben bis zu 900 € aus eigenen Mitteln finanzierbar
- Monatlich Gäste bewirten leistbar
- Finanzielle Belastung durch Wohnkosten
- Vorhandensein eines Farbfernsehers
- Vorhandensein eines Festnetztelefons
- Vorhandensein eines Internet-Anschlusses
- Vorhandensein eines privaten Pkws
- Feuchtigkeit, Fäulnis, Undichtheit in Haus/Wohnung
- Allgemeiner Gesundheitszustand
- Zufrieden mit dem Leben
- Zufrieden mit der Hauptbeschäftigung

²⁹⁾ $P^{(t)}$... Anzahl der Personen in der jeweiligen Folgebefragungsstichprobe.

Um allzu große Schwankungen der Nonresponse-Gewichte zu vermeiden, wurden Extremwerte auf einen vertretbaren Bereich eingeschränkt (engl. „trimming“). Bei knapp 2% der Personen in den Folgebefragungen wurden die für die Nonresponse angepassten Basis-Gewichte $RB060_p^{(t)}$ getrimmt. Formel (11) zeigt die Bedingung, nach der die Extremwertbeschränkung erfolgt ist.³⁰⁾ Jedes Basis-Gewicht des aktuellen und des Vorjahres wird in Bezug zum Mittelwert des Basis-Gewichts des aktuellen oder des Vorjahres gesetzt.

$$1/2 \leq \frac{RB060_p^{(t)} / \overline{RB060_p^{(t)}}}{RB060_p^{(t-1)} / \overline{RB060_p^{(t-1)}}} \leq 2 \quad (11)$$

Für jene Fälle, bei denen das Ergebnis aus Formel (11) außerhalb der Grenzen liegt, wurden die Basis-Gewichte an diese Grenzen angepasst.

Schritt 3: Basis-Gewichte für Personen, die nicht zur Ursprungstichprobe gehören

In der Panel-Erhebung werden auch zwei Personengruppen angetroffen, die nicht zur ursprünglichen Erstbefragungstichprobe gehört haben: Neugeborene und Personen, die nach dem Ersterhebungsjahr in einen befragten Haushalt gezogen sind. Diese Personen spielen für die ökonomische Situation von Haushalten oft eine entscheidende Rolle und müssen daher in die Erhebung einbezogen werden. Sie müssen aber bei der Anpassung der Basis-Gewichte gesondert behandelt werden.

Nach den Richtlinien von Eurostat bekommen alle seit Beginn der ersten Panel-Rotation Neugeborenen das Basis-Gewicht ihrer Mutter. Für EU-SILC 2008 betrifft dies alle zwischen 2005 und 2008 auf die Welt gekommenen Kinder in privaten Haushalten, die Teil der Erhebung sind. Sollte die Mutter nicht im gemeinsamen Haushalt leben, erhalten die Neugeborenen das mittlere Basis-Gewicht der anderen Haushaltsmitglieder. Die Einschusswahrscheinlichkeit von Personen, die in Stichprobenhaushalte eingezogen sind (engl. „co-residents“), ist nicht bekannt. Ihre Basis-Gewichte werden daher mit dem Wert Null festgesetzt.

Schritt 4: Gewichte-Teilung für alle Mitglieder des Haushalts

Die Haushaltsgewichte $w_b^{(t)}$ in den Folgebefragungen ergeben sich als die über den Haushalt gemittelten Basis-Gewichte:³¹⁾

³⁰⁾ Eine ähnliche Bedingung wurde von Eurostat empfohlen (Eurostat 2010, S. 39 f.). Aufgrund der empirischen Verteilung der Nonresponse-Gewichte konnte jedoch ein etwas strengeres Kriterium gewählt werden. Dadurch, dass bei jedem Gewichtungsschritt höchstens eine Verdoppelung bzw. Halbierung des ursprünglichen Gewichtungsfaktors erfolgt, kann die Streuung der resultierenden Gewichte gedämpft und die Sensitivität der Indikatoren gegenüber dem Gewichtungsverfahren minimiert werden.

³¹⁾ Der mit dem Kleinbuchstaben w_b bezeichnete Faktor ist nicht für die aktuellen Randverteilungen kalibriert und unterscheidet sich dadurch von dem in Großbuchstaben bezeichneten kalibrierten Querschnitt-Gewicht W_b (siehe Formel 7).

$$w_b^{(t)} = \frac{1}{m} \sum_{i=1}^m RB060_i^{(t)} \quad t \in \{2,3,4\} \\ i \in \{1, \dots, m\}, \quad b \in \{1, \dots, n^{(t)}\}$$

Mit Hilfe dieser Gewichte-Teilung (engl. „weight sharing“) bekommen alle Haushaltsmitglieder einschließlich der hinzugezogenen („co-residents“) das gleiche Haushaltsgewicht. Somit haben alle Personen m in allen privaten Haushalten n , die im Panel enthalten sind, ein Gewicht größer als Null.

Schritt 5: Anpassungs-Gewichte für jede Rotation

Um Kohärenz mit verlässlichen externen Datenquellen zu ermöglichen, werden wieder für jede Rotation Anpassungs-Gewichte unter Verwendung der gleichen Variablen und externen Verteilungen wie in der Erstbefragung berechnet. Die Kalibrierung wird auf Haushaltsebene mit Haushalts- und aggregierten Personenmerkmalen durchgeführt. Nach der Kalibrierung erhält jede Person i im Haushalt b das gleiche kalibrierte Gewicht $W_{bi}^{(t)}$.

$$W_{bi}^{(t)} = W_b^{(t)} = g_b \times w_b^{(t)} \quad t \in \{2,3,4\} \quad (13)$$

Die separate Kalibrierung jeder einzelnen Rotation garantiert, dass wichtige Randverteilungen jeder Teilstichprobe mit jenen aus der Grundgesamtheit übereinstimmen. So entspricht beispielsweise die Gewichtsumme der Haushalts- bzw. Personengewichte der Anzahl der Privathaushalte bzw. der darin lebenden Personen in der Grundgesamtheit.

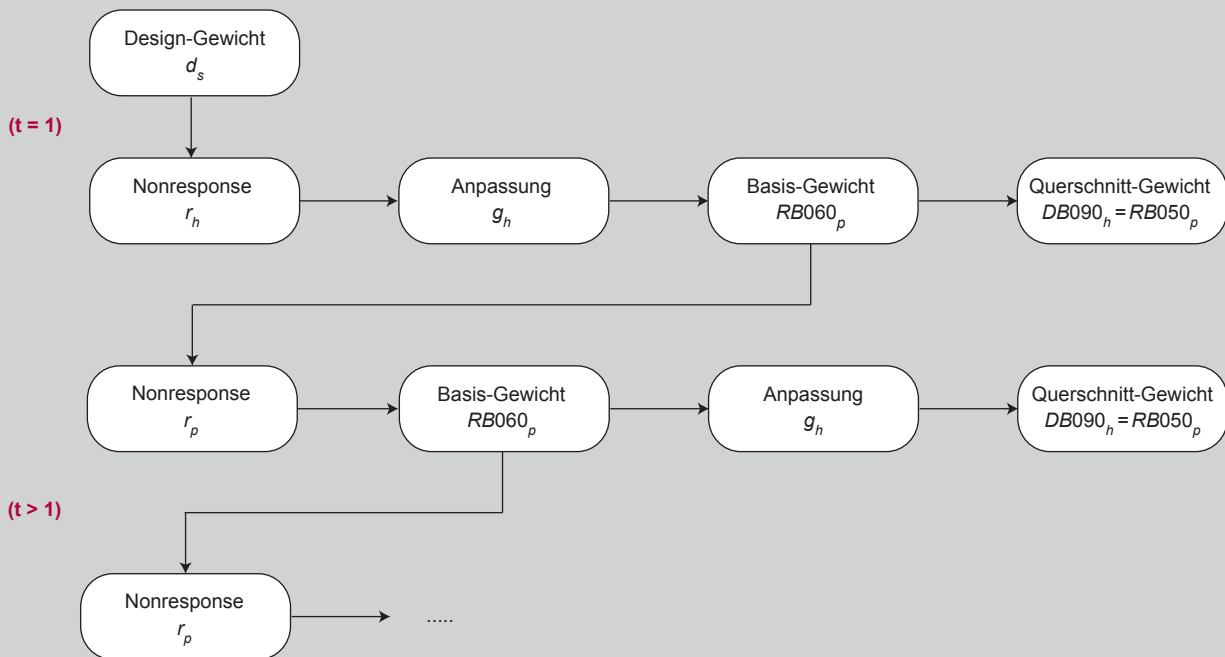
Das bisher beschriebene Gewichtungsverfahren wird auf alle vier Rotationsgruppen angewendet. Jedes Jahr tritt eine neue Rotation das erste Mal in das Panel ein und eine Rotation, die in den vorangegangenen Jahren schon viermal befragt wurde, verlässt das Panel. Grafik 2 stellt das Gewichtungsverfahren noch einmal im Zeitverlauf überblicksmäßig dar. Auf die Querschnitt-Gewichte $RB050$ und $DB090$ wird weiter unten noch genauer eingegangen. Dabei ist deutlich zu erkennen, dass sich im Ersterhebungsjahr ($t=1$) die Reihenfolge der Gewichtungsschritte von jener der Folgerhebungen ($t > 1$) unterscheidet.

Kombination von Teilstichproben

Jede Rotation repräsentiert die Grundgesamtheit. Für die Gesamtstichprobe müssen die Gewichte in geeigneter Weise kombiniert werden. Zuerst wird die Summe der Personengewichte auf die Anzahl der Personen in der jeweiligen Rotationsstichprobe skaliert (normiert).

Behandlung von Zuwanderern

Per Definition kann die Panel-Erhebung neu zugewanderte Personen nicht erfassen bzw. werden diese, falls sie in einen bestehenden Stichprobenhaushalt einziehen, mit dem Wert



Q: EU-SILC 2008

Null gewichtet. Beispielsweise können in der EU-SILC-Erhebung des Jahres 2008 jene Personen, die im Jahr 2007 zugewandert sind, nur in der Erstbefragungsstichprobe des Jahres 2008 erfasst werden. Personen, die hingegen im Jahr 2006 zugewandert sind, können immerhin in zwei Teilstichproben, nämlich der aktuellen Erstbefragung und der Rotation, die im Jahr 2007 erstmals befragt wurde, enthalten sein. Zuwanderer aus dem Jahr 2005 sind hingegen in drei von insgesamt vier Teilstichproben repräsentiert. Um dieser Verzerrung zu begegnen, werden die Gewichte der Zuwanderer, die sogenannte „IN-Population“, im Zuge der Kombination der vier Rotationen hochskaliert. Im zuletzt genannten Beispiel hätte der Skalierungsfaktor die Größe $4/3$, weil die Gruppe der 2005 zugewanderten Personen nur in der Grundgesamtheit von drei der vier Rotationen enthalten waren.

Da der Zeitpunkt der Migration nicht in der Erhebung erfasst wird, muss er mit Hilfe der im Zuge der Stichprobenziehung verwendeten Meldedaten geschätzt werden. Die Schätzung beruht auf der Anzahl an neu zugewanderten Personen in einem Stichprobenhaushalt.³²⁾ Der durchschnittliche Skalierungsfaktor wird dann auf alle Personen im Haushalt übertragen. Dieses Verfahren kann lediglich eine grobe Annäherung liefern. Gleichzeitig werden durch

³²⁾ Bei der Gewichtung nicht berücksichtigt wird, wenn Personen, aus einem Anstaltshaushalt (z.B. Studentenheim) in einen privaten-Haushalt einziehen. Die eindeutige Identifikation von Anstaltspersonen ist gegenwärtig nicht möglich.

die Mittelung über Haushalte allzu extreme Gewichtungsfaktoren vermieden.

Anpassungs-Gewichte für die Zusammenführung aller Haushalte aus vier Rotationen

Am Ende des Gewichtungsverfahrens für den Querschnitt der EU-SILC-Erhebung 2008 steht noch eine abschließende Kalibrierung der Gewichte des gesamten Querschnitts. Die Gewichte W_{b_i} sind innerhalb jedes Haushalts b gleich, die Kalibrierung wird wiederum auf Haushaltsebene für die Haushaltsgewichte W_b mit aggregierten Personenmerkmalen durchgeführt. Die dafür notwendigen Anpassungsgewichte g_b erhält man wiederum mit Hilfe des SAS-Makros CALMAR. Die für die Anpassung verwendeten externen Randverteilungen sind die gleichen wie für die zuvor beschriebenen Kalibrierungen. Für die endgültigen Querschnittsgewichte verwendet Eurostat die Bezeichnung $DB090$ auf Haushaltsebene und $RB050$ auf Personenebene, wobei alle Personen innerhalb eines Haushalts das gleiche Gewicht erhalten.

$$DB090_b = RB050_p = g_b \times W_b \tag{14}$$

Streuung der verschiedenen Gewichtungsfaktoren

Die Verteilung der jeweiligen Gewichtungsfaktoren für die Erstbefragung 2008 ist aus *Tabelle 2* abzulesen. *Spalte A* bezieht sich auf die Design-Gewichte der 3.052 für die Erstbefragung 2008 ausgewählten Adressen, also die Rotation

Gewichte EU-SILC, Haushalte der Erstbefragung (R 4/08)

Tabelle 2



Gewichtungsfaktoren	Design-Gewicht	Nonresponse-Gewicht	Anpassungs-faktor	Zwischenergebnis ¹⁾
	A	B	C	A x B x C
n	3.052	1.861	1.861	1.861
Mittelwert	1.155,4	1,6	1,1	1.916,4
Standardabweichung	207,9	0,1	0,3	567,8
Variationskoeffizient	0,18	0,09	0,23	0,30

Q: EU-SILC 2008. - 1) Wird als Zielvariable RB060 auf Personenebene gespeichert und dient den weiteren Gewichtungsschritten als Basis-Gewicht.

R4/08 aus *Grafik 1*. Im Mittel beträgt der Auswahlatz für die Erstbefragung 1/1.155. Die Gewichte variieren durch die Überrepräsentation für bestimmte Zielgruppenhaushalte und aufgrund von Rundungen bei der Festlegung der pro Schicht zu ziehenden Haushalte. Der kleinste Auswahlatz beträgt 1/1.719, während Zuwandererhaushalte mit erhöhtem Armutsgefährdungsrisiko in Wien einen Auswahlatz von 1/511 und dementsprechend das geringste Design-Gewicht unter den insgesamt 70 Schichten erhalten haben. Die Summe aller Gewichte entspricht der Anzahl der Haushalte in der Auswahlgrundlage des Zentralen Melderegisters.³³⁾ Um den für die verschiedenen Gewichtungsstufen vergleichbaren Variationskoeffizienten als Maß für die Streuung der Gewichte zu erhalten, wird die Standardabweichung der Gewichte durch den Mittelwert dividiert. Für die Design-Gewichte liegt der Variationskoeffizient bei 0,18.

Die zweite *Spalte (B)* zeigt die Verteilung der Gewichte für selektive Ausfälle. Nur Haushalte mit einem erfolgreichen Haushaltsinterview erhalten ein Gewicht. Der Variationskoeffizient dieses Faktors ist mit 0,09 nur etwa halb so groß wie jener der Design-Gewichte. Die vergleichsweise gerin-

³³⁾ Vgl. *Formel (2)*: Die Summe der inversen Auswahlwahrscheinlichkeiten über alle Schichten ergibt die Größe der Grundgesamtheit:

$$\sum_{s \in K} \sum_{h \in s} \frac{N_s}{n_s} = \sum_{s \in K} N_s = N$$

gere Streuung der Nonresponse-Gewichte unterstreicht, dass die Selektivität des Ausfalls in der Erstbefragung in Bezug auf die beobachteten Merkmale keine besonders große Rolle spielt.

Spalte C stellt den Faktor der Änderung der Gewichte durch Anpassung an externe Daten dar. Mit einem Variationskoeffizienten von 0,23 trägt die Kalibrierung am stärksten zur Variabilität der Haushaltsgewichte in der Erstbefragung bei.

Die *vierte Spalte* in *Tabelle 2* zeigt die Haushaltsgewichte, die in der Erstbefragung als Produkt aus Design-Gewicht, Nonresponse-Gewicht und Anpassungsfaktor berechnet werden. Diese Gewichte sind im Mittel deutlich höher als die Design-Gewichte, weil sich ausschöpfungsbedingt die für die Hochrechnung verwendete Gewichtsumme auf eine kleinere Anzahl an Beobachtungen aufteilt. Auch der Variationskoeffizient hat sich auf 0,30 deutlich erhöht, wobei dafür hauptsächlich der Anpassungsfaktor aus der Kalibrierung verantwortlich ist.

Tabelle 3 zeigt die Streuung der jeweiligen Gewichtungsfaktoren für die Folgebefragungsrotationen. Im Unterschied zur Erstbefragung ist der Ausgangspunkt des Gewichtungsprozesses nicht das haushaltsbezogene Design-Gewicht, sondern das bereits für Nonresponse-Gewichte bereinigte Basis-Gewicht *RB060* der aus dem Vorjahr erfolgreich weiterverfolgten Personen für die jeweilige Rotation (*Spalte A*).

Die Nonresponse-Gewichte der aus dem Vorjahr weiterverfolgten Personen weisen bei der ersten Folgebefragung mit 0,19 einen mehr als doppelt so hohen Variationskoeffizienten auf wie bei der Erstbefragung (*Spalte B von Tabelle 3*). Zwar ist der Anteil der Antwortverweigerung bei den Folgebefragungen geringer, die Variabilität ist jedoch besser durch Respondentenmerkmale erklärbar als bei der Erstbefragung. Dies ist vor allem darauf zurückzuführen, dass durch die

Gewichte EU-SILC, Personen der Folgebefragungen

Tabelle 3



Gewichtungsfaktoren		Basis-Gewicht <i>RB060</i> ⁽¹⁾	Nonresponse-Gewicht	Zwischenergebnis nach Gewichte-Teilung ¹⁾	Anpassungs-faktor	Zwischenergebnis
		A	B	C	D	C x D
1. Folgebefragung Rotation 3/07	n	3.513	3.513	3.620	3.620	3.620
	Mittelwert	1.761,77	1,24	2.276,66	0,99	2.276,66
	Standardabweichung	553,03	0,23	834,94	0,14	1.034,32
	Variationskoeffizient	0,31	0,19	0,37	0,14	0,45
2. Folgebefragung Rotation 4/06	n	2.911	2.911	3.035	3.035	3.035
	Mittelwert	2.050,10	1,26	2.715,49	0,99	2.715,49
	Standardabweichung	963,86	0,30	1.482,17	0,20	1.628,28
	Variationskoeffizient	0,47	0,24	0,55	0,21	0,60
3. Folgebefragung Rotation 1/05	n	2.545	2.545	2.692	2.692	2.692
	Mittelwert	2.395,00	1,22	3.061,49	0,97	3.061,49
	Standardabweichung	1.154,64	0,31	1.639,97	0,31	2.316,95
	Variationskoeffizient	0,48	0,25	0,54	0,32	0,76

Q: EU-SILC 2008. - 1) Ergebnis der Spalten AxB mit anschließender Gewichte-Teilung.

vorhandene Längsschnittinformation mehr erklärende Variablen im Modell verwendet werden und so eine viel feingliedrigere Schätzung der Bleibewahrscheinlichkeiten möglich ist.³⁴⁾ Die *Spalte C von Tabelle 3* zeigt das Produkt aus dem Basis-Gewicht des Vorjahres und dem Nonresponse-Gewicht. In diesem Zwischenschritt werden alle erfassten Personen, einschließlich der Neugeborenen und der in einen Stichprobenhaushalt eingezogenen Mitbewohnenden einbezogen und eine Gewichte-Teilung auf Haushaltsebene vorgenommen. Dieser Schritt teilt jedem Haushaltsmitglied ein Gewicht größer als Null zu und reduziert die Streuung der Gewichte. Obwohl die Variationskoeffizienten in *Spalte A und B* verhältnismäßig groß sind, ist wegen der Gewichte-Teilung der Haushaltsmitglieder der Wert in *Spalte C* mit 0,37 nur ein wenig größer als beim Basis-Gewicht aus dem Vorjahr. Da die Basis-Gewichte bereits einmal (in der Erstbefragung) kalibriert wurden, variieren auch die Anpassungsfaktoren weniger stark. Das Zwischenergebnis nach erfolgter Kalibrierung ergibt für alle Haushaltsmitglieder identische Personengewichte, mit deren Hilfe eine Hochrechnung aus einer Teilstichprobe, nämlich der ersten Folgebefragungsrotation R3/07, möglich wäre.

Mit zunehmender Zahl der Folgebefragungen erhöht sich der Mittelwert der Gewichte wegen der Verkleinerung der Panel-Rotationen über die Zeit und der erneuten Nonresponse-Gewichtung. Im Verlauf der vier Jahre steigen außerdem die Variationskoeffizienten der Basis-Gewichte, der Nonresponse-Gewichte und der Anpassungsfaktoren aus der Kalibrierung. Die größten Gewichte entstammen stets der Rotation, die sich am längsten im Panel befindet.

Für die Berechnung des Querschnitt-Gewichts müssen die Erstbefragungsrotationen und die drei Folgebefragungsrotationen zusammengespielt werden. *Tabelle 4* zeigt in *Spalte A* dieses **gemeinsame Zwischenergebnis** aller Rotationen.

Da nun Fälle aus allen Rotationen vorhanden sind, ist der Mittelwert dieses vorläufigen Querschnitt-Gewichts geringer als bei den Zwischenergebnissen in den *Tabellen 2 und 3* und den Basis-Gewichten in der *Tabelle 3*. Der Variationskoeffizient ist hingegen größer als bei den Basis-Gewichten der Folgebefragungsrotationen. Dies ist auf die separat für jede Rotation erfolgte Kalibrierung zurückzuführen. Die Erhöhung der Gewichte für Zuwandererhaushalte in *Spalte B* hat auf die Streuung der endgültigen Gewichte in der *letzten Spalte von Tabelle 4* im Zuge der Kombination der vier Rotationen einen nennenswerten Einfluss. Allerdings ist dieser „IN-Populationsfaktor“ äußerst rechtsschief. Für rund 95,7% der Personen ist dieser Faktor gleich dem Minimum

³⁴⁾ Das für die Nonresponse-Gewichtung verwendete logistische Regressionsmodell hat mit 14% ein höheres Pseudo-R² als bei der Erstbefragung (Pseudo-R²=9%) und kann die Antwortwahrscheinlichkeit pro Personengruppe besser vorhersagen.

Kombination der Rotationen für das Querschnittsgewicht RB050

Tabelle 4



Gewichtungsfaktoren	Personen im Querschnitt			
	Zwischenergebnisse aller Rotationen	IN-Populationsfaktor	Anpassungsfaktor	Endgültiges Querschnittsgewicht RB050
	A	B	C	A x B x C
n	13.631	13.631	13.631	13.631
Mittelwert	604,6	1,03	0,98	604,6
Standardabweichung	321,5	0,20	0,03	319,6
Variationskoeffizient	0,53	0,20	0,03	0,53

Q: EU-SILC 2008

1, das Maximum beträgt 4. Da die Rotationen bereits separat kalibriert wurden, liefert der Anpassungsfaktor (*Spalte C*) bei der abschließenden, gemeinsamen Kalibrierung nur mehr einen kleinen Beitrag zur Streuung des endgültigen Querschnitt-Gewichts RB050. Insgesamt liegt der Variationskoeffizient dieses für die Hochrechnung von EU-SILC-Querschnittergebnissen verwendeten Gewichtungsfaktors bei 0,53. Die Streuung der EU-SILC-Gewichte ist vergleichbar zur Streuung der Mikrozensus-Gewichte, für die im 2. Quartal 2008 der Mittelwert bei 169 und der Variationskoeffizient bei 0,57 liegen.

Das Querschnitt-Gewicht RB050 dient zur unverzerrten deskriptiven Analyse der EU-SILC Daten und ist in den unentgeltlich zu beziehenden EU-SILC-Nutzerdaten enthalten.³⁵⁾ Bei der Verwendung von gewichteten Daten ist zu beachten, dass Standardverfahren zur Berechnung der Fehlervarianzen bei einfachen Zufallsstichproben in der Regel nicht anwendbar sind.³⁶⁾

Quellen

Cochran, W.G. (1977): "Sampling Techniques". John Wiley & Sons, Inc. New York.

Eurostat (2010): "Description Of Target Variables: Cross-sectional and Longitudinal". EU-SILC document 065 (2008 Operation). Abrufbar unter: http://circa.europa.eu/Public/irc/dsis/eusilc/library?l=/guidelines_questionnaire/operation_guidelines/silc065_version/_EN_1.0_&ca=d (5.5.2010)

Groves, R.M. / Fowler, F.J., Jr. / Couper, M.P. / Lepkowski, J.M. / Singer, E. / Tourangeau, R. (2004): "Survey Methodology". John Wiley & Sons, Inc. New Jersey.

³⁵⁾ In den Nutzerdaten wird RB050 als *hgew* bezeichnet.

³⁶⁾ Gewichte spielen eine entscheidende Rolle für den sogenannten Designeffekt. Dieser gibt an, um welchen Faktor das jeweilige Design, d.h. insbesondere Mehrstufigkeit bzw. Schichtung sowie Allokation und Gewichtung, die Varianz im Verhältnis zur einfachen Zufallsstichprobe erhöht bzw. verringert. Der Designeffekt ist für jedes Merkmal unterschiedlich. Für die Armutsgefährdungsquote wurde der Designeffekt näherungsweise mit 1,33 geschätzt (vgl. STATISTIK AUSTRIA 2010, S. 16).

Kish, L. (1990): "Weighting: Why, When and How?". In: American Statistical Association. 1990 Proceedings of the Survey Research Methods. Alexandria, VA. 121-130. Abrufbar unter: http://www.amstat.org/sections/srms/Proceedings/papers/1990_018.pdf. (5.5.2010)

Osier, G. / Museux, J.M. / Seoane, P. / Verma, V. (2006): "Cross-sectional and longitudinal weighting for the EU-SILC rotational design". Eurostat/University of Siena. Luxembourg/Siena. Abrufbar unter: <http://www.iser.essex.ac.uk/files/survey/ulsc/methodological-research/mols-2006/scientific-social-programme/papers/Osier.pdf>. (5.5.2010)

Rendtel, U. (1995): „Panelmortalität und Panelrepräsentativität“. Frankfurt: Campus.

Rösch, G. (1994): „Kriterien der Gewichtung einer nationalen Bevölkerungsstichprobe“. In: Gabler, S. / Hoffmeyer-Zlotnik, J.H.P. / Krebs, D.: Gewichtung in der Umfragepraxis. Opladen: Westdeutscher Verlag GmbH. 7-26.

STATISTIK AUSTRIA (2009a). „Arbeitskräfteerhebung: Ergebnisse des Mikrozensus 2008“. Wien. Abrufbar unter: www.statistik.at/dynamic/wcmsprod/idcplg?IdcService=GET_NATIVE_FILE&dID=57149&dDocName=041730 (5.5.2010)

STATISTIK AUSTRIA (2009b): „Einkommen, Armut und Lebensbedingungen: Ergebnisse aus EU-SILC 2008 in Österreich“. Wien. Abrufbar unter: http://www.statistik.at/web_de/static/eu-silc_2008eingliederungsindikatoren__armutsgefaehrdung_in_oesterreich_bd_043350.pdf (05.05.2010)

STATISTIK AUSTRIA (2010): „Methoden und Vergleiche zu EU-SILC“. Abrufbar unter: http://www.statistik.at/web_de/static/methoden_und_vergleiche_zu_einkommen__armut_und_lebensbedingungen_eu_silc_2_044097.pdf (5.5.2010)

Till, M. (2001): „Stichprobe und Gewichtung des Europäischen Haushaltspanels in Österreich“. Diplomarbeit. Universität Wien. Wien.

Summary

In Austria EU-SILC (Community Statistics on Income and Living Conditions) was carried out for the first time in the year 2003 as a purely cross sectional sample based on a contractual agreement. Since then it has been used for regular reporting on income, poverty and living conditions. EU-SILC provides the main data source for indicators on social inclusion such as the at-risk-of-poverty rate. In the year 2004 EU-SILC was implemented according to a binding EU-regulation. A rotational panel was started in order to fulfill the requirements of the longitudinal and the cross-sectional component in one integrated design. With the EU-SILC operation 2007 the rotational Panel with four subsamples was fully implemented. Every year one rotational group which has been traced over a period of four years is replaced by a new (cross-sectional) rotation of private households. This design implies a complex weighting procedure, in which different selection probabilities and selective non-response patterns have to be compensated. Finally weights are calibrated to reduce bias and ensure that the results match known marginal distributions from the Labour Force Survey or register data. The article outlines the procedures recommended by Eurostat, explains the necessary steps to obtain weights for one survey year and presents the dispersion of weights derived in each step.