

Dipl.-Sozialwissenschaftlerin Anja Afentakis, Dipl.-Mathematiker Wolf Bihler

# Das Hochrechnungsverfahren beim unterjährigen Mikrozensus ab 2005

*Ab 2005 wird der Mikrozensus als kontinuierliche Erhebung durchgeführt, um künftig Quartals- und Jahresdurchschnittsergebnisse veröffentlichen zu können. Durch diesen Umstieg – von einer festen auf eine gleitende Berichtswoche – musste das Hochrechnungsverfahren völlig neu konzipiert werden.*

*Auch in Zukunft erfolgt die Hochrechnung der Mikrozensusergebnisse in zwei Stufen: Zunächst wird anhand von Informationen über die Haushalte, die nicht geantwortet haben, ein Kompensationsfaktor berechnet. In einer zweiten Stufe werden die mit dem Kompensationsfaktor gewichteten Stichprobenverteilungen ausgewählter Hilfsvariablen an Eckwerte aus der laufenden Bevölkerungsfortschreibung und dem Ausländerzentralregister angepasst. Hierbei wurde der Hochrechnungsrahmen zur Verbesserung der Ergebnisqualität um Altersklassen und ausgewählte Staatsangehörigkeiten erweitert. Weil zum Zeitpunkt der Hochrechnung keine aktuellen Eckwerte zur Verfügung stehen, müssen diese geschätzt werden.*

*Das neue quartalsweise Hochrechnungsverfahren kann nur noch auf einem Viertel des bisherigen Stichprobenumfangs basieren. Um zu schwach besetzte Anpassungsklassen zu vermeiden, werden die Kompensations- und Hochrechnungsfaktoren ab 2005 durch ein modernes Kalibrierungsverfahren (Generalized Regression Estimation) berechnet, das eine Anpassung an getrennte Randverteilungen ermöglicht.*

## Vorbemerkung

Bis 2004 wurde der Mikrozensus, die EU-weit größte Haushaltsstichprobe, für eine feste Berichtswoche durchgeführt. Die Angaben aller befragten Haushalte bezogen sich auf diese Berichtswoche, die normalerweise in der zweiten Aprilhälfte lag. Die EU-Verordnung über die Arbeitskräfteerhebung<sup>1)</sup> schreibt eine kontinuierliche Erhebung vor, bei der die Berichtswochen gleichmäßig über das ganze Jahr verteilt sind und die neben Jahresergebnissen auch Quartalsergebnisse liefert. Deutschland hatte bei dieser Erhebung, die in den Mikrozensus integriert ist, die Ausnahmegenehmigung, bis 2004 eine jährliche Befragung mit einer festen Berichtswoche im Frühjahr durchzuführen. Mit dem neuen Mikrozensusgesetz für die Jahre 2005 bis 2012<sup>2)</sup> wurde die rechtliche Grundlage geschaffen, den Mikrozensus als kontinuierliche Erhebung durchzuführen. Mit der kontinuierlichen Erhebung werden folgende Ziele verfolgt:

- Es sollen zeitraumbezogene Ergebnisse, nämlich Jahresdurchschnitte und Quartalsdurchschnitte, anstelle der bisherigen Ergebnisse für eine feste Berichtswoche ermittelt werden. Ferner sollen Veränderungen zwischen zwei aufeinander folgenden Quartalen analysiert werden können.
- Für Jahresdurchschnitte sollen die Stichprobenfehler nicht höher ausfallen als bisher für Ergebnisse der Be-

1) Verordnung (EG) Nr. 577/98 des Rates vom 9. März 1998 zur Durchführung einer Stichprobenerhebung über Arbeitskräfte in der Gemeinschaft (Amtsbl. der EG Nr. L 77, S. 3), zuletzt geändert durch Verordnung (EG) Nr. 2257/2003 vom 25. November 2003 (Amtsbl. der EU Nr. L 336, S. 6).

2) Gesetz zur Durchführung einer Repräsentativstatistik über die Bevölkerung und den Arbeitsmarkt sowie die Wohnsituation der Haushalte (Mikrozensusgesetz 2005 – MZG 2005) vom 24. Juni 2004 (BGBl. I S. 1350).

richtswoche. Die Präzisionsanforderungen an Quartalsergebnisse gemäß Artikel 3 Absatz 2 der EU-Verordnung<sup>3)</sup> sollen eingehalten werden.

- Die Ergebnisse sollen schneller als bisher – nämlich 90 Tage nach Quartalsende – verfügbar sein.

Während der Umstieg auf die kontinuierliche Erhebung am Auswahlverfahren nur geringfügige Änderungen erforderte, musste das Hochrechnungsverfahren völlig neu konzipiert werden. Das neue Hochrechnungsverfahren wird in diesem Aufsatz vorgestellt. In einem einleitenden Kapitel werden die Änderungen beim Auswahlverfahren beschrieben.<sup>4)</sup>

## 1 Auswahl

### 1.1 Rotationsschema

Veränderungen zwischen zwei aufeinander folgenden Quartalen können in der Regel präziser gemessen werden, wenn die Stichproben beider Quartale sich überlappen, das heißt wenn es eine gemeinsame Schnittmenge gibt. Je größer der Überlappungsgrad ist, umso präziser können die Veränderungen gemessen werden. Allerdings besteht ein Zielkonflikt zwischen der möglichst präzisen Messung von Quartalsveränderungen und der möglichst präzisen Messung von Jahresdurchschnitten. Folgendes Beispiel soll diesen Zielkonflikt verdeutlichen: Angenommen, der unterjährige Mikrozensus würde nach einem Rotationsschema mit einer 50%igen Überlappung zwischen zwei aufeinander folgenden Quartalen als auch mit einer 75%igen Überlappung zwischen zwei aufeinander folgenden Jahren durchgeführt. Ein Stichprobenhaushalt würde dann vier Jahre lang in jeweils zwei aufeinander folgenden Quartalen am Mikrozensus teilnehmen. Bei diesem Rotationsschema reduzierte sich für ein Merkmal mit einem Korrelationskoeffizienten von 0,7 zwischen zwei aufeinander folgenden Quartalen der Zufallsfehler für die Differenz um 19% gegenüber überschneidungsfreien Quartalsstichproben gleichen Umfangs und für ein Merkmal mit einem Korrelationskoeffizienten von 0,9 sogar um 29%. Der Zufallsfehler für Jahresdurchschnitte würde sich aber um maximal 37% erhöhen, da sich die Zahl der jährlich einbezogenen Haushalte fast halbiert.<sup>5)</sup> Da für Jahresergebnisse keine höheren Zufallsfehler als bisher hingenommen werden sollen, fiel die Entscheidung gegen eine unterjährige Rotation. Ohne unterjährige Rotation kann mit gleicher Präzision der Jahresdurchschnitte gerechnet werden wie bisher für die Ergebnisse der festen Berichtswoche.<sup>6)</sup> Der bisherige jährliche Austausch eines Viertels der Stichprobe hat sich bewährt und wird beibehalten. Ein Stichprobenhaushalt

verbleibt also wie bisher vier Jahre in der Stichprobe und wird jedes Jahr nur einmal befragt. Um die Präzisionsanforderungen der EU-Verordnung für Quartalsveränderungen unter diesem für die Messung von Quartalsveränderungen recht ungünstigen Stichprobendesign erfüllen zu können, kann allerdings ab 2005 die Arbeitskräftestichprobe nicht mehr als Unterstichprobe des Mikrozensus mit einem Auswahlatz von 0,45% durchgeführt werden. Mit dem vollen Auswahlatz von 1% können die Präzisionsanforderungen der EU-Verordnung aber voraussichtlich gerade erreicht werden. Wegen des Verzichts auf eine unterjährige Rotation kann – was das jährliche Volumen der Stichprobe betrifft – der derzeitige Auswahlplan beibehalten werden<sup>7)</sup>. Zu klären bleibt noch die Frage, wie die jährliche Stichprobe auf die 52 Wochen des Jahres verteilt wird.

### 1.2 Zeitliche Aufteilung der Stichprobe

Aus stichprobenmethodischer Sicht wäre ein Verfahren vorzuziehen, bei dem die jährliche Stichprobe nach einem Zufallsverfahren in 52 Teilstichproben zerlegt wird. Für jede Berichtswoche hätte man dann eine Zufallsstichprobe mit einem Auswahlatz von  $\frac{1}{52}\%$ . Dieses Vorgehen ist aber bei einer auf persönliche Interviews gestützten Erhebung wie dem Mikrozensus wegen der dann enorm hohen Wegezeiten und Fahrtkosten nicht realisierbar. Für einen bestimmten Zeitraum (z. B. ein Quartal oder einen Monat) müssen die Stichprobenauswahlbezirke<sup>8)</sup> in irgendeiner Weise räumlich geklumpt werden, um den Zeitaufwand und die Fahrtkosten für die Interviewerinnen und Interviewer in einem vertretbaren Rahmen zu halten. Diese Klumpung erhöht aber für viele Merkmale (Merkmale, die räumlich korreliert sind) tendenziell den Zufallsfehler für Quartalsergebnisse.

Untersucht wurden folgende zwei Möglichkeiten:

- *Aufteilung auf der Basis eines Quartalsvolumens:* Die jährliche Stichprobe wird nach einem Zufallsverfahren in vier gleich große Teile für die Quartale zerlegt. Je Quartalsstichprobe werden die Stichprobenauswahlbezirke zu so genannten *Interviewerpaketen* zusammengefasst. Ein Interviewerpaket umfasst eine Menge von Stichprobenauswahlbezirken, die eine Interviewerin bzw. ein Interviewer innerhalb eines Monats abuarbeiten hat. Die Interviewerpakete eines Quartals werden anschließend nach einem Zufallsverfahren in drei Teile für die Monate zerlegt.
- *Aufteilung auf der Basis eines Monatsvolumens:* Die jährliche Stichprobe wird nach einem Zufallsverfahren in

3) Für eine Personengruppe auf nationaler Ebene, die 5% der Bevölkerung im erwerbsfähigen Alter umfasst, darf der Standardfehler für die Differenz zwischen dem hochgerechneten Quartalswert und dem Wert des Vorquartals – bezogen auf den hochgerechneten Quartalswert – höchstens 2% betragen.

4) Zu den organisatorischen Aspekten der Umstellung des Mikrozensus auf eine kontinuierliche Erhebung siehe Lotze, S./Breiholz, H.: „Zum neuen Erhebungsdesign des Mikrozensus (Teil 1)“ in WiSta 5/2002, S. 359 ff. und dies.: „Zum neuen Erhebungsdesign des Mikrozensus (Teil 2)“ in WiSta 6/2002, S. 454 ff.

5) Der Anstieg des Zufallsfehlers fällt etwas geringer aus, wenn die Variabilität zwischen den Quartalen innerhalb eines Haushalts gegenüber der Variabilität zwischen den Haushalten nicht vernachlässigt werden kann.

6) Siehe Grohmann, H.: „Zum Übergang vom Berichtswochenkonzept zum Konzept unterjähriger Erhebungen beim Mikrozensus“, Wiesbaden 1991, S. 11 ff.

7) Zwar ist ein neuer Auswahlplan wünschenswert – der derzeitige basiert auf den Ergebnissen der Volkszählung 1987 bzw. in den neuen Bundesländern auf einer Auswertung des Bevölkerungsregisters der ehemaligen DDR –, jedoch zurzeit mangels einer geeigneten Auswahlgrundlage nicht sinnvoll. Zum derzeitigen Auswahlplan siehe Meyer, K.: „Zum Auswahlplan des Mikrozensus ab 1990“ in Gabler, S. u. a.: „Gewichtung in der Umfragepraxis“, Opladen 1994, S. 106 ff.

8) Auswahlbezirke sind die der Stichprobenziehung zugrunde liegenden Einheiten. Sie sind künstlich abgegrenzte Flächenstücke, die mehrere in der Regel benachbarte Gebäude umfassen; bei größeren Gebäuden ist nur ein Gebäude – ganz oder teilweise – enthalten. Ein Auswahlbezirk enthält durchschnittlich etwa 9 Wohnungen.

zwölf gleich große Teile für die Monate zerlegt. Je Monatsstichprobe werden die Stichprobenauswahlbezirke zu Interviewerpaketen zusammengefasst.

Da für die Bildung von Interviewerpaketen auf der Basis eines Quartalsvolumens mehr Stichprobenauswahlbezirke zur Verfügung stehen als auf der Basis eines Monatsvolumens, liegen die Bezirke eines Interviewerpaketes näher beieinander, was zu geringeren Reisekosten für die Interviewerin bzw. den Interviewer führt. Die stärkere Klumpung führt aber zu höheren Zufallsfehlern für die Quartalsergebnisse. Fehlerabschätzungen mit simuliertem Material aus sieben Bundesländern auf der Basis des Mikrozensus 1999 ergaben, dass bei der Aufteilung auf der Basis eines Quartalsvolumens die erwarteten Standardfehler für die Quartalsergebnisse zum Teil um mehr als 20% höher liegen als bei der Aufteilung auf der Basis eines Monatsvolumens<sup>9)</sup>. Daher entschied man sich trotz der höheren Kosten für die Aufteilung auf der Basis eines Monatsvolumens.

Technisch wurde die Zerlegung der jährlichen Stichprobe in zwölf Teile wie folgt realisiert: Bei der Grundausswahl des Mikrozensus<sup>10)</sup> waren für jede Mikrozensus-Stichprobe – in Voraussicht einer künftigen kontinuierlichen Erhebung – die Auswahlbezirke in jeder Schicht beginnend mit einem Zufallsstart fortlaufend in der Reihenfolge der Sortierung für die Auswahl modulo 48 nummeriert worden. Jeweils vier Nummern des resultierenden Kennzeichens wurden nun zusammengefasst und bilden ein Monatsvolumen.

Unterhalb der Monatsebene kann die Aufteilung auf die Wochen nicht nach einem Zufallsverfahren erfolgen. Der Interviewer bzw. die Interviewerin können die Aufteilung selbst vornehmen und ihre Wege minimieren, also zum Beispiel benachbarte Auswahlbezirke in derselben Woche abarbeiten. Vorgegeben ist lediglich, die Interviews insgesamt gleichmäßig auf die Wochen zu verteilen.

Die Berichtswoche eines Stichprobenhaushalts liegt nicht immer innerhalb des ausgewählten Monats. Durch verzögerte Beantwortung (Urlaub, Selbstaussfüllung) kann die Berichtswoche auch in einen späteren Monat fallen, da sie dem Haushalt nicht fest vorgegeben ist, sondern immer die Kalenderwoche vor dem Interviewzeitpunkt bzw. dem Zeitpunkt des Ausfüllens des Fragebogens ist. Bei einer dem Haushalt fest vorgegebenen Berichtswoche könnte man die Termine für Quartalsergebnisse nicht einhalten, ohne eine dramatisch höhere Non-Response-Rate, die zurzeit bei etwa 3% liegt, zu riskieren. Die Auswirkungen der Verschiebungen der Berichtswoche werden zu untersuchen sein, wenn die Daten der ersten Quartale vorliegen. Vermutlich werden sich die Verschiebungen zum großen Teil in ihren Auswirkungen kompensieren.

## 2 Methodische Grundlagen

### 2.1 Ziele

Bei der Entwicklung des Hochrechnungsverfahrens waren folgende Rahmenbedingungen zu beachten:

- *Quartalsweise Hochrechnung je Bundesland:* Da die Ergebnisse für Quartalsdurchschnitte 90 Tage nach Quartalsende vorliegen sollen, muss auch die Hochrechnung quartalsweise erfolgen. Außerdem soll die Hochrechnung – wie bisher – nicht zentral im Statistischen Bundesamt, sondern in den einzelnen Statistischen Landesämtern erfolgen.
- *Anpassung an bekannte Eckwerte der Grundgesamtheit aus der laufenden Bevölkerungsfortschreibung:* Wie bisher soll grundsätzlich eine gebundene Hochrechnung mit Anpassung an Eckwerte aus der laufenden Bevölkerungsfortschreibung durchgeführt werden. Sie hat in erster Linie den Zweck, systematische Fehler durch unerkannte Antwortausfälle zu verringern. Außerdem lassen sich die Zufallsfehler derjenigen Erhebungsmerkmale, die mit den Bezugsmerkmalen hoch korreliert sind, verringern. Es wird angenommen, dass die verwendeten Eckwerte der Bezugsmerkmale aus der laufenden Bevölkerungsfortschreibung korrekt sind und etwaige systematische Fehler in den Eckwerten so gering sind, dass sie vernachlässigt werden können. Durch die Anpassung wird auch eine gewisse Konsistenz der in der amtlichen Statistik verfügbaren Bevölkerungsdaten erreicht.
- *Gleicher Hochrechnungsfaktor für alle Personen eines Haushalts:* Bisher gab es zwei verschiedene Hochrechnungsfaktoren für Haushalts- und Personenauswertungen. Das bisherige Hochrechnungsverfahren lieferte zunächst Personenfaktoren. Der Hochrechnungsfaktor eines Haushalts war dann der Mittelwert der Personenfaktoren aller Haushaltsmitglieder. Die Verwendung zweier Faktoren führte zu Inkonsistenzen zwischen Haushalts- und Personentabellen.

Bisher erfolgte im Mikrozensus die Hochrechnung durch eine Anpassung an Eckwerte aus der laufenden Bevölkerungsfortschreibung gegliedert nach disjunkten Klassen einer Kreuzkombination regionaler Gebietseinheiten, den so genannten regionalen Anpassungsschichten<sup>11)</sup>, mit der Staatsangehörigkeit (deutsch/nicht deutsch) und dem Geschlecht.<sup>12)</sup> Das neue quartalsweise Hochrechnungsverfahren kann nur noch auf einem Viertel des bisherigen Stichprobenumfangs basieren. Das bisherige Hochrechnungsverfahren lässt sich nicht mehr realisieren, da viele Anpassungsklassen zu schwach besetzt wären. Sehr kleine Besetzungszahlen in den Anpassungszellen führen dazu, dass die Schätzung instabil wird. Neben einer Erhöhung

9) Siehe Bihler, W.: "Problems of Converting the Microcensus into a Continuous Survey" in „Rotierende Stichproben“, Band 21 der Schriftenreihe „Spektrum Bundesstatistik“, Wiesbaden 2002, S. 102 ff.

10) Bei den späteren ergänzenden Auswahlen von Neubaubezirken wird analog verfahren.

11) Regionale Anpassungsschichten sind Zusammenfassungen von Kreisen, bei ausreichender Größe auch einzelne Kreise oder bei sehr großen Städten auch Teile. Eine Anpassungsschicht umfasst in der Regel mindestens 500 000 Einwohner. Bundesweit gibt es 132 regionale Anpassungsschichten.

12) Zum bisherigen Hochrechnungsverfahren siehe Heidenreich, H.-J.: „Hochrechnung des Mikrozensus ab 1990“ in Gabler, S. u. a.: „Gewichtung in der Umfragepraxis“, Opladen 1994, S. 112 ff.

des Zufallsfehlers wäre auch die näherungsweise Erwartungstreue der Schätzung nicht mehr gewährleistet. So genannte Kalibrierungsverfahren ermöglichen es, an Randverteilungen getrennt anzupassen, sodass beispielsweise die hochgerechnete Zahl der Männer, der Frauen sowie der Deutschen und der Ausländer mit den Ergebnissen der laufenden Bevölkerungsfortschreibung übereinstimmt, die Eckwerte für die Kombinationen, etwa für die männlichen Deutschen, aber nicht notwendigerweise getroffen werden. Die Fallzahlproblematik wird wegen der in der Regel ausreichend starken Besetzung der Randverteilungen auf diese Weise vermieden. Zu den Kalibrierungsverfahren gehört auch die so genannte verallgemeinerte Regressionsschätzung<sup>13)</sup>, die gegenüber anderen Verfahren relativ einfache Schätzungen der Zufallsfehler ermöglicht.

## 2.2 Der verallgemeinerte Regressionsschätzer

Der verallgemeinerte Regressionsschätzer für den Totalwert  $t_y$  einer interessierenden Variablen  $Y$  lässt sich darstellen als

$$(1) \hat{t}_y = \hat{t}_{y,HT} + \hat{\mathbf{B}}' \cdot (\mathbf{t}_x - \hat{\mathbf{t}}_{x,HT}),$$

wobei

$$(2) \hat{t}_{y,HT} = \sum_{k=1}^n d_k y_k$$

der frei hochgerechnete Totalwert für die Variable  $Y$  ist („Horvitz-Thompson-Schätzer“) mit

$y_k$ : Wert der Variablen  $Y$  für Person  $k$ ,

$n$ : (Netto-)Stichprobenumfang,

$d_k$ : Kehrwert der Wahrscheinlichkeit für Person  $k$ , in die Stichprobe zu gelangen,

$\mathbf{t}_x$ : Vektor der bekannten Totalwerte der Ausprägungen der Hilfsvariablen („Eckwerte“),

$\hat{\mathbf{t}}_{x,HT}$ : Vektor der aus der Stichprobe frei hochgerechneten Schätzwerte für  $\mathbf{t}_x$ ,

$$(3) \hat{\mathbf{t}}_{x,HT} = \sum_{k=1}^n d_k \mathbf{x}_k$$

mit

$\mathbf{x}_k$ : Vektor der Ausprägungen der Hilfsvariablen für Person  $k$ .

$\hat{\mathbf{B}}$  ist der Vektor der geschätzten Regressionskoeffizienten:

$$(4) \hat{\mathbf{B}} = \left( \sum_{k=1}^n d_k \mathbf{x}_k \mathbf{x}_k' \right)^{-1} \left( \sum_{k=1}^n d_k \mathbf{x}_k y_k \right).$$

Es handelt sich also um eine mit  $d_k$  gewichtete multiple Regression. Wenn man beispielsweise an die Randverteilungen Geschlecht und Staatsangehörigkeit (deutsch/nicht

deutsch) anpasst, besteht der Vektor  $\mathbf{x}_k$  aus vier Komponenten,  $\mathbf{x}_k = (x_{k1}, x_{k2}, x_{k3}, x_{k4})'$ .  $x_{k1}$  ist 1, falls Person  $k$  männlich ist, und andernfalls Null.  $x_{k2}$  bis  $x_{k4}$  werden mit den Ausprägungen weiblich, deutsch und nicht deutsch analog gebildet.<sup>14)</sup> Welche Hilfsvariablen konkret gewählt wurden, wird in Abschnitt 3.2 erläutert. Im Fall des Mikrozensus wird  $d_k$  nicht als der Kehrwert der reinen Ziehungswahrscheinlichkeit aufgefasst, das heißt als der Wahrscheinlichkeit, dass die Person  $k$  für den Mikrozensus ausgewählt wird. Da der Einfluss der Antwortausfälle nicht vernachlässigt werden soll, wird noch die Wahrscheinlichkeit berücksichtigt, dass eine Person, die ausgewählt ist, auch antwortet. Die Wahrscheinlichkeit, in den Mikrozensus zu gelangen, ist also das Produkt aus der Ziehungswahrscheinlichkeit und der Antwortwahrscheinlichkeit. Damit gilt für den Kehrwert  $d_k$

$$(5) d_k = \frac{1}{\pi_k \hat{\theta}_k}$$

mit

$\pi_k$ : Ziehungswahrscheinlichkeit für Person  $k$  (= 0,25%),

$\hat{\theta}_k$ : geschätzte Antwortwahrscheinlichkeit für Person  $k$ .

Die Schätzmethode für die Antwortwahrscheinlichkeiten wird im nächsten Kapitel ausgeführt. Der Regressionsschätzer ist eine lineare Schätzfunktion und kann dargestellt werden als

$$(6) \hat{t}_y = \sum_{k=1}^n \left( 1 + (\mathbf{t}_x - \hat{\mathbf{t}}_{x,HT})' \left( \sum_{k=1}^n d_k \mathbf{x}_k \mathbf{x}_k' \right)^{-1} \mathbf{x}_k \right) d_k y_k = \sum_{k=1}^n w_k y_k.$$

Somit ergibt sich der zu berechnende Hochrechnungsfaktor  $w_k$  als

$$(7) w_k = \left( 1 + (\mathbf{t}_x - \hat{\mathbf{t}}_{x,HT})' \left( \sum_{k=1}^n d_k \mathbf{x}_k \mathbf{x}_k' \right)^{-1} \mathbf{x}_k \right) d_k.$$

Dieser Hochrechnungsfaktor wird an das Einzelmaterial angefügt, sodass wie üblich flexibel für die verschiedensten Zielvariablen ausgewertet werden kann.

Der Regressionsschätzer hat die Eigenschaft, dass die Eckwerte getroffen werden, wenn sie aus der Stichprobe hochgerechnet werden:

$$(8) \hat{\mathbf{t}}_x = \mathbf{t}_x.$$

Alle Personen eines Haushalts sollen den gleichen Hochrechnungsfaktor erhalten. Die Hilfsvariablen sind aber ausschließlich Personenmerkmale. Daher muss der Vektor  $\mathbf{x}_k$  modifiziert werden: Anstelle der Ausprägungen auf der Personenebene werden die Haushaltsmittelwerte genommen. Im obigen Beispiel wäre dann  $x_{k1}$  – die Komponente für die Anpassung an die Zahl der Männer – die durchschnittliche Zahl der Männer in dem Haushalt, zu dem Person  $k$  gehört. Der Vektor  $\mathbf{x}_k$  ist also für alle Personen eines Haushalts gleich, was zu gleichen Hochrechnungsfaktoren innerhalb

13) "Generalized Regression Estimator" (GREG) in der englischsprachigen Literatur.

14) Zu beachten ist in diesem Beispiel, dass die vier Komponenten linear abhängig sind, was dazu führt, dass die Matrix in Formel (4) nicht invertierbar ist. Daher muss eine der vier Komponenten weggelassen werden. Auch bei den Eckwerten kann auf eine Ausprägung verzichtet werden, da sich die vierte aus den übrigen drei berechnen lässt.

des Haushalts führt. Die Eingangsgewichte  $d_k$  sind ohnehin für alle Personen eines Haushalts gleich. Durch die Mittelwertbildung werden die Eckwerte weiterhin getroffen, wenn sie aus der Stichprobe hochgerechnet werden. Für Personen in Gemeinschaftsunterkünften ist kein einheitlicher Hochrechnungsfaktor erforderlich. Für diese Personen bleiben die Hilfsvariablen daher auf der Personenebene.

### 2.3 Beschränkung der Hochrechnungsfaktoren

Beim Regressionsschätzer ist es theoretisch denkbar, dass die resultierenden Hochrechnungsfaktoren negativ werden können. Die recht anschauliche Interpretation, dass der Hochrechnungsfaktor einer Stichprobenperson angibt, wie viele Personen in der Grundgesamtheit diese Person vertritt, wäre dann nicht mehr möglich. Negative Hochrechnungsfaktoren sollten daher nicht auftreten. Außerdem sollten die resultierenden Hochrechnungsfaktoren eine nicht zu starke Streuung aufweisen, da dies tendenziell den Zufallsfehler erhöht. Durch die höheren Anforderungen an das neue Hochrechnungsverfahren – insbesondere wegen der einheitlichen Hochrechnungsfaktoren für Personen und Haushalte – neigen die Hochrechnungsfaktoren dazu, sehr stark zu streuen. Die Korrekturfaktoren  $g_k$ , die definiert sind als Verhältnis von Hochrechnungsfaktor und Eingangsgewicht ( $= w_k/d_k$ ), können durch die geeignete Wahl eines Dämpfungsfaktors  $q_k$  zwischen 0 und 1 beschränkt werden, indem der Vektor der Regressionskoeffizienten aus Formel (4) modifiziert wird:

$$(9) \hat{\mathbf{B}}_q = \left( \sum_{k=1}^n d_k q_k \mathbf{x}_k \mathbf{x}'_k \right)^{-1} \left( \sum_{k=1}^n d_k q_k \mathbf{x}_k y_k \right).$$

Der Dämpfungsfaktor wird in einem iterativen Verfahren unter Berücksichtigung der vorgegebenen Unter- und Obergrenze des Korrekturfaktors berechnet.<sup>15)</sup> Für den Mikrozensus wurde die Untergrenze auf 0,01 und die Obergrenze auf 5 festgelegt.

### 2.4 Behandlung der Antwortausfälle

Antwortausfälle können nicht vernachlässigt werden. Modellhaft wird angenommen, dass die Antwortgewährung eine Zufallsauswahl von Haushalten auf einer zweiten Stufe ist. Die Wahrscheinlichkeit, dass ein Stichprobenhaushalt antwortet, ist unabhängig von den Antwortwahrscheinlichkeiten der übrigen Haushalte. Durch diese Sichtweise als zweistufiges Verfahren kann auch die Hochrechnung in zwei Schritten erfolgen: Zunächst wird von der Nettostichprobe der Antwortenden auf die Bruttostichprobe hochgerechnet, anschließend von der Bruttostichprobe auf die Grundgesamtheit. Auch für den ersten Hochrechnungsschritt wird das oben dargestellte Verfahren der Regressionsschätzung eingesetzt. Allerdings ist hier der Haushalt die Einheit

und nicht die Person. Die Eingangsgewichte sind auf Eins gesetzt. Die Eckwerte, an die angepasst wird, sind die Haushaltszahlen der vollen Bruttostichprobe (antwortende und nicht antwortende Haushalte zusammen). Aus Beobachtungen der Interviewerin bzw. des Interviewers sind Hilfsvariablen für die Antwortausfälle bekannt, sodass die Eckwerte gegliedert werden können. Die Hilfsvariablen sind im nachfolgenden Abschnitt 3.1 erläutert. Der Kehrwert des aus dem ersten Schritt berechneten Faktors kann als Schätzwert für die Antwortwahrscheinlichkeit  $\hat{\theta}_k$  interpretiert werden. Zusammen mit der Ziehungswahrscheinlichkeit  $\pi_k$  dient er gemäß Formel (5) als Eingangsgewicht für den in Abschnitt 2.3 beschriebenen zweiten Schritt der Hochrechnung.

Bei den Antwortausfällen handelt es sich beispielsweise um allein stehende Personen mit einem längeren Krankenhaus- oder Auslandsaufenthalt oder um Haushalte, die zum Zeitpunkt der Berichtswoche noch im Auswahlbezirk gemeldet waren, jedoch bedingt durch Fortzug nicht mehr von der Interviewerin bzw. vom Interviewer angetroffen werden. Zwischen dem ersten Kontaktversuch der Interviewerin bzw. des Interviewers und der Signierung eines Haushalts als Antwortausfall liegen in der Regel mehrere Wochen. Zunächst erhalten Haushalte, die auch nach drei Kontaktversuchen nicht angetroffen wurden bzw. die Erhebungsteilnahme verweigert haben, den so genannten „Selbstaussfüllerbogen“ und werden um eine schriftliche Teilnahme gebeten. Als Antwortausfall wird ein Haushalt erst dann signiert, wenn auch nach mehreren Aufforderungen sowie nach der Androhung von Zwangsgeld kein ausgefüllter Erhebungsbogen beim zuständigen Statistischen Landesamt eingegangen ist. Durch dieses zeitaufwändige Vorgehen werden einige Haushalte erst im jeweils nachfolgenden Berichtsquartal als Befragungsausfälle gekennzeichnet und bei der Berechnung der Kompensationsfaktoren berücksichtigt.<sup>16)</sup> Es wird jedoch davon ausgegangen, dass sich die ausgefallenen Haushalte zwischen den Berichtsquartalen in ihrer Struktur nur zufällig voneinander unterscheiden und daher die zeitlich verschobene Berechnung der Kompensationsfaktoren keine systematische Auswirkung auf die Mikrozensusergebnisse hat. Empirische Untersuchungen der Ausfallstruktur werden vorgenommen, wenn die Daten mehrerer Quartale vorliegen.

## 3 Verwendete Modelle

### 3.1 Hilfsvariablen für die Kompensation der Antwortausfälle

Die Hilfsvariablen für die Kompensation der Antwortausfälle werden vor dem Hintergrund zweier Kriterien ausgewählt: Es sind Hilfsvariablen zu bevorzugen, in deren Klassen überdurchschnittlich hohe Ausfallquoten vorliegen. Zudem muss es für die Interviewer/-innen möglich sein, die Ausprägungen der Hilfsvariablen für die ausgefallenen

15) Der verwendete Algorithmus ist beschrieben in Nieuwenbroeck, N./Boonstra, H.J.: "Bascula 4.0 Reference Manual", Heerlen 2001, S. 57 f.

16) Im Gegensatz dazu hätten auch alle Haushalte als Antwortausfälle gekennzeichnet und in die Berechnung der Kompensationsfaktoren einbezogen werden können, die nicht innerhalb des vom Stichprobenplan vorgesehenen Berichtsquartals an der Erhebung teilgenommen haben. Allerdings könnten dann die Termine für die Bereitstellung der Quartalsergebnisse nicht eingehalten werden. Aus diesem Grund und weil hierdurch die Ausfallquote enorm ansteigen würde, hat sich das Statistische Bundesamt gemeinsam mit den Statistischen Ämtern der Länder für das oben beschriebene Vorgehen entschieden.

Haushalte zu ermitteln, ohne die betroffenen Haushalte zu befragen.<sup>17)</sup> In den vergangenen Jahren kamen überdurchschnittlich hohe Ausfallquoten bei den Einpersonenhaushalten und bei Haushalten mit einer nicht deutschen Haushaltsbezugsperson<sup>18)</sup> vor. Dabei handelte es sich bei den ausgefallenen Einpersonenhaushalten häufig um Haushalte mit Männern unter 60 Jahren und um solche mit Frauen im Alter von 60 Jahren und mehr.

In Übersicht 1 sind die Hilfsvariablen aufgeführt, in deren Abhängigkeit die Kompensationsfaktoren für die Quartals- und Jahresdurchschnittsergebnisse ab 2005 berechnet werden. Bis auf die Hilfsvariablen „Rotationsviertel“ und „Neubauschicht (ja/nein)“ werden ab 2005 die gleichen Merkmale verwendet wie bisher. Das Rotationsviertel wird bei der Berechnung der Kompensationsfaktoren als Hilfsvariable berücksichtigt, weil bei der Erstbefragung die Ausfallquote in der Regel höher ist als in den Folgebefragungen. Ein zentrales Hilfsmerkmal bei der Kompensation der ausgefallenen Haushalte stellt die „Haushaltsgröße“ dar, weil für dieses Merkmal keine externen Daten aus der laufenden Bevölkerungsfortschreibung für eine gebundene Hochrechnung zur Verfügung stehen.

Übersicht 1: Modell für die Kompensation der Antwortausfälle bei Privathaushalten

Regionale Ebene	Kompensationsterme
Bundesland	– Rotationsviertel – Neubauschicht (ja/nein)
Regionale Anpassungsschicht	– Haushaltsgröße (Ein-, Zwei-, Drei- und Mehrpersonenhaushalte) – Staatsangehörigkeit der Haushaltsbezugsperson (deutsch/nicht deutsch) – Wohnsitz der Haushaltsbezugsperson (Hauptwohnung/Nebenwohnung) <i>zusätzlich für Einpersonenhaushalte:</i> – Geschlecht – Alter (unter 60 Jahre/60 Jahre und mehr)
Regionale Untergruppe	– Privathaushalte insgesamt

Bislang wurden die Kompensationsfaktoren durch eine Anpassung an die kombinierte Verteilung der Eckwerte berechnet. Diese waren nach disjunkten Klassen auf der Ebene der regionalen Untergruppen<sup>19)</sup> gegliedert. Um sehr kleine Besetzungszahlen in den Anpassungszellen zu vermeiden, wird ab 2005 an getrennte Randverteilungen angepasst. Im Abschnitt 2.1 wurde bereits darauf verwiesen, dass die Schätzung bei sehr kleinen Besetzungszahlen in den Anpassungszellen instabil wird. Daher wird durch das Hochrechnungsprogramm geprüft, ob je Anpassungszelle mindestens zehn befragte Haushalte vorliegen. Trifft dies nicht zu, so wird die Stichprobenverteilung des betreffenden Merkmals automatisch auf der nächsthöheren regionalen Ebene – Regierungsbezirk oder Bundesland – an die Eckwerte an-

gepasst. Hiervon sind oftmals die Hilfsvariablen „Staatsangehörigkeit der Haushaltsbezugsperson (deutsch/nicht deutsch)“ und „Wohnsitz der Haushaltsbezugsperson (Hauptwohnsitz/Nebenwohnsitz)“ betroffen. Liegen auch auf der Bundesländerebene, also der größten regionalen Einheit, weniger als zehn befragte Haushalte in einer Anpassungsklasse vor, so wird die betreffende Hilfsvariable nicht bei der Berechnung der Kompensationsfaktoren berücksichtigt. Somit stellen die in Übersicht 1 aufgeführten Hilfsvariablen und regionalen Ebenen lediglich das Ausgangsmodell für die Kompensation der ausgefallenen Haushalte dar. Dieses Ausgangsmodell wird in den einzelnen Bundesländern in Abhängigkeit von den Zellbesetzungen unterschiedlich realisiert.

Antwortausfälle von Personen aus Gemeinschaftsunterkünften werden bei der Berechnung der Kompensationsfaktoren ebenfalls berücksichtigt. Hier stellen die Eckwerte die Zahl der befragten und ausgefallenen Personen aus Gemeinschaftsunterkünften je Regierungsbezirk dar. Auch hier prüft das Hochrechnungsprogramm, ob je Regierungsbezirk mindestens zehn befragte Personen aus Gemeinschaftsunterkünften vorliegen.

### 3.2 Hilfsvariablen für die gebundene Hochrechnung

Bezogen auf die Hilfsvariablen für die gebundene Hochrechnung macht das Statistische Amt der Europäischen Gemeinschaften (Eurostat) den Mitgliedstaaten sehr moderate Vorgaben: Es wird dazu geraten, bei der Berechnung der Gewichte für die Hochrechnung die Auswahlwahrscheinlichkeiten und externe Eckwerte über die Verteilung der Grundgesamtheit nach Geschlecht, Alter (5-Jahres-Altersgruppen) und Region [Ebene NUTS II<sup>20)</sup>] zu berücksichtigen, sofern die Eckdaten für hinreichend verlässlich gehalten werden.<sup>21)</sup>

In Übersicht 2 werden die Hilfsvariablen für die gebundene Hochrechnung ab 2005 dargestellt. Die Hilfsvariablen „Alter“ und „Staatsangehörigkeit“ mit den Ausprägungen deutsch, türkisch, EU-25 und nicht EU-25 nach Geschlecht wurden neu in den Hochrechnungsrahmen aufgenommen. Die Hilfsvariable „Alter“ wird nur in drei Anpassungsklassen unterteilt, weil eine feinere Klassifikation tendenziell zu einer stärkeren Streuung der Hochrechnungsfaktoren und somit zu einer instabilen Schätzung der Ergebnisse führt. Der Grund für die differenziertere Anpassung an die Staatsangehörigkeiten ist darin zu sehen, dass die nicht deutsche Bevölkerung bezogen auf ihre Antwortwahrscheinlichkeit eine recht inhomogene Anpassungsklasse bildet. Dies verdeutlicht die Tabelle 1. Hier werden die nach einer freien Hochrechnung resultierende Verteilung der Staatsangehörigkeiten nach Geschlecht und die dazugehörigen Eckwerte für das zweite Vierteljahr 2005 dargestellt. Für die freie Hochrechnung wurde der Faktor „Landesbevölkerung

17) Einige Statistische Landesämter ermitteln die Ausprägungen der Hilfsvariablen über die ausgefallenen Haushalte zusätzlich durch die Einwohnermelderegister.

18) Die Haushaltsbezugsperson ist die erste im Erhebungsbogen eingetragene Person. Der Fragebogen gibt als Reihenfolge vor: Eheleute, Kinder, Verwandte, Familienfremde.

19) Regionale Untergruppen umfassen einzelne oder mehrere Kreise. Eine regionale Untergruppe besteht aus mindestens 100 000 Einwohnern. Bundesweit gibt es insgesamt 379 regionale Untergruppen.

20) Die Ebene NUTS II entspricht in Deutschland den Regierungsbezirken.

21) Siehe Verordnung (EG) Nr. 577/98 des Rates vom 9. März 1998 zur Durchführung einer Stichprobenerhebung über Arbeitskräfte in der Gemeinschaft, Artikel 3 Absatz 5.

Übersicht 2: Modell für die Hochrechnung der Quartals- und Jahresdurchschnittsergebnisse

Regionale Ebene	Hochrechnungsterme	Quelle
Bundesland	– Alter (unter 15, 15 bis 44, 45 Jahre und mehr) differenziert nach Geschlecht – Staatsangehörigkeit (deutsch, türkisch, EU-25, nicht EU-25) differenziert nach Geschlecht – Zeit- und Berufssoldaten einschließlich Bundes- und Bereitschaftspolizei, Grundwehrdienstleistende, Zivilbevölkerung – Bevölkerung insgesamt je Monat	Laufende Bevölkerungsfortschreibung Ausländerzentralregister Bundesministerium der Verteidigung, Bundesministerium des Innern, Bundespolizei Laufende Bevölkerungsfortschreibung
Regierungsbezirk	– Staatsangehörigkeit (deutsch, nicht deutsch) differenziert nach Geschlecht	Laufende Bevölkerungsfortschreibung
Regionale Anpassungsschicht	– Bevölkerung insgesamt	Laufende Bevölkerungsfortschreibung

geteilt durch befragte Personen am Hauptwohnsitz je Bundesland“ gewählt. Bei den türkischen Frauen und Männern weichen die frei hochgerechneten Stichprobenverteilungen weniger von den Eckwerten ab als bei der übrigen nicht deutschen Bevölkerung. Weil die nicht deutsche Bevölkerung bisher in einer gemeinsamen Klasse gegliedert nach Geschlecht an die Eckwerte angepasst wurde, waren die Türken nach der gebundenen Hochrechnung überrepräsentiert und die restliche nicht deutsche Bevölkerung weiterhin unterrepräsentiert. Dies gilt insbesondere für die nicht deutsche Bevölkerung mit einer Staatsangehörigkeit außerhalb der EU25-Staaten.

Tabelle 1: Unterschiedlicher Anpassungsbedarf bei der nicht deutschen Bevölkerung nach Geschlecht  
Ergebnis des Mikrozensus 2. Vj 2005  
1 000

Anpassungsklassen	Bevölkerung am Hauptwohnsitz		
	Eckwert	freie Hochrechnung <sup>1)</sup>	Differenz
männlich, türkisch . . . .	1011,5	750,0	261,5
männlich, EU-25 . . . . .	1228,3	877,1	351,2
männlich, nicht EU-25 . .	1579,7	1065,1	514,6
weiblich, türkisch . . . . .	881,0	727,5	153,5
weiblich, EU-25 . . . . .	1092,7	857,8	234,9
weiblich, nicht EU-25 . .	1534,2	1178,6	355,6

1) Hier wurden die Stichprobenfallzahlen mit dem Faktor „Landesbevölkerung geteilt durch befragte Personen am Hauptwohnsitz“ hochgerechnet.

Bei der gebundenen Hochrechnung werden die Hochrechnungsterme durch das Programm automatisch reduziert, wenn in einer Zelle weniger als zehn befragte Personen vorliegen. Auch hier soll eine instabile Schätzung durch zu geringe Fallzahlen in den Anpassungsklassen vermieden werden. Besonders in den neuen Bundesländern wird auf der Länderebene häufig nicht an die Staatsangehörigkeit (deutsch, türkisch, EU-25, nicht EU-25) nach Geschlecht, sondern nur an die deutsche und nicht deutsche Staatsangehörigkeit nach Geschlecht angepasst. Zudem werden die Grundwehrdienstleistenden in einigen Bundesländern oftmals mit den Zeit- und Berufssoldaten einschließlich Bundes- und Bereitschaftspolizei in einer gemeinsamen Klasse zusammengefasst. In einigen Bundesländern liegen

anschließend immer noch weniger als zehn befragte Personen je Klasse vor, sodass die Hilfsvariable ganz aus dem Hochrechnungsrahmen entfernt wird. Wie das in Übersicht 2 dargestellte Modell für die Hochrechnung der Quartals- und Jahresdurchschnittsergebnisse<sup>22)</sup> in den einzelnen Bundesländern im zweiten Vierteljahr 2005 im Einzelnen realisiert wurde, ist in Übersicht 3 auf S. 1046 abgebildet.

#### 4 Schätzung der Eckwerte

In Übersicht 2 werden auch die Quellen genannt, aus denen die Eckwerte für die Hochrechnung stammen. Sie stellen, wie die Mikrozensusergebnisse ab 2005, Quartalsdurchschnitte dar. Weil zum Zeitpunkt der Hochrechnung keine aktuellen Eckwerte für das betreffende Berichtsquartal aus der laufenden Bevölkerungsfortschreibung und dem Ausländerzentralregister zur Verfügung stehen, müssen die benötigten Eckwerte geschätzt werden<sup>23)</sup>. Hierbei ist zu berücksichtigen, dass die geschätzten Eckwerte trotz der unterschiedlichen Quellen<sup>24)</sup> und regionalen Ebenen in sich konsistent sein müssen. Beispielsweise muss die Zahl der nicht deutschen Frauen und Männer je Regierungsbezirk in der Summe mit der Zahl der nicht deutschen Frauen und Männer je Bundesland übereinstimmen.

In Schaubild 1 auf S. 1046 wird die Schätzung der Eckwerte beschrieben. Zuerst wird je Bundesland die „Bevölkerung insgesamt“ durch die Schätzmethode nach Winters für das Hochrechnungsquartal geschätzt. Die Basis hierfür sind der jeweils letzte verfügbare Wert für die Landesbevölkerung aus der laufenden Bevölkerungsfortschreibung sowie unabhängige Zeitreihen über die Zahlen der Geborenen, Gestorbenen und der Zu- und der Fortzüge. Der letzte verfügbare Wert für die Landesbevölkerung steht vier bis fünf Monate nach dem Berichtsmonat zur Verfügung. Die Zeitreihen bestehen aus mindestens 24 Monatswerten und haben jeweils eine eigene Saisonfigur.

Die Eckwerte werden anschließend aus den relativen Anteilen der entsprechenden Bevölkerungsgruppe an der zuvor geschätzten Landesbevölkerung gebildet: Die Statistischen Ämter der Länder übermitteln an das Statistische Bundes-

22) Die für die Quartalsergebnisse berechneten Hochrechnungsfaktoren werden geviertelt und als Hochrechnungsfaktoren für die Jahresdurchschnittsergebnisse verwendet.  
 23) Eine Ausnahme bilden die Eckwerte für die Zeit- und Berufssoldaten, die Grundwehrdienstleistenden und die Bundespolizei. Diese werden vom Bundesministerium für Verteidigung und von der Bundespolizei quartalsweise an das Statistische Bundesamt übermittelt. Vom Bundesministerium des Innern wird dem Statistischen Bundesamt jeweils für den Stichtag 31.12. die Zahl der Bereitschaftspolizisten zur Verfügung gestellt. Diese Werte werden für das gesamte folgende Jahr als Eckwerte verwendet.  
 24) Das Ausländerzentralregister und die laufende Bevölkerungsfortschreibung weisen unterschiedliche Zahlen für die in Deutschland lebende nicht deutsche Bevölkerung aus.

Übersicht 3: Realisiertes Hochrechnungsmodell nach Bundesländern im 2. Vierteljahr 2005

Bundesländer	Hochrechnungsterm					
	Bundesland				Regierungsbezirk	regionale Anpassungsschicht
	Bevölkerung insgesamt je Monat	Alter (unter 15, 15 bis 44, 45 und mehr Jahre) nach Geschlecht	Staatsangehörigkeit (deutsch, türkisch, EU-25, nicht EU-25) nach Geschlecht	Zeit- und Berufssoldaten einschließlich Bundes- und Bereitschaftspolizei/ Grundwehrdienstleistende/Zivilbevölkerung	Staatsangehörigkeit (deutsch, nicht deutsch) nach Geschlecht	Bevölkerung insgesamt
Schleswig-Holstein .....	•	•	•	R3	X	•
Hamburg .....	•	•	•	—	X	X
Niedersachsen .....	•	•	•	R3	•	•
Bremen .....	•	•	R1, R2	—	X	•
Nordrhein-Westfalen .....	•	•	•	R3	•	•
Hessen .....	•	•	•	R3	•	•
Rheinland-Pfalz .....	•	•	•	R3	•	•
Baden-Württemberg .....	•	•	•	R3	•	•
Bayern .....	•	•	•	R3	•	•
Saarland .....	•	•	•	—	X	X
Berlin .....	•	•	•	R3	X	•
Brandenburg .....	•	•	R2	R3	X	•
Mecklenburg-Vorpommern ..	•	•	R2	R3	X	•
Sachsen .....	•	•	R2	R3	•	•
Sachsen-Anhalt .....	•	•	R2	R3	R4	•
Thüringen .....	•	•	R2	R3	X	•

Erläuterungen:

- Anpassung erfolgt
- Anpassung erfolgt auch nicht reduziert
- X Bundesland ist nicht in Regierungsbezirke bzw. regionale Anpassungsschichten unterteilt

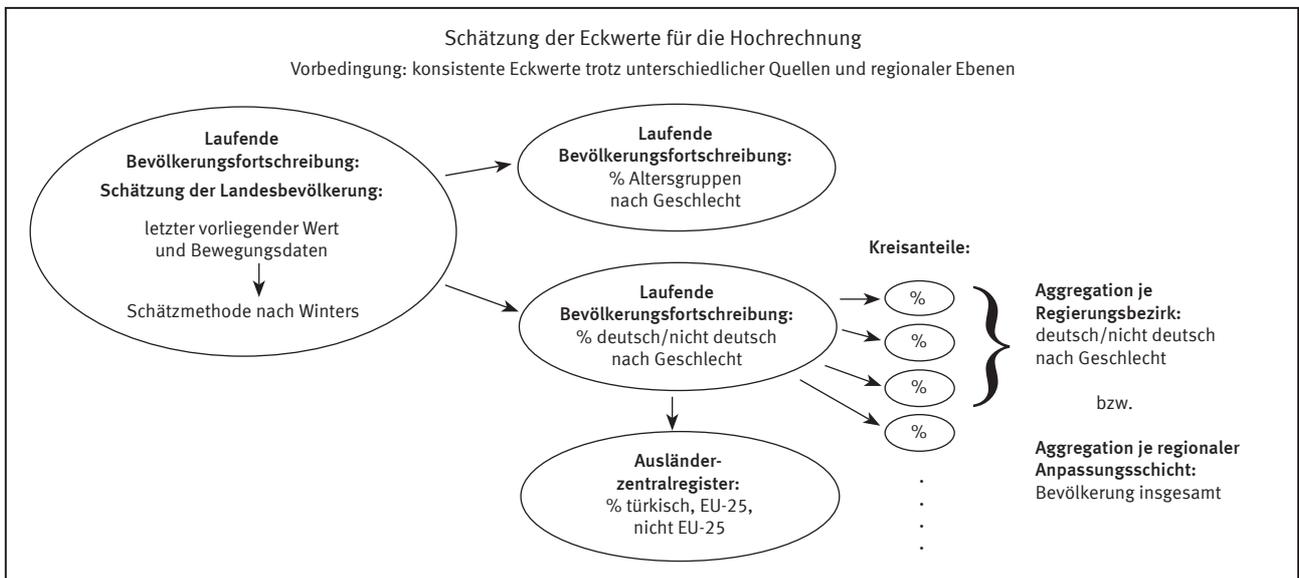
Anpassung erfolgt reduziert:

- R1: Staatsangehörigkeit (deutsch, türkisch, EU-25, nicht EU-25)
- R2: Staatsangehörigkeit (deutsch, nicht deutsch) nach Geschlecht
- R3: Zeit- und Berufssoldaten einschließlich Bundes- und Bereitschaftspolizei, Grundwehrdienstleistende/Zivilbevölkerung
- R4: Geschlecht je Regierungsbezirk

amt monatlich – mit einer zeitlichen Verzögerung von vier bis fünf Monaten – die Bevölkerung nach den für die Hochrechnung benötigten Altersgruppen und nach Geschlecht sowie die deutsche und nicht deutsche Bevölkerung nach Geschlecht. Diese Angaben bilden die Grundlage für die Ableitung der entsprechenden Anteile. Untersuchungen haben gezeigt, dass sich diese im Zeitverlauf nur sehr langsam verändern und somit eine gute Schätzgrundlage darstellen.

Um Sprünge zum Jahreswechsel zu vermeiden (diese können infolge einer einmaligen Verbuchung der Einbürgerungen am Jahresende bzw. beim Übergang in eine neue Altersgruppe entstehen), werden die Anteile mithilfe der gleitenden Mittelwerte gebildet. Die zuvor geschätzte Landesbevölkerung wird anhand der Anteile auf die entsprechenden Bevölkerungsgruppen umgerechnet. Danach werden die Anteile der nicht deutschen Bevölkerung aus dem

Schaubild 1



Ausländerzentralregister<sup>25)</sup> auf die zuvor ermittelte nicht deutsche Bevölkerung aus der laufenden Bevölkerungsfortschreibung umgerechnet. Auch die Kreisanteile sind im Zeitverlauf relativ stabil. Sie liegen ebenfalls mit einer zeitlichen Verzögerung von vier bis fünf Monaten vor und werden auf die zuvor auf Länderebene ermittelte deutsche und nicht deutsche Bevölkerung nach Geschlecht umgerechnet. Anschließend werden die Kreiswerte je Regierungsbezirk bzw. je regionaler Anpassungsschicht aggregiert.

Alternativ zu dem beschriebenen Vorgehen könnten die Eckwerte auch auf Kreisebene geschätzt und anschließend je regionaler Anpassungsschicht, Regierungsbezirk und Bundesland aggregiert werden. Untersuchungen haben jedoch gezeigt, dass eine Schätzung der nicht deutschen Bevölkerung nach Geschlecht auf Kreisebene großen Schwankungen unterliegt. Der Grund hierfür sind die sehr geringen Fallzahlen.

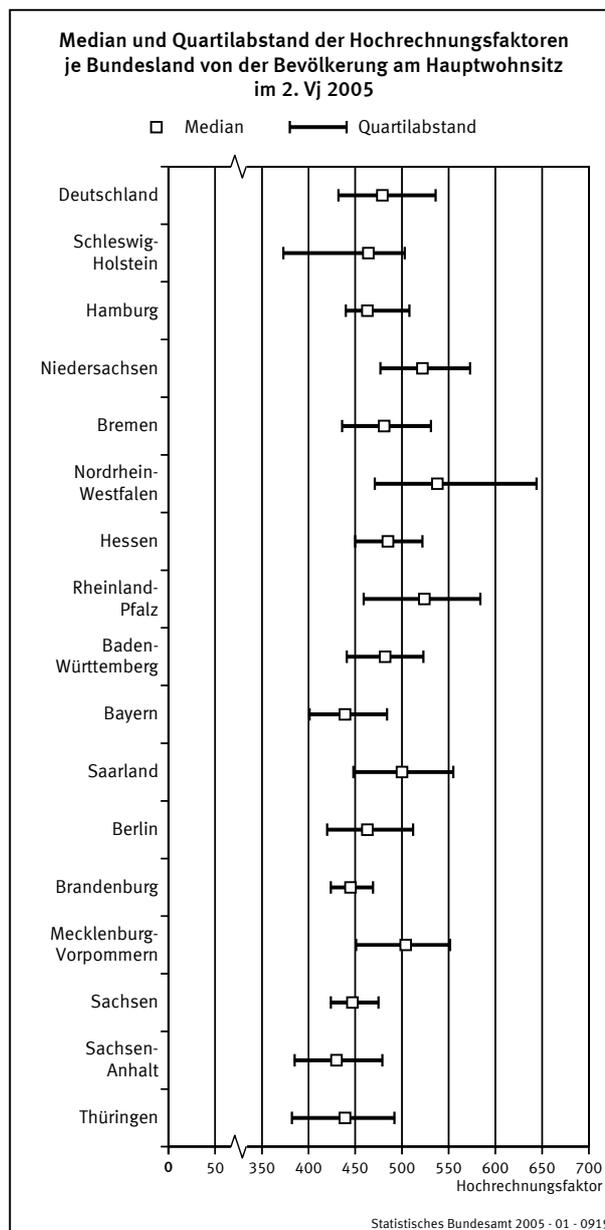
### 5 Hochrechnungsfaktoren im zweiten Vierteljahr 2005

Zum jetzigen Zeitpunkt (Mitte Oktober 2005) liegen im Statistischen Bundesamt die Daten für das zweite Quartal 2005 vor. In Schaubild 2 sind der Median und der Quartilabstand der berechneten Hochrechnungsfaktoren für Deutschland und je Bundesland für das zweite Quartal 2005 abgebildet. Der Median der Hochrechnungsfaktoren variiert zwischen den Bundesländern. Dies ist vor allem auf die unterschiedliche Zahl der realisierten Interviews je Bundesland zurückzuführen. Pro Quartal sollten in jedem Bundesland etwa 0,25% der Bevölkerung befragt werden. Dies entspricht einem durchschnittlichen Hochrechnungsfaktor von etwa 400. Weil in allen Bundesländern weniger als 0,25% der Bevölkerung befragt wurde, sind die Hochrechnungsfaktoren im Durchschnitt größer als 400. Auch der Quar-

Tabelle 2: Statistik über die Hochrechnungsfaktoren je Bundesland (Bevölkerung am Hauptwohnsitz)  
Ergebnis des Mikrozensus 2. Vj 2005

Bundesland	Arithmetisches Mittel	Standardabweichung	Median	1. Quartil	3. Quartil
Schleswig-Holstein .	460	105	464	373	503
Hamburg .....	489	82	463	440	508
Niedersachsen .....	530	90	522	477	573
Bremen .....	516	202	481	436	531
Nordrhein-Westfalen	580	165	538	471	644
Hessen .....	493	77	485	450	522
Rheinland-Pfalz ....	530	112	524	459	584
Baden-Württemberg	495	78	482	441	523
Bayern .....	454	89	439	401	484
Saarland .....	512	108	500	448	555
Berlin .....	484	102	463	420	512
Brandenburg .....	461	191	445	424	469
Mecklenburg-Vorpommern .....	510	96	504	451	551
Sachsen .....	460	107	447	424	475
Sachsen-Anhalt ....	439	71	430	385	479
Thüringen .....	446	117	439	382	492
Deutschland ...	501	124	479	432	536

Schaubild 2



tilabstand der Hochrechnungsfaktoren variiert zwischen den Bundesländern. In Nordrhein-Westfalen ist der Quartilabstand am größten. Weitere Statistiken über die Hochrechnungsfaktoren je Bundesland befinden sich in der Tabelle 2.

Das Schaubild 3 auf S. 1048 zeigt den Median und den Quartilabstand der Hochrechnungsfaktoren je Anpassungskategorie. Der Median variiert deutlich stärker zwischen den Anpassungsklassen als zwischen den Bundesländern. Die nicht deutsche Bevölkerung hat im Durchschnitt deutlich höhere Hochrechnungsfaktoren als die deutsche Bevölkerung. Auch in Schaubild 3 wird deutlich, dass eine getrennte Anpassung der nicht deutschen Bevölkerung nach türkisch,

25) Das Ausländerzentralregister weist die nicht deutsche Bevölkerung nach Geschlecht jeweils zum Jahresende (Stichtag: 31.12.) aus. Für die Schätzung der Staatsangehörigkeit (türkisch, EU-25, nicht EU-25) nach Geschlecht werden das ganze folgende Jahr über die gleichen Anteile verwendet. Eine Ausnahme bilden Berlin und Hamburg: Hier wird eine eigene Auszählung der Melderegister am Jahresende vorgenommen und für die Schätzung der Staatsangehörigkeit (türkisch, EU-25, nicht EU-25) nach Geschlecht verwendet.

Schaubild 3

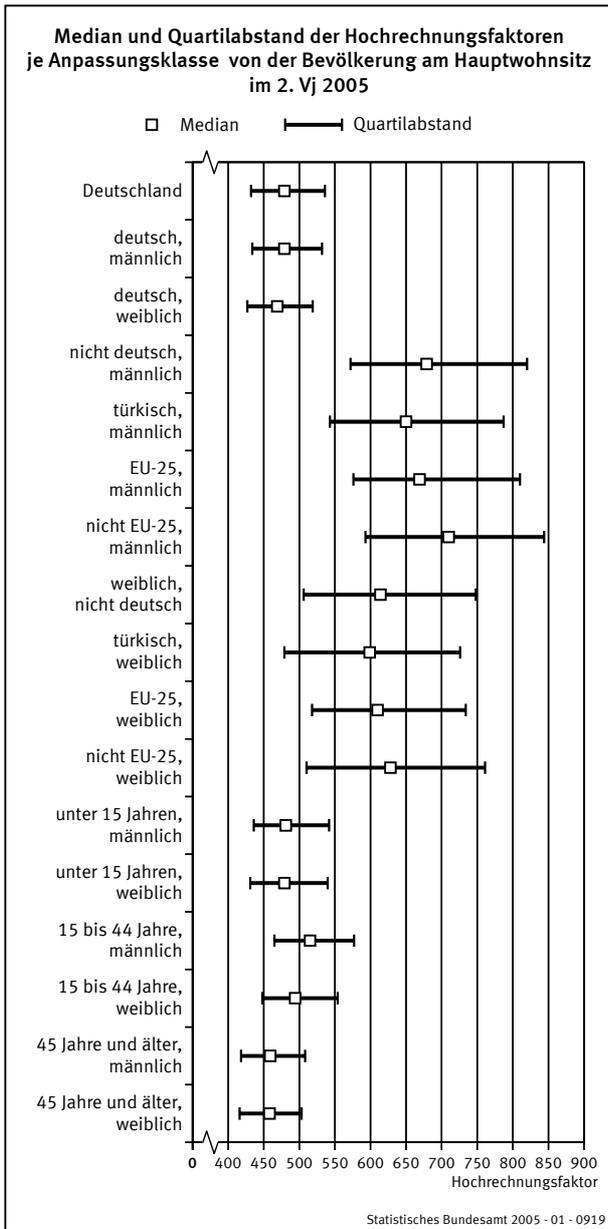


Tabelle 3: Statistik über die Hochrechnungsfaktoren je Anpassungsklasse (Bevölkerung am Hauptwohnsitz) Ergebnis des Mikrozensus 2. Vj 2005

Anpassungsklasse	Arithmetisches Mittel	Standardabweichung	Median	1. Quartil	3. Quartil
deutsch, männlich ..	494	100	479	434	532
deutsch, weiblich ..	483	101	469	427	519
nicht deutsch, männlich .....	721	245	679	572	820
türkisch .....	687	206	650	543	787
EU-25 .....	718	229	669	576	810
nicht EU-25 .....	748	277	710	593	844
nicht deutsch, weiblich .....	644	215	614	506	748
türkisch .....	625	203	599	479	726
EU-25 .....	644	205	610	518	734
nicht EU-25 .....	656	228	628	510	761
männlich, unter 15 Jahren .....	509	128	481	436	542
weiblich, unter 15 Jahren .....	504	135	479	431	540
männlich, 15 bis 44 Jahre .....	542	138	515	465	577
weiblich, 15 bis 44 Jahre .....	517	125	494	448	554
männlich, 45 Jahre und älter .....	480	112	459	418	508
weiblich, 45 Jahre und älter .....	473	105	458	416	503
Deutschland ...	501	124	479	432	536

EU-25 und nicht EU-25 sinnvoll ist, weil sich der Median in den genannten Klassen leicht unterscheidet. Die Hochrechnungsfaktoren der nicht deutschen Bevölkerung weisen zudem eine größere Streuung auf als die Hochrechnungsfaktoren in den übrigen Anpassungsklassen. Weitere Statistiken über die Hochrechnungsfaktoren je Anpassungsklasse enthält Tabelle 3.

## 6 Die eingesetzte Software „Bascula“

Im Abschnitt 2.1 wurden die Rahmenbedingungen für das Hochrechnungsverfahren ab 2005 (gleiche Hochrechnungsfaktoren für alle Personen eines Haushalts, getrennte Anpassung an Randverteilungen, Beschränkung der Hochrech-

nungsfaktoren) bereits erläutert. Diese Anforderungen können durch das bisher verwendete Großrechnerprogramm datenverarbeitungstechnisch nicht umgesetzt werden. Das Blaisemodul „Bascula“ stellt jedoch alle benötigten Funktionalitäten zur Verfügung.

Zudem bietet Bascula die Möglichkeit, für die Anpassungsklassen eine minimale Zellbesetzung vorzugeben. Wird der vorgegebene Wert unterschritten, so reduziert Bascula automatisch den betreffenden Term. Hierbei wird geprüft, ob die Hilfsvariablen der reduzierten Terme gleichzeitig in einem anderen Term des Hochrechnungsmodells enthalten sind. Ist dies der Fall, so werden die reduzierten Terme aus dem Hochrechnungsmodell entfernt, da es sich um redundante Informationen handelt.

Weiter ist es möglich, in Bascula eine maximal tolerierte Differenz zwischen den hochgerechneten Werten und den Eckwerten einzustellen. Hierdurch werden die Hochrechnungsfaktoren in der Regel innerhalb der vorgegebenen Grenzen berechnet, selbst wenn die Stichprobenverteilungen stark von den Eckwerten abweichen.

Die Kompensations- und Hochrechnungsfaktoren können im Dialog oder durch einen Batch-Prozess berechnet werden. Für die Mitarbeiterinnen und Mitarbeiter der Statistischen Landesämter wurde ein vollständig automatisierter Batch-Prozess programmiert. Zu jedem Hochrechnungslauf wird durch „Bascula“ automatisch eine Log-Datei generiert. Diese Log-Datei zeigt beispielsweise an, ob das Programm fehlerfrei durchgelaufen ist, ob und wie die Terme reduziert wurden, wie stark die Stichprobenwerte von den Eckwerten abweichen und ob die vorgegebenen Eckwerte durch die berechneten Hochrechnungsfaktoren getroffen werden. Das neue Hochrechnungsprogramm hat sich bereits im ersten und zweiten Vierteljahr des Jahres 2005 bewährt. [uu](#)

## Auszug aus Wirtschaft und Statistik

© Statistisches Bundesamt, Wiesbaden 2005

Für nichtgewerbliche Zwecke sind Vervielfältigung und unentgeltliche Verbreitung, auch auszugsweise, mit Quellenangabe gestattet. Die Verbreitung, auch auszugsweise, über elektronische Systeme/Datenträger bedarf der vorherigen Zustimmung. Alle übrigen Rechte bleiben vorbehalten.

Herausgeber: Statistisches Bundesamt, Wiesbaden

Schriftleitung: Johann Hahlen  
Präsident des Statistischen Bundesamtes  
Verantwortlich für den Inhalt:  
Brigitte Reimann,  
65180 Wiesbaden

- Telefon: +49 (0) 6 11/75 20 86
- E-Mail: [wirtschaft-und-statistik@destatis.de](mailto:wirtschaft-und-statistik@destatis.de)

Vertriebspartner: SFG Servicecenter Fachverlage  
Part of the Elsevier Group  
Postfach 43 43  
72774 Reutlingen  
Telefon: +49 (0) 70 71/93 53 50  
Telefax: +49 (0) 70 71/93 53 35  
E-Mail: [destatis@s-f-g.com](mailto:destatis@s-f-g.com)

Erscheinungsfolge: monatlich



Allgemeine Informationen über das Statistische Bundesamt und sein Datenangebot erhalten Sie:

- im Internet: [www.destatis.de](http://www.destatis.de)

oder bei unserem Informationsservice  
65180 Wiesbaden

- Telefon: +49 (0) 6 11/75 24 05
- Telefax: +49 (0) 6 11/75 33 30
- [www.destatis.de/kontakt](http://www.destatis.de/kontakt)