

Pischner, Rainer

Working Paper — Digitized Version

Eine konsistente Haushalts- und Personengewichtung für die DDR-Basisbefragung des SOEP und für die Ost-Pilotstudie des Wohlfahrtssurveys

DIW Discussion Papers, No. 18a

Provided in Cooperation with:

German Institute for Economic Research (DIW Berlin)

Suggested Citation: Pischner, Rainer (1991) : Eine konsistente Haushalts- und Personengewichtung für die DDR-Basisbefragung des SOEP und für die Ost-Pilotstudie des Wohlfahrtssurveys, DIW Discussion Papers, No. 18a, Deutsches Institut für Wirtschaftsforschung (DIW), Berlin

This Version is available at:

<https://hdl.handle.net/10419/103365>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



Diskussionspapiere
Discussion Papers

Diskussionspapier Nr. 18a

**Eine konsistente
Haushalts- und Personengewichtung
für die DDR-Basisbefragung des SOEP und
für die Ost-Pilotstudie des Wohlfahrtssurveys**

von Rainer Pischner

Die in diesem Papier vertretenen Auffassungen liegen ausschließlich in der Verantwortung des Verfassers und nicht in der des Instituts.

Opinions expressed in this paper are those of the author and do not necessarily reflect views of the Institute.

Diskussionspapier Nr. 18a

**Eine konsistente
Haushalts- und Personengewichtung
für die DDR-Basisbefragung des SOEP und
für die Ost-Pilotstudie des Wohlfahrtssurveys**

von Rainer Pischner

Eine konsistente Haushalts- und Personengewichtung für die DDR-Basisbefragung des SOEP und für die Ost-Pilotstudie des Wohlfahrtssurveys¹

0 Problemaufriß

Aktuelle statistische Informationen über die sozio-demographische Struktur der privaten Haushalte liegen für die ehemalige DDR bzw. für die neuen Bundesländern (einschl. Berlin-Ost) nur spärlich vor². inzwischen stehen aktuelle Umfragedaten aus dem sozio-ökonomischen Panel (SOEP-Ost) und der Pilotstudie zum Wohlfahrtssurvey aus dem Jahr 1990 zur Verfügung, die diesen Mangel beheben sollen. Natürlich enthalten diese Erhebungen gewisse demographische Verzerrungen, wie sie bei allen Stichproben vorkommen, auch wenn sie repräsentativ angelegt wurden. So werden meist Ein-Personenhaushalte und ältere Menschen untererfaßt. Um die Vergleichbarkeit der beiden Stichproben zu ermöglichen, bietet es sich an, auf einem gemeinsamen Hochrechnungskonzept basierend, konsistente Personen- und Haushaltsstrukturen zu schätzen.

Im folgenden wird am Beispiel der genannten Umfragen gezeigt, wie mit Hilfe von wenigen exogenen Daten über Personenstrukturen zunächst für Personen und Haushalte des SOEP eine in sich konsistente Hochrechnung erstellt wird. Auf Basis dieses Konzeptes und unter Einbeziehung der Ergebnisse wird dann die Gewichtung des Wohlfahrtssurveys vorgenommen.

1 Der vorliegende Beitrag ist eine überarbeitete und ^{'''}(um den Wohlfahrtssurvey) erweiterte Fassung des DIW-Diskussionspapiers Nr. 18 (Pischner 1991).

2 Die letzten Daten der amtlichen Statistik der DDR beziehen sich auf das Jahr 1988, wobei es sich aber nur um Fortschreibungsdaten handelt. Vgl. Bedau und Vortmann (1990).

1 Die Stichprobenanlage des SOEP-Ost und des Wohlfahrtssurveys.

Konzeption und Feldarbeit der DDR-Basisbefragung werden an anderer Stelle von Schupp und Wagner ausführlich dargestellt³. Eine ausführliche Beschreibung des Stichprobenplans ist zu finden im Methodenbericht von Infratest.⁴ Über den Wohlfahrtssurvey berichten Habich und Landua⁵. Deshalb sollen im folgenden nur einige wesentliche Punkte zu den beiden Stichprobenanlagen erwähnt werden.

1.1 Zur Stichprobe des SOEP-Ost

Die Grundgesamtheit der DDR-Basisbefragung des SOEP bilden alle Privathaushalte der damaligen DDR mit deutscher Bezugsperson bzw. deutschem Haushaltsvorstand und den in ihnen wohnenden Personen, die zum Befragungszeitpunkt (Juni 1990) mindestens 16 Jahre alt waren.

Der Stichprobenplan basiert auf dem von Infratest/München erstellten Mastersample. Dieses ist als geschichtete, mehrphasige Stichprobe aufgebaut und umfaßt 753 Sample-Points, die proportional zur Einwohnerzahl durch Schichtung von 217 Land- bzw. Stadtkreisen und fünf Gemeindegrößenklassen gezogen wurden. Aus ihnen wurden Gemeinden durch eine systematische Auswahl mit Zufallsstart ermittelt. Die Auswahl der Startadressen erfolgte proportional zur Zahl der benötigten Sample-Points je Gemeinde. Startadressen für den Random-Walk sind durch systematische Auswahl nach Zufallsstart aus der zentralen Einwohnerdatei gewonnen worden. Von den Startadressen ausgehend sind zehn Befragungshaushalte und zwei Reservehaushalte ermittelt worden, von denen schließlich acht Interviews zu realisieren waren (vgl. hierzu auch weiter unten den Abschnitt 2).

3 Schupp und Wagner (1991).

4 Infratest Sozialforschung (1990).

5 Habich und Landua (1991).

1.2 Die Stichprobenanlage der Pilotstudie des Wohlfahrtssurveys-Ost

Der Wohlfahrtssurvey-Ost - im weiteren auch als WfS bezeichnet - ist wie das SOEP eine Replikation der bundesrepublikanischen Vorbildstudie. Die November-Stichprobe des WfS-Ost war 1990 als Pilotstudie angelegt. Sie sollte bereits erste Auswirkungen der Vereinigung messen.

Basis dieser Stichprobe ist eine Zufallsziehung von Adressen aus der Gesamtheit der wahlberechtigten Bevölkerung auf Grundlage der zentralen Einwohnermeldekartei. Aus insgesamt 10.000 Adressen wurden für den WfS 2.000 Adressen zur Verfügung gestellt. 212 Start- und weitere 100 Reserveadressen wurden aus diesem Reservoir geschöpft, indem durch systematische Zufallsauswahl jede neunte Adresse für den WfS gewonnen wurde. Wie beim SOEP wurden diese als Ausgangspunkt für den Random-Walk verwendet. Insgesamt wurden 1.050 Bruttoadressen eingesetzt. 24 qualitätsneutrale Ausfälle ergaben ein bereinigtes Brutto von 1.026 Adressen, von denen schließlich 735 zu Interviews führten. Dies ergibt die hohe Ausschöpfungsquote von 72 vH.

2 Warum ist eine Gewichtung von Zufalls-Stichproben erforderlich?

Die Notwendigkeit der Gewichtung repräsentativ angelegter Stichprobe soll am Beispiel der DDR-Basiserhebung im Rahmen des sozio-ökonomischen Panels verdeutlicht werden:

Anfang Juli 1990 endete die Feldarbeit für die erste Welle des SOEP-Ost⁶. Diese Stichprobe wurde trotz der sehr kurzen Vorbereitungszeit dem bundesdeutschen Standard entsprechend als bevölkerungsrepräsentative Stichprobe angelegt.

So sorgfältig Erhebungen geplant und durchgeführt werden, während der Feldarbeit kommt es zu zufälligen Fehlern und meist

6 Vgl. Schupp und Wagner (1990).

auch zu systematischen Verzerrungen, die eine Beeinträchtigung der Stichprobenrepräsentativität zur Folge haben: So standen zu Beginn der Feldarbeit 3.616 Adressen zur Verfügung, die über das für die DDR erstellte Mastersample durch Adressenziehung und Random-Walk ermittelt wurden⁷. Von diesem unbereinigten Adressenbestand wurden 162 Anschriften nicht benötigt, da die Interviewer bereits ihre erforderlichen acht Interviews in den jeweiligen Sample-Points realisieren konnten. Bei weiteren 50 Adressen fanden sie leere Wohnungen vor. Schließlich fielen vollständige Sample-Points mit insgesamt 290 Anschriften aus der Stichprobe, weil sie - bedingt durch die kurze Feldzeit - nicht rechtzeitig bearbeitet werden konnten. Die insgesamt 502 Ausfälle gelten zwar als qualitätsneutral, d.h. ohne Einfluß auf die Repräsentativität, doch dürften zumindest die 290 Totalausfälle zu Verzerrungen im Design der Stichprobe, d.h. zu regionalen Ungleichgewichten, geführt haben. Jedenfalls verblieben 3.114 Adressen, die als bereinigtes Brutto die Ausgangsbasis der Befragung bildeten.

Nach Abschluß der Feldarbeit lagen für 2.179 Haushalte und 4.453 Personen ab 16 Jahren Interviews vor. Dies entspricht einer - auf das bereinigte Brutto bezogen - Ausschöpfungsquote von 70 vH. Von den ursprünglich 3616 Adressen, die das Gebiet der DDR repräsentativ abbildeten, wurden indes nur 60 vH realisiert. Zwar ist dies ein gutes, verglichen mit dem SOEP-West (57 vH) sogar ein sehr gutes Ergebnis, dennoch ist die Repräsentativität der Stichprobe durch die Ausfälle, wenn sie auch teilweise qualitätsneutral waren, systematisch beeinträchtigt. Hinzu kommen unsystematische Stichprobenfehler, die zufallsbedingt sind.

Obwohl unter bestimmten Voraussetzungen erwartungstreue Schätzungen für die Parameter bedingter Verteilungen auch aus nichtrepräsentativen Stichproben gewonnen werden können, ist zu-

7 Vgl. von Rosenblatt und Schupp (1990).

zumindest für die Analyse von Kontingenztabellen, die für eine leicht interpretierbare Darstellung von Sachverhalten unverzichtbar sind, eine Entzerrung der Stichprobenstrukturen, also eine Gewichtung erforderlich. Zur Gewinnung von Aussagen über die Grundgesamtheit der Haushalte und Personen sind Hochrechnungsfaktoren notwendig. Diese unterscheiden sich in der Regel aber nur durch einen einzigen Faktor, mit dem die Gewichtungsvektoren noch zu multiplizieren sind. Deshalb werden in diesem Aufsatz die Begriffe Gewichte und Hochrechnungsfaktoren synonym verwendet. Sollte eine Unterscheidung notwendig sein, wird explizit im Text darauf hingewiesen.

3 Methodische Grundlagen der Gewichtung

In diesem Aufsatz geht es vornehmlich um die Durchführung der Gewichtung bzw. Hochrechnung für die Basiserhebung des DDR-Panels und des Wohlfahrtssurveys. Deshalb wird auch der theoretische Teil des Gewichtungsverfahrens, der für das bundesdeutsche SOEP an anderer Stelle erläutert wird, hier sehr knapp gehalten^{8 9}:

Den Hochrechnungen sämtlicher Stichproben liegt der Grundgedanke von Horvitz/Thompson¹⁰ zu Grunde, daß die inverse Auswahlwahrscheinlichkeit eines Stichprobenelements einen unverzerrten Schätzwert für den Hochrechnungsfaktor jener Elemente darstellt. Für das SOEP-Ost soll hier unter Auswahlwahrscheinlichkeit die Wahrscheinlichkeit verstanden werden, daß von einem beliebigen privaten Haushalt in der DDR mit deutscher Bezugsperson ein ausgefüllter Haushaltsfragebogen sowie von allen Personen, die 16 Jahre und älter sind, ein Personen-

8 Einen umfassenden Überblick zur Berechnung konsistenter Hochrechnungsrahmen geben die Beiträge von Galler (1977) und Merz (1983).

9 Zu den theoretischen Grundlagen der Gewichtung und Hochrechnung für die Panels im Quer- und Längsschnitt sei verwiesen auf Rendtel (1987).

10 Siehe hierzu die grundlegenden Arbeiten von Horvitz und Thompson (1952).

fragebogen vorliegt. Für den WfS ist es die Wahrscheinlichkeit, daß ein ausgefüllter Personenfragebogen von genau einer Person vorliegt, die 18 Jahre und älter ist.

Diese Wahrscheinlichkeiten setzen sich aus vielen (bedingten) Einzelwahrscheinlichkeiten zusammen:

- Auswahlwahrscheinlichkeit der Gemeinden und der Sample-Points,
- Wahrscheinlichkeit der Kontaktaufnahme,
- Antwortbereitschaft der Haushalte bzw. Personen.

Hätte jeder Haushalt und jede Person dieselbe Auswahlwahrscheinlichkeit, wären sämtliche Gewichte identisch und gleich eins; die Hochrechnungsfaktoren ergäben sich als reziproker Auswahlssatz, also als Grundgesamtheit dividiert durch die Zahl der Stichprobenelemente. Wie bereits weiter oben angedeutet, sieht die Realität freilich anders aus. Bis tatsächlich die erforderliche Zahl von Interviews vorliegt, können vielerlei Einflüsse die Stichprobe verzerrt haben. Die wesentlichsten Fehlerquellen sind:

- Die Verteilung der realisierten Sample-Points spiegelt regional nicht die wahre Haushalts- und Personenverteilung wider. **Abhilfe:** Anpassung der Gewichte für Sample-Points an die Regionalverteilung.
- Zwar mögen die Sample-Points gleichmäßig über die DDR verteilt gewesen sein; wenn die Ausschöpfung dagegen ungleichmäßig ausfiel, werden Bezirke ungleich stark repräsentiert. **Abhilfe:** Anpassung der Gewichte für Sample-Points um eben jenes Verhältnis (Ausgleich zwischen Brutto- und Nettostichprobe).
- Die Strukturen in den einzelnen Sample-Points zeigen stärkere Klumpungseffekte als erwartet, Haushalte mit speziellen demographischen Eigenschaften werden dadurch über- oder unterrepräsentiert. **Abhilfe:** Berücksichtigung von demographischen Merkmalsverteilungen.

Für die Gewichtung der Basiserhebung des SOEP in der alten Bundesrepublik sind diese Schritte durchgeführt worden. Dabei zeigte sich, daß der erwähnte Ausgleich zwischen Brutto- und Nettostichprobe allein schon zu einer deutlichen Erhöhung der Varianz in den Hochrechnungsfaktoren führte. Da durch eine sol-

che Anpassung die Stichprobe vornehmlich nur regional entzerzt wird, regionale Verteilungen aber an anderer Stelle berücksichtigt werden, wurde für die DDR-Stichproben zugunsten geringerer Stichprobenvarianz auf diesen Schritt verzichtet.

Bei der Gewichtung von SOEP-Ost und WfS-Ost wurde folgendermaßen vorgegangen:

Zunächst wurde ein Hochrechnungsrahmen für die Basis-Erhebung des SOEP-Ost entwickelt. Dies ergab sich einerseits aus dem zeitlichen Vorsprung des SOEP-Ost gegenüber dem Wohlfahrtssurvey. Auch ist die Stichprobe des SOEP-Ost - bezogen auf die befragten Personen - etwa um den Faktor Sechs größer als die des Wohlfahrtssurveys und kann deshalb genauer hochgerechnet werden, indem mehr Restriktionen, d.h. detailliertere Randverteilungen, berücksichtigt werden als beim WfS.

Es werden für die Basisbefragung des SOEP-Ost zwei Gewichtsvektoren für Haushalts- und Personengewichte benötigt. Diese sollen zumindest folgende Bedingungen erfüllen:

1. Da Personen und Haushalte dieselbe Auswahlwahrscheinlichkeit besitzen, müssen die Gewichte aller Personen eines Haushalts mit dem Haushaltsgewicht übereinstimmen¹¹.
2. Die Summe aller Hochrechnungsfaktoren für die Haushalte muß die Gesamtsumme aller Haushalte in der DDR, diejenige der Personen die Wohnbevölkerung in der DDR ergeben.
3. Die demographische Struktur der DDR soll möglichst gut auf die Stichprobe übertragen werden.
4. Die Varianz der Gewichte soll unter den Voraussetzungen 1 bis 3 minimal sein.

Es wurde versucht, diese Bedingungen durch folgendes Vorgehen zu entsprechen: Da das Design der Stichprobe bereits repräsentativ angelegt wurde, wäre im Idealfalle keine Designgewichtung erforderlich. Deshalb bietet es sich an, von einem Gewichtungsvektor für Haushalte auszugehen, der das Ergebnis einer freien Hochrechnung ist. Seine Varianz ist zunächst gleich Null. An-

¹¹ Beim SOEP werden alle Haushaltsmitglieder befragt, die 16 Jahre und älter sind. Die Verweigerung nur einer Befragungsperson führt zum Totalausfall des Haushalts. Daher weisen a posteriori Personen und Haushalte dieselbe Auswahlwahrscheinlichkeit auf.

schließlich wird der Vektor nach dem Prinzip der Minimierung des Informationsverlusts, also möglichst geringfügig korrigiert, so daß vorgegebene Randbedingungen erfüllt werden. Diese Randbedingungen werden im folgenden Abschnitt zum Hochrechnungsrahmen näher beschrieben.

4 Die Gewichtung der DDR-Basisbefragung

4.1 Der Hochrechnungsrahmen

Jede Hochrechnung, jede Gewichtung kann auch bei bester methodischer Grundlage nur so gut sein wie der ihr zugrundegelegte Hochrechnungsrahmen, d. h. so gut wie die Informationen über die anzustrebenden Randverteilungen und Ecksummen.

Seit Mitte 1989 verließen hunderttausende Bürger der DDR ihr Land, um sich in der Bundesrepublik eine neue Existenz aufzubauen. Die letzten zuverlässigen Angaben zur Demographie der Wohnbevölkerung datieren vom Jahresende 1989. Danach hatte die DDR zum 31.12.1989 eine Bevölkerungsbestand von 16,43 Mill. Personen. Die Wanderungen rissen danach nicht ab, auch nach der Befragung im Juni 1990 siedelten noch viele tausend DDR-Bürger in die Bundesrepublik über.¹² Es erhebt sich die Frage, warum unter diesen Umständen überhaupt an Eckwerte angepaßt wird, die ohnehin überholt oder gar falsch sind. Im wesentlichen sind zwei Gründe zu nennen:

1. Der Zeitpunkt der DDR-Basiserhebung differiert zeitlich um ungefähr ein halbes Jahr von der aktuellen amtlichen Statistik der DDR zum Jahresende 1989. Im Normalfall, d.h. bei einer in ruhigen Bahnen verlaufenden Bevölkerungsbewegung, wären diese sechs Monate Differenz nahezu unerheblich. Die starken Übersiedlungen im ersten Halbjahr 1990 ergaben indes

12 Das statistische Amt der DDR geht von ca. 200 Tausend Personen bis Ende Juni 1990 aus, berücksichtigt aber nur ordnungsgemäß abgemeldete Personen. Dagegen spricht das Statistische Bundesamt Wiesbaden von ca. 300 Tausend Bürgern; hier dürften aber Doppelzählungen und Rückwanderungen nicht berücksichtigt sein. Die genannten Zahlen können deshalb als Unter- und Obergrenze der Wanderung im 1. Halbjahr 1990 interpretiert werden.

einen Wanderungsverlust zwischen einem und zwei vH, die sich nicht proportional auf die Bevölkerungsgruppen verteilen. Die Bevölkerungsstruktur änderte sich in Richtung eines größeren Anteils älterer Menschen; freilich liegen keine exakten Informationen darüber vor. Dennoch ist aus Gründen der Vergleichbarkeit mit anderen Analysen und Untersuchungen der Bezug auf offizielle Zahlen unverzichtbar, auch wenn diese Daten überholt sind. Es bleibt dem Anwender immer noch überlassen, auf Basis der Hochrechnung Aktualisierungen der Rahmen-daten vorzunehmen.

2. Wie weiter unten noch an Beispielen gezeigt werden kann, werden auch in echten Zufallsstichproben Subpopulationen über- oder untererfaßt. Diese Verzerrungen sind teilweise erheblich stärker ausgefallen als die Veränderungen, die durch die turbulente demographische Entwicklung in den ersten sechs Monaten von 1990 induziert wurden.

Der Hochrechnungsrahmen für die Basiserhebung des SOEP-Ost gliedert sich in drei Teile:

- Verteilung der Wohnbevölkerung nach Geschlecht und Bezirken,
- Verteilung der Wohnbevölkerung nach Geschlecht, Alter und Familienstand,
- Gesamtzahl der Haushalte in der DDR.

Die regionale Gewichtung, die nach Bezirken vorgenommen wird, weist einige Besonderheiten auf, die erklärungsbedürftig sind:

Im Gegensatz zur Bundesrepublik war die DDR politisch in 15 Bezirke geteilt und nicht nach Ländern gegliedert. Die Bezirke wurden erst mit der Vereinigung im Oktober 1990 durch folgende Zusammenlegungen wieder in Länder zurückgeführt:

Länder	Bezirke
Mecklenburg-Vorpommern	Rostock, Schwerin, Neubrandenburg
Brandenburg	Potsdam, Frankfurt/Oder, Cottbus
Sachsen-Anhalt	Magdeburg, Halle
Thüringen	Erfurt, Gera, Suhl
Sachsen	Dresden, Leipzig, Chemnitz
nachrichtlich:	
Berlin	Berlin(West) und Berlin(Ost)

Allerdings wurde die Basisbefragung am 30.6.90 zu einem Zeitpunkt abgeschlossen, als die endgültige Zuordnung einiger Kreise

zu den neuen Ländern wegen anstehender Volksentscheidungen noch nicht erfolgt war. Die dunklen Flächen in Abbildung 1 zeigen die Regionen, die sich für eine Eingliederung in ein anderes Bundesland entschieden haben und deshalb nicht den oben aufgeführten Zuordnungen entsprechen.

Dieser Wechsel der Zuordnungen schlug sich auch in der Stichprobe nieder: Von 2.179 Netto-Haushalten gibt es in 71 Fällen (3,3 vH) Inkonsistenzen zwischen Bezirks- und Länderabgrenzungen. Solche Differenzen konnten in der amtlichen Statistik der DDR zum 13.12.89 definitionsgemäß nicht vorkommen. Aus diesem Grunde war es erforderlich, die regionale Anpassung nach Bezirken und nicht nach den neuen Ländern vorzunehmen, zumal die Bezirksstruktur auch wesentlich feiner als die Länderstruktur ist.

Für die 15 Bezirke der DDR weist das Statistische Amt der DDR für Ende 1989 die Wohnbevölkerung nach Geschlecht aus. Hieraus ergeben sich $2 \cdot (15-1) = 28$ Restriktionen¹³.

Tabelle 1 ist die vorgegebene Regionalverteilung zu entnehmen:

Für jeden Bezirk - getrennt nach dem Geschlecht - sind folgende Merkmale aufgelistet: Zahl der Personen am 31.12.1989, Anzahl der Personen in der Stichprobe und das Verhältnis aus beiden Größen (= inverse Auswahlwahrscheinlichkeit).

Insgesamt wies die Statistik Ende 1989 eine Wohnbevölkerung von 16.433.796 Personen aus. Einschließlich der Kinder umfaßt das SOEP-Ost 6.044 Personen. Damit ergibt sich eine durchschnittliche Auswahlwahrscheinlichkeit von 0,03678 vH, d.h. eine Person im SOEP-Ost "repräsentiert" 2.719 Personen der Gesamtbevölkerung. Dieser Wert wäre der Faktor für eine freie Hochrechnung. Getrennt nach den Geschlechtern ergeben sich die in der Tabelle ausgewiesenen Werte. Sie weichen nur geringfügig

13 Da im folgenden Restriktionsblock die Gesamtzahl der Männer und Frauen in der DDR definiert wird, ist die Verteilung nach Bezirken und Geschlecht schon durch 14 Bezirke hinreichend definiert.

Abb 1: Länder- und Bezirksgrenzen im Gebiet der DDR 1990



Tabelle 1

Wohnbevölkerung in der DDR und im SOEP -Ost - nach Bezirken und Geschlecht -						
Bezirke	Männer			Frauen		
	Per- sonen DDR	Stich- probe SOEP	DDR/ SOEP	Per- sonen DDR	Stich- probe SOEP	DDR/ SOEP
Rostock	444731	162	2745	465109	169	2752
Schwerin	285808	123	2324	304363	102	2984
Neubrandenburg	301168	110	2738	314599	109	2886
Potsdam	537263	186	2889	573947	188	3053
Frankfurt/Oder	344944	118	2923	361172	126	2866
Cottbus	426093	181	2354	449488	199	2259
Magdeburg	594518	284	2093	643389	303	2123
Halle	836539	251	3333	911491	281	3244
Erfurt	586458	187	3136	636439	187	3403
Gera	347943	135	2577	380136	144	2640
Suhl	262320	138	1901	282951	155	1825
Dresden	812012	281	2890	901074	313	2879
Leipzig	631097	259	2437	702045	270	2600
Chemnitz	853468	326	2618	964019	338	2852
Berlin(Ost)	611244	208	2928	670274	211	3177
Insgesamt	7873300	2949	2670	8560496	3095	2766
Quellen: Das Sozio-ökonomische Panel (Ost), Statistisches Amt der DDR.						

von dem Durchschnittswert ab. Die Männer sind mit einem Faktor von 2.670¹⁴ um 1,8 vH überrepräsentiert, die Frauen mit einem Faktor von 2.766 um etwa 1,7 vH untererfaßt¹⁵.

Regional gibt es weitaus größere Abweichungen. Zu 20 vH am stärksten unterrepräsentiert sind die Frauen im Bezirk Erfurt

14 Man könnte diesen Wert auch als "empirischen Hochrechnungsfaktor" bezeichnen, da die Multiplikation der Stichprobenelemente mit diesem Faktor ex definitione die vorgegebene Randsumme ergibt.

15 Generell gilt: Liegen die Werte in den Spalten "DDR/SOEP" über dem Durchschnittswert von 2.719, liegt Untererfassung vor, da ein Stichprobenelement mehr Personen zu repräsentieren hat. Werte kleiner als 2.719 zeugen von einer überproportionalen Ausschöpfung der jeweiligen Merkmalskombination.

(HF=3.403). Mit fast 50 vH über dem Soll wurden im Bezirk Suhl die Frauen übererfaßt (HF=1.825).

Zu den typischen Randverteilungen, an die Stichproben angepaßt werden, zählt die gemeinsame Verteilung der Personen nach **Alter, Geschlecht und Familienstand**.

Tabelle 2 zeigt die demographischen Restriktionen, die für die Gewichtung berücksichtigt wurden. Die Ergebnisse vermitteln folgendes Bild: Wie schon erwähnt, sind die Frauen um ca. 1,7 vH untererfaßt¹⁶, eine Abweichung, die angesichts der erwähnten stürmischen Entwicklung nicht sehr ins Gewicht fällt. Anders ist die Lage zu beurteilen, wenn zur Bewertung der Repräsentativität die Verteilung nach dem Familienstand herangezogen wird. Hier heben sich die Abweichungen von den vorgegebenen Randverteilungen deutlicher hervor. So sind Verheiratete um ca. 4 vH überrepräsentiert, für diese Gruppe vermutlich kein schlechtes Ergebnis. Auch für die Ledigen errechnet sich insgesamt nur eine relativ geringe Abweichung vom Erwartungswert (ca. +6 vH). Dieses Ergebnis täuscht aber über große strukturelle Abweichungen innerhalb dieser Gruppe hinweg: Die jüngeren Ledigen, die überwiegend noch bei ihren Eltern leben, sind mit 19,4 vH stark überrepräsentiert, die ältere Gruppe mit -20,5 vH entsprechend untererfaßt. Die im Saldo gemessene Übererfassung der Ledigen und Verheirateten hat eine entsprechende Untererfassung der Verwitweten und der Geschiedenen zur Konsequenz. Letztere Subpopulation umfaßt mit 2,26 Mill. Personen etwa ein Siebtel der Wohnbevölkerung. Somit errechnet sich eine Minderausschöpfung von nahezu einem Drittel. Statt der zu erwartenden gut 830 verwitweten oder geschiedenen Personen, befinden sich im SOEP-Ost nur 560 Frauen und Männer in dieser Gruppe. Diese Untererfassung

16 Diese und die weiteren prozentualen Abweichungen ergeben sich durch Vergleich der Ist- und Sollzahlen in der Stichprobe: Insgesamt wurden 3.095 Frauen erhoben. Aufgrund der oben erwähnten Auswahlwahrscheinlichkeit von 0,03678 vH wären im SOEP-Ost 3.149 Frauen zu erwarten gewesen. Somit sind 54 von 3.149, also ca. 1,7 vH Frauen, zu wenig erfaßt worden.

Tabelle 2

**Wohnbevölkerung in der DDR und im SOEP -Ost
- nach Familienstand, Alter und Geschlecht -**

Altersklassen	Männer			Frauen		
	Per- sonen DDR	Stich- probe SOEP	DDR/ SOEP	Per- sonen DDR	Stich- probe SOEP	DDR/ SOEP
Männer und Frauen insgesamt						
0 -4 J.	545191	229	2381	517254	188	2751
5 -9 J.	577826	272	2124	551517	246	2242
10 -14 J.	518660	264	1965	491694	240	2049
15 -19 J.	502511	195	2577	478014	199	2402
20 -24 J.	633170	194	3264	596900	190	3142
25 -29 J.	719458	247	2913	672170	274	2453
30 -34 J.	649482	270	2405	611288	285	2145
35 -39 J.	631305	247	2556	601112	254	2367
40 -44 J.	414361	197	2103	404597	209	1936
45 -49 J.	579616	203	2855	582704	186	3133
50 -54 J.	598589	216	2771	609967	214	2850
55 -59 J.	456572	158	2890	482116	155	3110
60 -64 J.	353307	100	3533	474354	134	3540
65 -69 J.	250538	68	3684	468452	120	3904
70 -74 J.	126770	34	3729	261042	70	3729
75 -79 J.	162990	33	4939	369108	69	5349
80 -99 J.	152954	22	6952	388207	62	6261
Insgesamt	7873300	2949	2670	8560496	3095	2766
Ledige Männer und Frauen*						
0 -4 J.	545191	229	2381	517254	188	2751
5 -9 J.	577826	272	2124	551517	246	2242
10 -14 J.	518660	264	1965	491694	240	2049
15 -19 J.	500938	194	2582	467691	199	2350
20 -24 J.	510339	161	3170	344060	109	3157
25 -29 J.	283954	79	3594	137160	58	2365
30 -34 J.	117518	33	3561	58982	19	3104
35 -44 J.	100277	27	3713	56829	13	4371
45 -54 J.	59509	11	5410	53287	7	7612
55 -99 J.	29264	6	4877	171088	39	4387
Insgesamt	3243476	1276	2542	2849562	1118	2549
nachrichtlich:						
0 -19 J.	2142615	959	2234	2028156	873	2323
20 -99 J.	1100861	317	3473	821406	245	3353

* Ledige Männer und Frauen wurden nicht direkt in den Hochrechnungsrahmen einbezogen. Sie sind hier nur der Vollständigkeit halber aufgeführt.

noch Tabelle 2

Verheiratete Männer und Frauen

Altersklassen	Männer			Frauen		
	Per- sonen DDR	Stich- probe SOEP	DDR/ SOEP	Per- sonen DDR	Stich- probe SOEP	DDR/ SOEP
0 -24 J.	116913	34	3439	244999	75	3267
25 -29 J.	392229	156	2514	475388	202	2353
30 -34 J.	467143	218	2143	478827	239	2003
35 -39 J.	490313	215	2281	480438	218	2204
40 -44 J.	336099	173	1943	325393	182	1788
45 -49 J.	421230	184	2615	467694	162	2887
50 -54 J.	513394	198	2593	478856	173	2768
55 -59 J.	402822	145	2778	352096	110	3201
60 -64 J.	315007	87	3621	299208	97	3085
65 -69 J.	219497	61	3598	222942	57	3911
70 -74 J.	105647	28	3773	87641	24	3652
75 -99 J.	201270	36	5591	127947	16	7997
Insgesamt	4041564	1535	2633	4041429	1555	2599

Verwitwete Männer und Frauen

0 -49 J.	10678	5	2136	39826	13	3064
50 -59 J.	24339	6	4057	98730	25	3949
60 -64 J.	17739	9	1971	99229	17	5837
65 -69 J.	20406	6	3401	166902	43	3881
70 -74 J.	16410	5	3282	136663	38	3596
75 -79 J.	37412	3	5776	238233	53	4495
80 -99 J.	66192	13	6637	297646	50	5953
Insgesamt	193176	47	4110	1077229	239	4507

Geschiedene Männer und Frauen

0 -29 J.	50307	12	4192	75586	19	3978
30 -34 J.	63895	18	3550	69871	27	2588
35 -39 J.	69267	14	4948	76088	26	2926
40 -44 J.	46025	13	3540	52792	16	3299
45 -49 J.	59551	10	5955	71213	19	3748
50 -54 J.	48678	10	4868	62413	26	2401
55 -99 J.	57361	14	4097	184313	50	3686
Insgesamt	395084	91	4342	592276	183	3236

nachrichtlich nur für Frauen:*

55 -64 J.	.	.	.	80822	31	2607
65 -99 J.	.	.	.	103491	19	5447

Quellen: Das Sozio-ökonomische Panel (Ost), Statistisches Amt der DDR.

* Diese Unterteilung wurde nur für die Frauen vorgenommen, da die Zellenbesetzung bei den Männern in diesen Altersklassen zu gering war.

ist nicht - wie vielleicht zu erwarten - allein auf das höhere Durchschnittsalter jenes Bevölkerungskreises zurückzuführen, zieht sich die Minderausschöpfung doch durch fast alle Altersklassen.

Die Alterstruktur der Stichprobe zeigt das erwartete Bild: Deutlich unterrepräsentiert sind - wie bei fast allen Erhebungen - vornehmlich ältere Bürger. Der Bruch zeigt sich etwa ab dem 60. Lebensjahr, ein Alter ab dem viele Menschen in Rente gehen, gesundheitlich nicht mehr auf voller Höhe sind und allmählich weniger aktiv am Leben teilnehmen. Deutlich erkennbar zeigt sich der steile Anstieg des empirischen Hochrechnungsfaktors. Zu niedrige Ausgeschöpfungsquoten finden sich auch bei den jungen Menschen um Zwanzig, die erfahrungsgemäß überdurchschnittlich oft verweigern. Dagegen sind Bürger zwischen 30 und 45 überrepräsentiert.

Insgesamt steuerten diese Verteilungen mit 86 Eckwerten den größten Teil zur Gesamtzahl der Restriktionen bei.

Die Datenbasis für die Gesamtzahl der Haushalte und ihre Struktur für das Jahr 1990 ist nur wenig abgesichert. Vornehmlich der Anteil der Ein-Personenhaushalte bleibt ungewiß.

Die amtliche Statistik bietet lediglich Ergebnisse der letzten Volkszählung aus dem Jahre 1981 sowie eine "amtliche Fortschreibung" an.¹⁷ Allein Wohnungsstatistiken jüngeren Datums ergänzen dieses Material.¹⁸ Danach schwanken die Anteile der Ein-Personenhaushalte an allen Privathaushalten zwischen 25,1 vH (Volkszählung 1981) und 29,1 vH (amtliche Fortschreibung 1988). Letzlich sind die amtlichen Zahlen aus der DDR schon vor Beginn der Übersiedlerwelle seit Herbst 1989 nicht sehr verlässlich. Die Auswirkungen der starken Wanderungsbewegungen auf die Haushaltsstrukturen sind schon gar nicht quantifizierbar. Gleichwohl ist

17 Bedau und Vortmann (1990), S.656.

18 Vgl. Bartholmai, Melzer und Schulz (1990).

der Anteil von 15,6 vH der Ein-Personenhaushalte im ungewichteten SOEP mit Sicherheit zu gering.

Für die Hochrechnung besteht allerdings auch das Problem, daß es zuverlässige Zahlen über die Haushaltsstruktur in der DDR nicht gibt. Deshalb soll die Stichprobe nur an die vom DIW geschätzte Ecksumme von 6.794 Tausend Haushalten angepaßt werden¹⁹. Es war Tabelle 2 zu entnehmen, daß die Gruppe der Ledigen, die 20 Jahre und älter sind, stark unterrepräsentiert ist. Dies ist eine Subpopulation, die erfahrungsgemäß häufig alleine wohnt. Damit ist die Wahrscheinlichkeit groß, daß durch eine Hochrechnung, die Geschlecht, Familienstand und Alter berücksichtigt, auch der Anteil der Ein-Personenhaushalte steigt. Unter Umständen erübrigt sich durch diese Kollinearität eine Korrektur der Haushaltsstruktur.

Mit der Festlegung der Gesamtzahl der Haushalte erhöht sich die Zahl der Restriktionen um eine auf insgesamt 115. Sollte künftig eine bessere Stützung der Haushaltszahl vorliegen, kann diese durch eine einfache skalare Multiplikation berücksichtigt werden. Sollten weiterhin noch zuverlässige Haushaltsstrukturinformationen bekannt werden, so würden diese ggfs. eine neue Hochrechnung erfordern, wenn die neue Struktur signifikant von der alten abweichen würde²⁰.

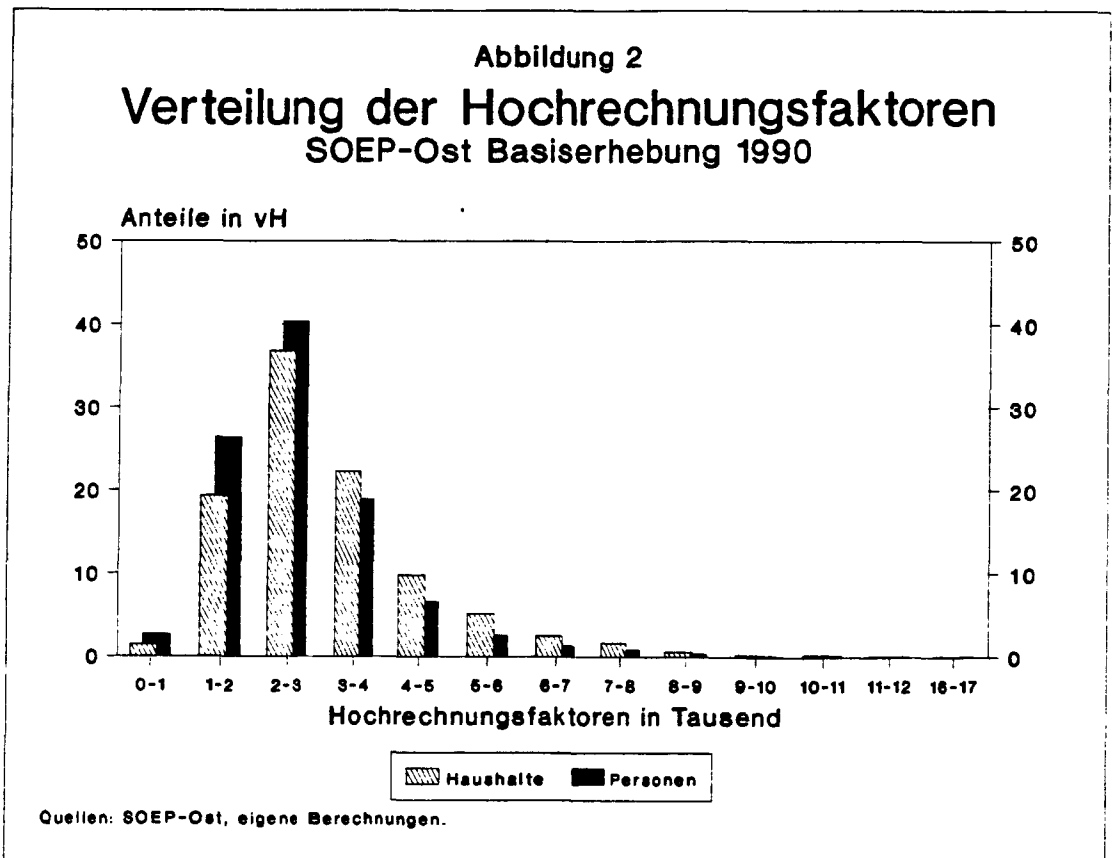
Nach Bestimmung der Haushaltsgewichte werden diese auf sämtliche Personen einschließlich der Kinder unter 16 Jahren der zugehörigen Haushalte übertragen. Auf diese Weise ist die Identität von Haushalts- und Personengewichte immer gewährleistet.

19 Bedau und Vortmann (1990).

20 Unter Umständen wird man anhand einer Hochrechnung für die 2. oder gar 3. Welle eine Rückrechnung auf die Basiserhebung vornehmen können. Dies setzt allerdings parallele Mikrozensusergebnisse für die neuen Bundesländer und den Ostteil Berlins voraus.

4.2 Ergebnisse der Gewichtung

Die Auswirkungen der Gewichtung werden an einigen Beispielen demonstriert, indem regionale und demographische Verteilungen vor und nach Gewichtung gezeigt und knapp kommentiert werden. Zunächst jedoch werden einige statistische Eigenschaften der Gewichte betrachtet.



4.2.1 Statistische Eigenschaften der Hochrechnungsfaktoren

Definitionsgemäß sind für jeden Haushalt die Personen und Haushaltsgewichte identisch. Unterschiedliche Haushaltsgrößen führen indes zu leicht differierenden Verteilungen. Sie sind in Abbildung 2 dargestellt.

Der niedrigste Wert für einen Hochrechnungsfaktor wurde mit 364 ermittelt. Dies heißt, daß die betreffenden Personen und Haushalte nur mit einem Achtel des Durchschnittsgewicht bei Analysen und Auswertungen berücksichtigt werden. Nur 31 Haushalte (1,4 vH) und 157 Personen (2,6 vH) repräsentieren weniger als 1.000 Mitbürger und besitzen somit nicht mehr als ein Drittel des Durchschnittsgewichts.

Auswertungen von Teilpopulationen bergen bei allzu großer Varianz der Hochrechnungsfaktoren verstärkt die Gefahr großer zufälliger Abweichungen, die zu Fehlinterpretationen verleiten können. Dies gilt insbesondere, wenn einzelne Personen und/oder Haushalte außerordentlich hohe Gewichte zugewiesen bekommen. Diese treten in der Basisbefragung DDR erfreulicherweise kaum auf. Nur ein einziger Ein-Personenhaushalt wird mit dem Gewicht von knapp 16.800 bewertet und übertrifft den Durchschnittswert um das 5,5-fache (Personengewicht: 6,2-fach). Sechs Haushalte (0,3 vH) mit insgesamt acht Personen (0,1 vH) repräsentieren mehr als 10.000 Haushalte bzw. Personen.

So drängt sich der weitaus größte Teil der Gewichte um ihren Mittelwert. Fast 60 vH liegen im Bereich von 2.000 bis 4.000; mehr als 95 vH fallen in das Intervall 1.000 bis 7.000. Dies wirkt sich auch in den Standardabweichungen aus: Sie beträgt für die Haushalte 1.511 und für die Personen 1.277.

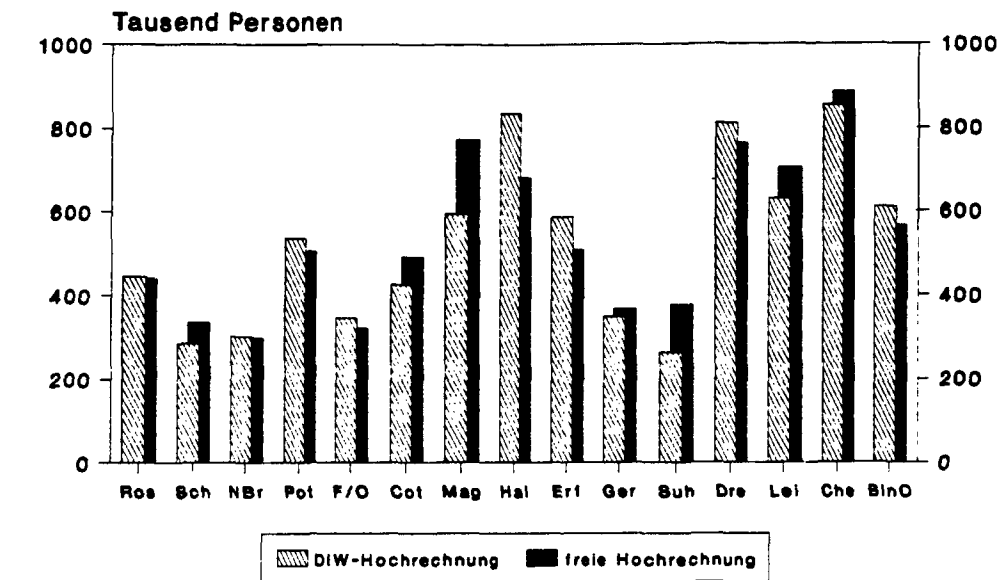
Die wichtigsten statistischen Kennziffern der Hochrechnungsfaktoren auf einen Blick:

Kennziffer	Haushalte	Personen
Anzahl	2.179	6.044
Summe	6.793.998	16.433.794
Arithmetisches Mittel	3.117,9	2.719,0
Standardabweichung	1.510,9	1.277,1
Variationskoeffizient	0,48	0,47
Minimum	363,9	363,9
Maximum	16.798,4	16.798,4

Abbildung 3a

Männer in der DDR nach Bezirken

Hochrechnung: SOEP-Ost Basiserhebung '90

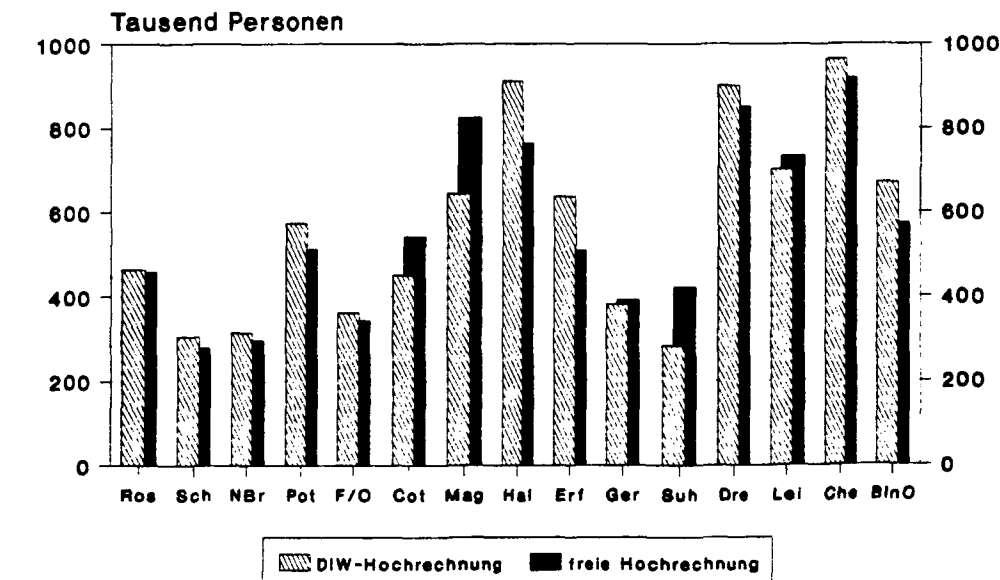


Quellen: SOEP-Ost, eigene Berechnungen, Statistisches Amt der DDR.

Abbildung 3b

Frauen in der DDR nach Bezirken

Hochrechnung: SOEP-Ost Basiserhebung '90



Quellen: SOEP-Ost, eigene Berechnungen, Statistisches Amt der DDR.

4.2.2 Die Regionalverteilung

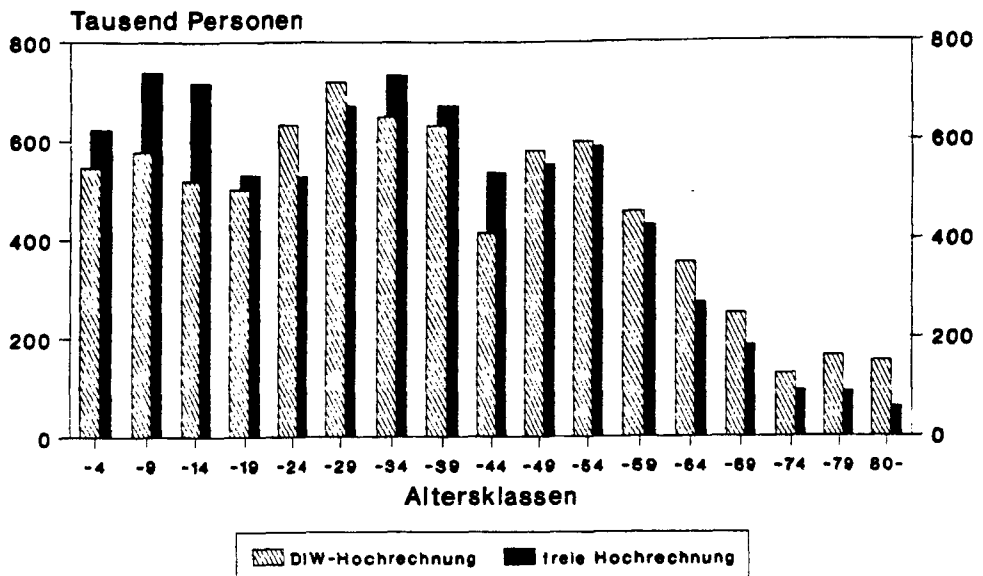
In diesem und im folgenden Abschnitt werden die Ergebnisse der Hochrechnung, die mit den vorgegebenen Randverteilungen übereinstimmen, denen einer freien Hochrechnung gegenübergestellt. Freie Hochrechnung heißt für die Personendaten, daß die Zahl der Stichprobenelemente mit dem arithmetischem Mittel der Personengewichte, also mit 2.719,0, multipliziert werden. Im wesentlichen werden die im Kapitel "Hochrechnungsrahmen" bereits diskutierten Eigenschaften graphisch verdeutlicht (Abbildungen 3a und 3b). Ersichtlich ist die demographische Verteilung durch die freie Hochrechnung (schwarze Flächen) gut wiedergegeben. Größere Abweichungen von der "wahren" Verteilung (schraffierte Flächen) zeigen sich nur in den Bezirken Magdeburg, Suhl, Halle und Cottbus.

4.2.3 Die demographische Verteilung

Die Abbildungen 4a und 4b zeigen die Verzerrungen in der Altersstruktur, wenn frei hochgerechnet wird. Auffällig ist der Geburtenausfall nach dem 2. Weltkrieg, der sich 1989/90 in der Klasse der 40-44 jährigen zeigt. Er wird in der Männer-Stichprobe zu schwach, bei den Frauen kaum nachgezeichnet. Besonders hebt die Graphik für die Frauen die unterschiedlichen Ausschöpfungsraten der älteren Bürger gegenüber denen der 30 bis 45-jährigen hervor.

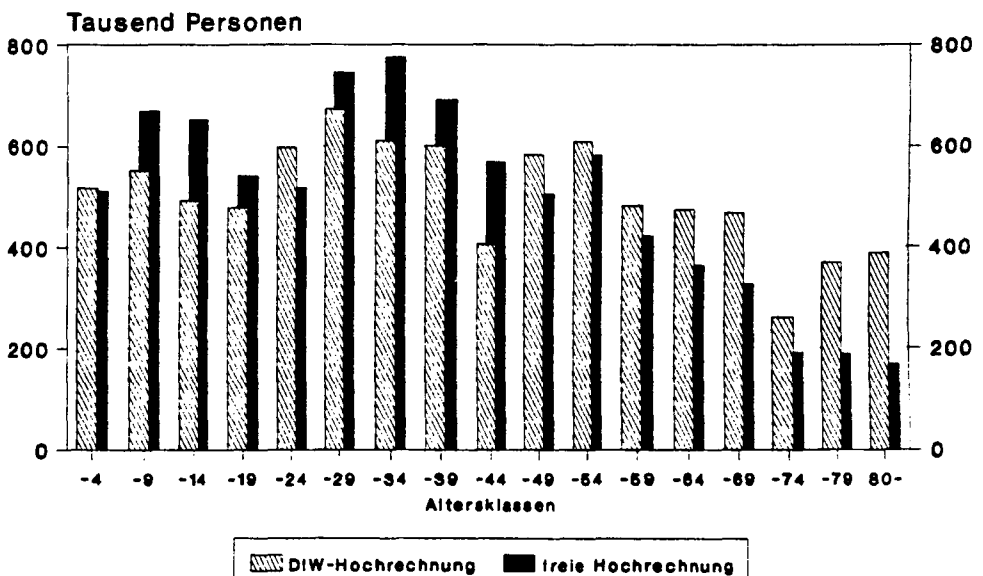
In den weiteren Abbildungen sind Aufgliederungen nach dem Familienstand vorgenommen werden. Auch diese Graphiken erklären sich von selbst. Erwähnt werden soll lediglich noch Abbildung 7b, die die verwitweten Frauen betrifft. Hier sieht man die generelle und die mit dem Alter zunehmende Untererfassung dieser Subpopulation.

Abbildung 4a
Männer in der DDR nach Alter
 Hochrechnung: SOEP-Ost Basiserhebung '90



Quellen: SOEP-Ost, eigene Berechnungen,
 Statistisches Amt der DDR.

Abbildung 4b
Frauen in der DDR nach Alter
 Hochrechnung: SOEP-Ost Basiserhebung '90

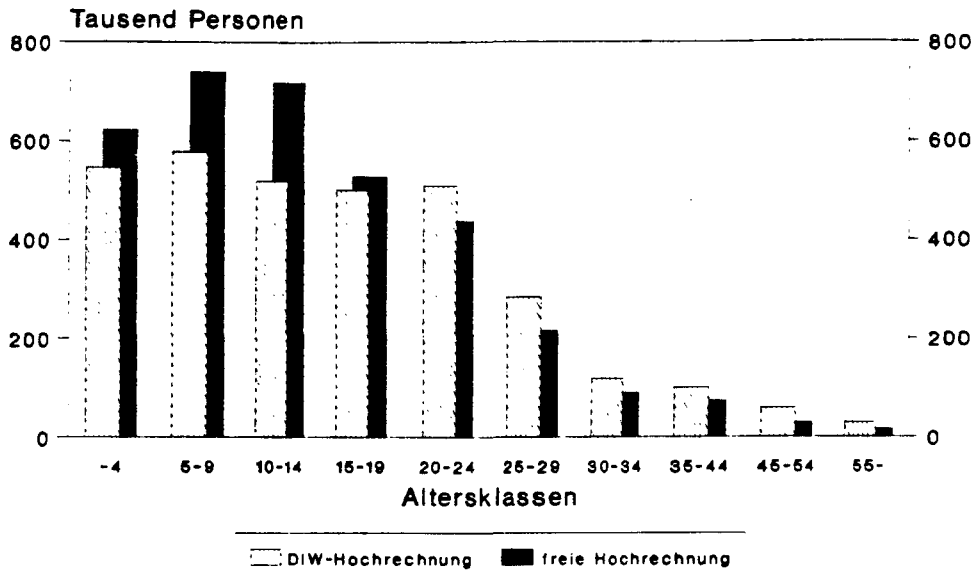


Quellen: SOEP-Ost, eigene Berechnungen,
 Statistisches Amt der DDR.

Abbildung 5a

Ledige Männer in der DDR nach Alter

Hochrechnung: SOEP-Ost Basiserhebung '90

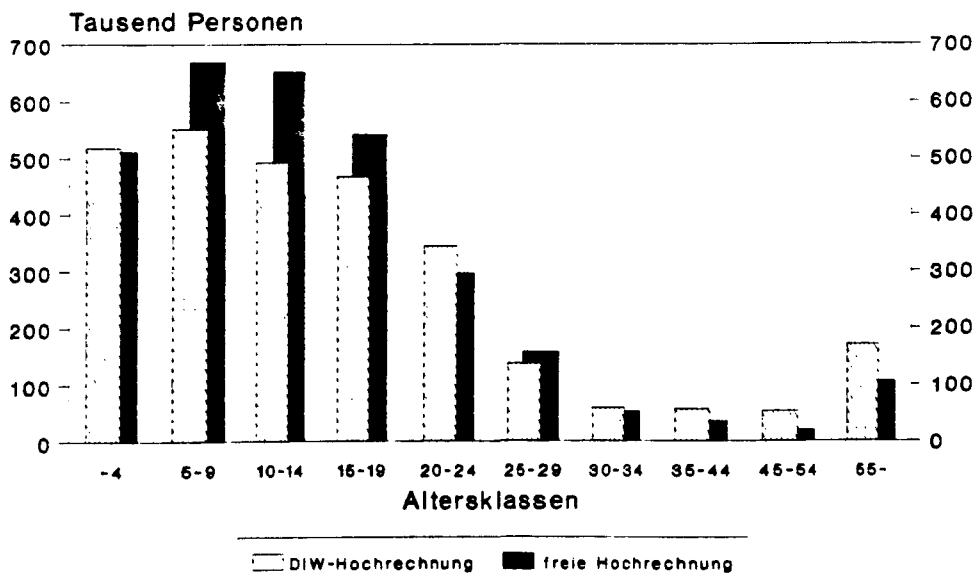


Quellen: SOEP-Ost, eigene Berechnungen,
Statistisches Amt der DDR.

Abbildung 5b

Ledige Frauen in der DDR nach Alter

Hochrechnung: SOEP-Ost Basiserhebung '90

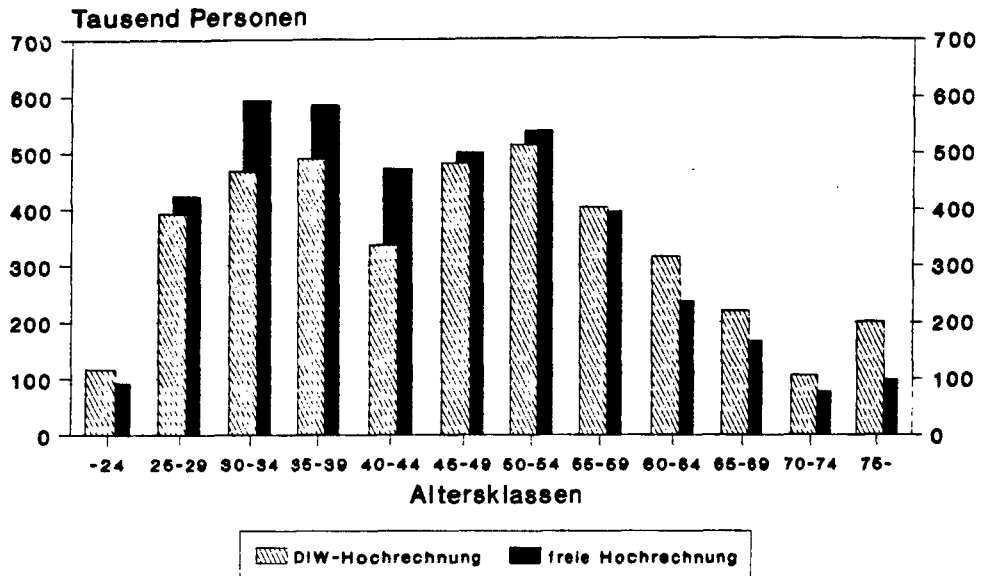


Quellen: SOEP-Ost, eigene Berechnungen,
Statistisches Amt der DDR.

Abbildung 6a

Verheiratete Männer nach Alter

Hochrechnung: SOEP-Ost Basiserhebung '90

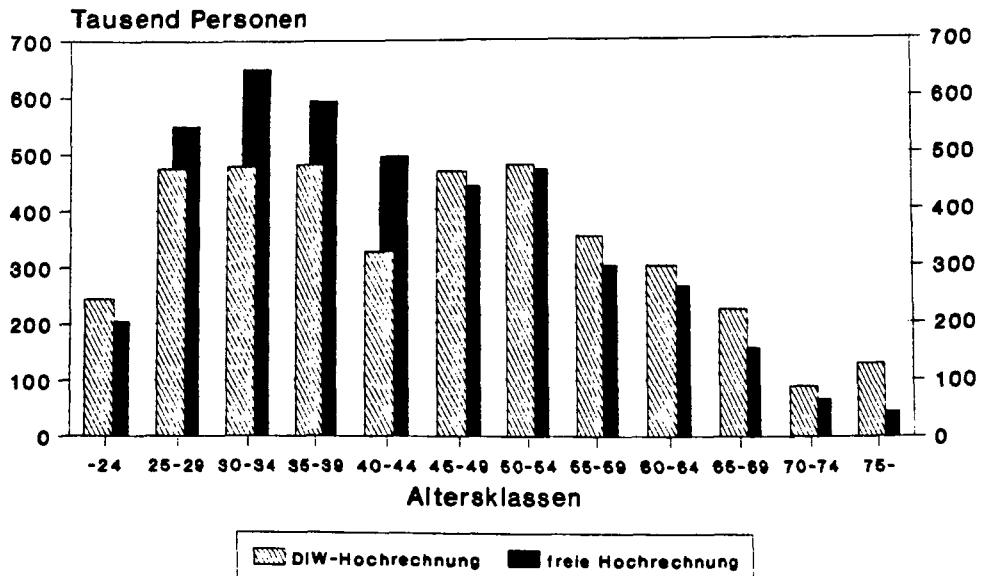


Quellen: SOEP-Ost, eigene Berechnungen,
Statistisches Amt der DDR.

Abbildung 6b

Verheiratete Frauen nach Alter

Hochrechnung: SOEP-Ost Basiserhebung '90

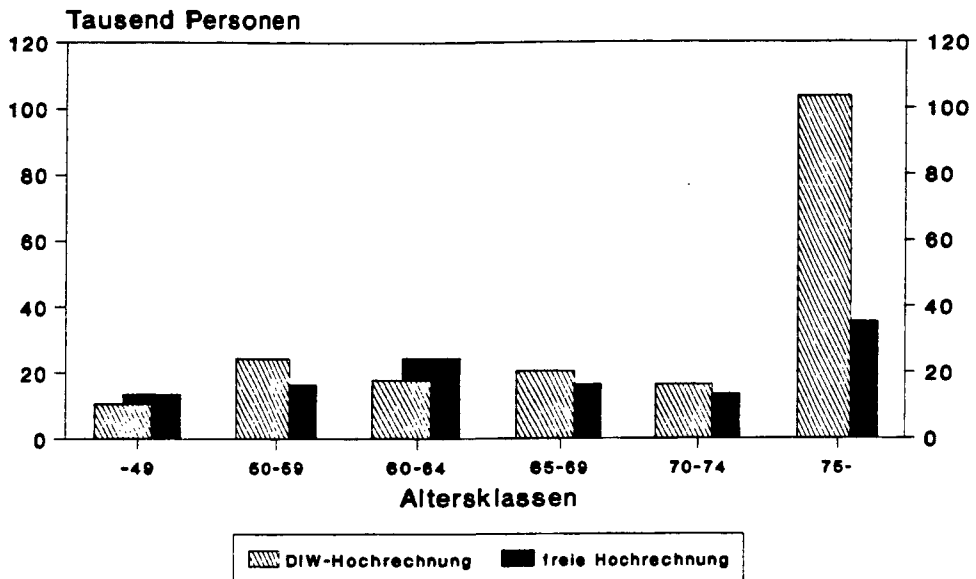


Quellen: SOEP-Ost, eigene Berechnungen,
Statistisches Amt der DDR.

Abbildung 7a

Verwitwete Männer in der DDR nach Alter

Hochrechnung: SOEP-Ost Basiserhebung '90

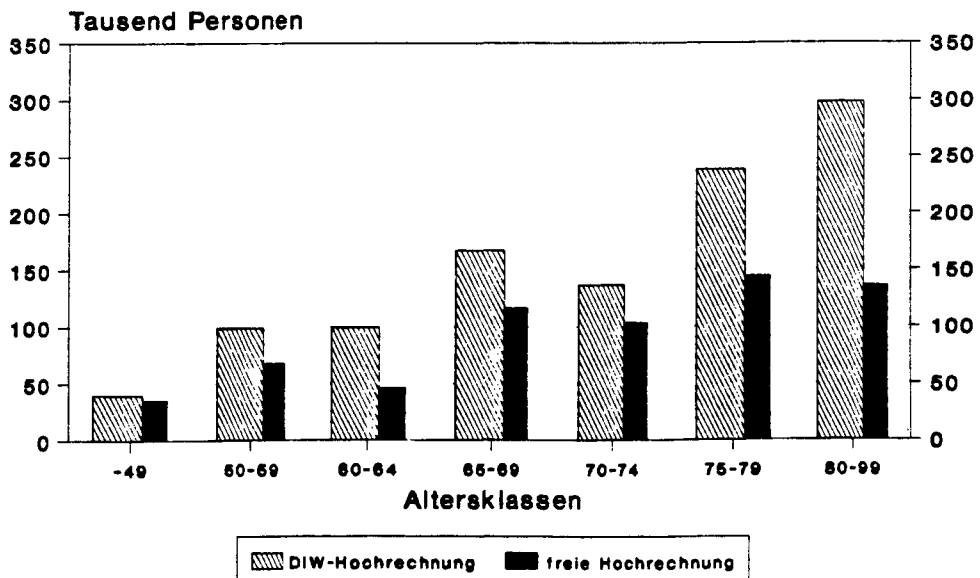


Quellen: SOEP-Ost, eigene Berechnungen,
Statistisches Amt der DDR.

Abbildung 7b

Verwitwete Frauen in der DDR nach Alter

Hochrechnung: SOEP-Ost Basiserhebung '90

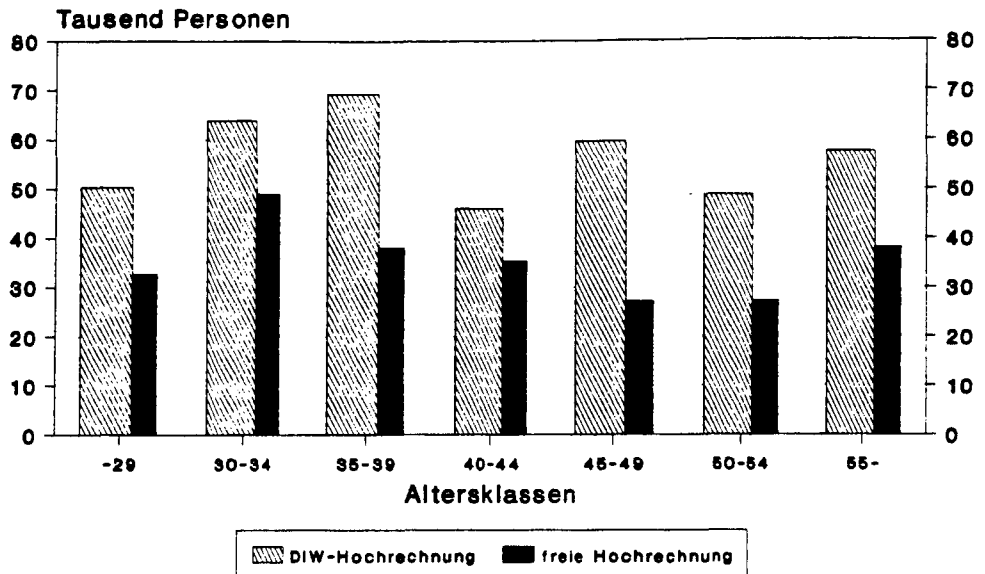


Quellen: SOEP-Ost, eigene Berechnungen,
Statistisches Amt der DDR.

Abbildung 8a

Geschiedene Männer nach Alter

Hochrechnung: SOEP-Ost Basiserhebung '90

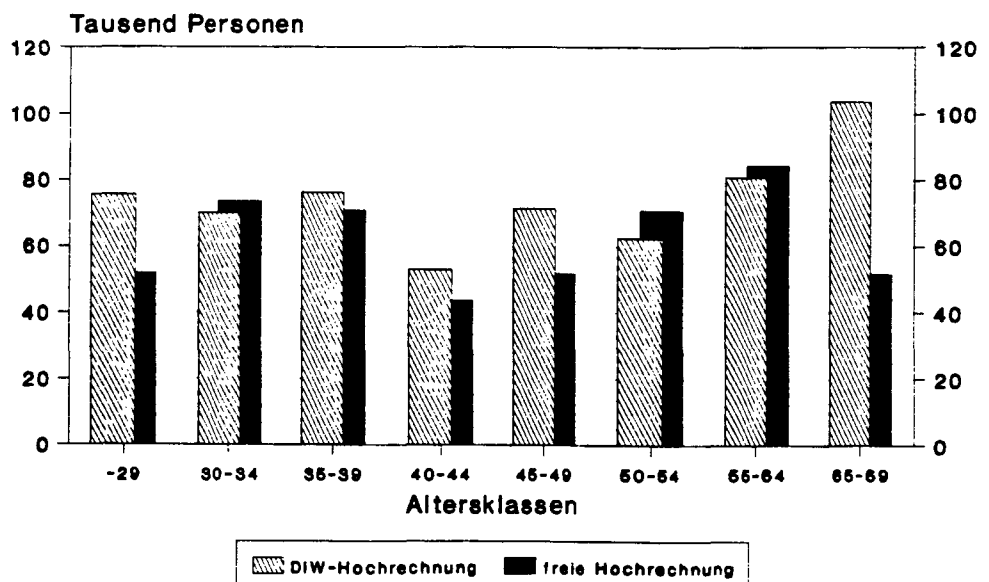


Quellen: SOEP-Ost, eigene Berechnungen,
Statistisches Amt der DDR.

Abbildung 8b

Geschiedene Frauen nach Alter

Hochrechnung: SOEP-Ost Basiserhebung '90



Quellen: SOEP-Ost, eigene Berechnungen,
Statistisches Amt der DDR.

4.2.4 Die Haushaltstruktur

Als "Restriktion" ist in die Hochrechnung für die Anpassung der Haushalte nur die Ecksumme eingegangen. Die Gründe hierfür sind im Abschnitt 4.1 "Hochrechnungsrahmen" genannten worden. Die Verteilung nach Haushaltsgröße mußte mangels geeigneter Rahmendaten allein dem Stichprobenergebnis sowie den Auswirkungen der übrigen Anpassung an Personenmerkmale überlassen werden:

In Abbildung 9 sind Haushaltsstrukturen dargestellt worden, die sich aus vier verschiedenen Schätzungen ergaben:

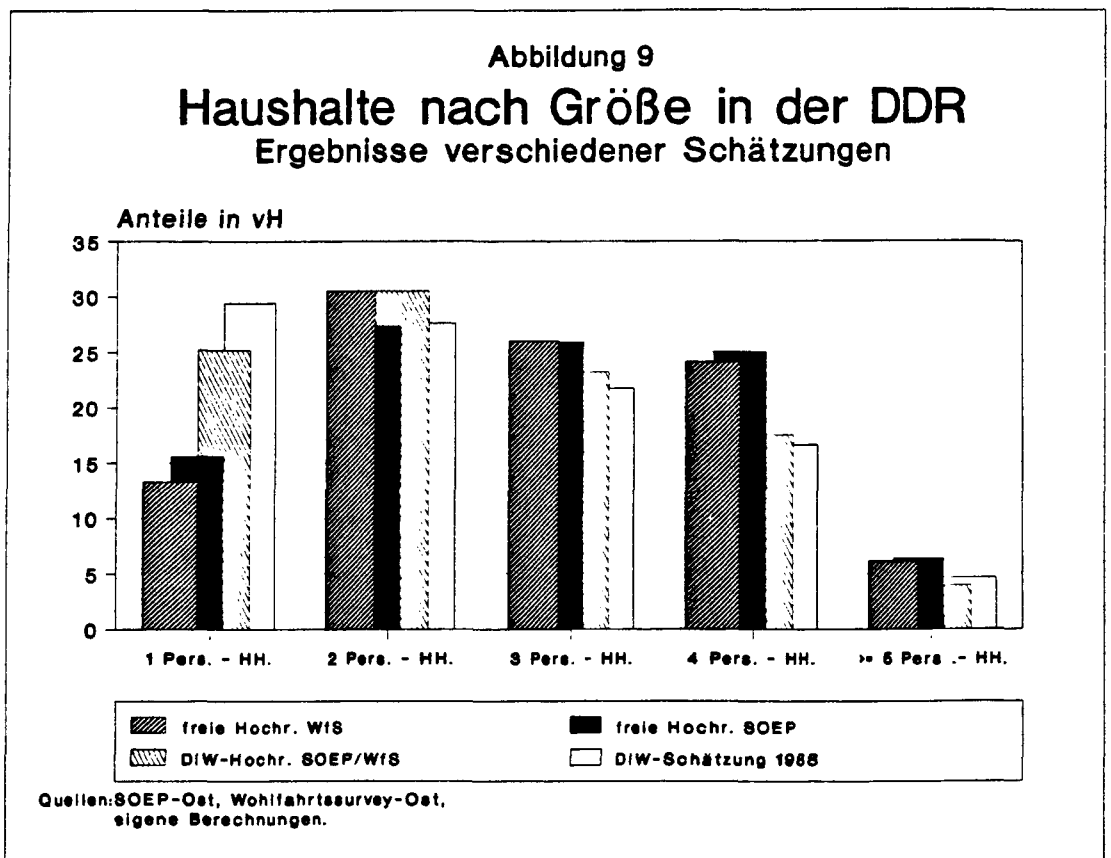
DIW-1988	Schätzung des DIW für das Jahr 1988 ²¹
DIW-Hochrechnung	Vorliegende Gewichtung des DIW mit 115 Restriktionen.
freie Hochrechnung SOEP-Ost	Ungewichtete Hochrechnung der Basiserhebungung DDR.
freie Hochrechnung Wfs-Ost	Ungewichtete Hochrechnung des Wohlfahrtssurveys.

In der Abbildung sind bereits Haushaltsstrukturen abgebildet, die sich auf den Wohlfahrtssurvey beziehen. Auf diese wird im Abschnitt 5 gesondert eingegangen werden.

Während sich die Anteile für die 2, 3 und 5 u.m.-Personenhaushalte relativ wenig voneinander unterscheiden, führen die Verfahren in den beiden übrigen Klassen doch zu erheblichen Abweichungen. Vor allem in den Ein-Personenhaushalten sind die Differenzen signifikant. Zwei Hauptgründe sind dafür verantwortlich, daß Ein-Personenhaushalte in nahezu jeder Stichprobe unterrepräsentiert sind: Zum einen ist die Antreffwahrscheinlichkeit und somit die Antwortwahrscheinlichkeit geringer, weil nur eine Person im Haushalt wohnt, zum anderen leben in ihnen häufig sehr junge oder relativ alte Personen, deren Verweigerungsquote - wenn auch meist aus unterschiedlichen Gründen - überdurchschnittlich hoch ist. Der erzielte Anteil von 15,6 vH, der mit dem Wert der freien Hochrechnung identisch ist, er-

21 Bedau und Vortmann (1990).

scheint auch unter diesen Umständen als ausgesprochen niedrig²². Auf der anderen Seite wird der Schätzwert des DIW für 1988 mit 29,4 vH eher als eine Obergrenze anzusehen sein²³. So gesehen ist das erzielte Ergebnis der Hochrechnung mit 25,2 vH als plausibel einzuschätzen, das sich vornehmlich aus der Anhebung der Gewichte für Ledige über 19 Jahre und ältere Personen ergeben hat.



- 22 Zum Vergleich: In der Basiserhebung des SOEP-West lag der Anteil der Ein-Personenhaushalte bei 22,6 vH.
- 23 Die DDR-Volkszählung aus dem Jahr 1981 ergab einen Anteil der Einpersonenhaushalte an allen Haushalten von rund 25 vH. Zwar ist es aufgrund von demographischen Verschiebungen und Verhaltensänderungen wahrscheinlich, daß dieser Anteil gestiegen ist. Ein Anteil von fast 30 vH ist angesichts der Wohnungsprobleme in der ehemaligen DDR indes eher unwahrscheinlich.

Inzwischen liegen für eine Teilmenge des SOEP-Ost Vergleichsdaten aus einer weiteren Erhebung im Beitrittsgebiet vor, die zum Vergleich mit dem SOEP-Ost herangezogen werden kann: Der "Arbeitsmarktmonitor", der im Auftrag des IAB/Nürnberg von Infratest Sozialforschung durchgeführt wird ist erstmalig im November 1990 erhoben worden. Diese schriftliche Befragung erfaßt die Bevölkerung im erwerbsfähigen Alter; das waren in der ersten Welle die Geburtsjahrgänge zwischen 1926 und 1974. Mit dieser Befragungsform werden grundsätzlich Ein-Personenhaushalte besser erreicht als mit Erhebungen, die auf persönlichen Interviews beruhen. Da mit 70 Vh eine sehr gute Rücklaufquote erzielt wurde können die durch den Arbeitsmarktmonitor ermittelten Haushaltszusammensetzungen durchaus als Maßstab herangezogen werden. Eine Auswertung des SOEP-Ost entsprechend der Stichprobenanlage des Monitors, ergibt folgendes Bild:

Tabelle 3

**Personenbezogene Auswertung (in vH aller Personen) der
Bevölkerung im erwerbsfähigen Alter im Beitrittsgebiet nach
Haushaltsgröße (gewichtete Auswertung)**

	Arbeitsmarkt- monitor	SOEP
	Erhebungszeit	
	November 1990	Juni 1990
Ein-Personen-Haushalt	8	7
Zwei-Personen-Haushalt	27	27
Drei-Personen-Haushalt	31	33
Vier-Personen-Haushalt	27	26
Fünf-u.mehr-Personen- Haushalt	7	7
Insgesamt	100	100

Quelle: Infratest Sozialforschung und DIW.

Ersichtlich stimmen beide Verteilungen gut überein; der leicht höhere Anteil der Ein-Personen-Haushalte erklärt sich gut durch die erwähnte schriftliche Befragungsform gegenüber dem Random-Walk, der bei dem SOEP-Ost angewendet wurde. Insgesamt scheinen beide Befragungen gleichermaßen verlässliche Ergebnisse zu vermitteln.

5 Die Gewichtung des Wohlfahrtssurveys

5.1 Der Hochrechnungsrahmen

Bevor man an die Erstellung des Hochrechnungsrahmens für den Wfs herangehen kann, müssen vor allem zwei Unterschiede gegenüber dem SOEP-Ost beachtet werden. Zum einen beträgt die Fallzahl des Wohlfahrtssurvey nur etwa ein Achtel des SOEP-Ost, wenn man die Zahl der in der Hochrechnung berücksichtigten Personen - beim SOEP also auch die Kinder - berücksichtigt. Dies hat zur Konsequenz, daß die feine Gliederung in der demographischen Struktur der Personen nicht übernommen werden kann. Ein zweites Problem ergibt sich aus der Stichprobenanlage. Der Wohlfahrtssurvey ist als Personenstichprobe angelegt, die Männer und Frauen erfaßt, die zum Zeitpunkt der Befragung 18 Jahre und älter - also volljährig - waren. Dieser Umstand erfordert grundsätzlich Gewichtungen allein auf dieser Ebene. Nun war es von Interesse, auch aus dem Wohlfahrtssurvey Informationen über die Verteilung der Haushaltsgröße zu gewinnen. In freier Hochrechnung ergibt sich für den Anteil der Ein-Personenhaushalte ein Wert von 13,3 vH (vgl. hierzu die Abbildung 9). Dieser liegt noch unter dem Wert für das SOEP-Ost (15,6 vH). Da im Wohlfahrtssurveys die demographische Struktur des Gesamthaushalts erfaßt worden war, war es möglich, analog zum Panel eine simultane Haushalts- und Personengewichtung vorzunehmen, die sich auf Angaben von 2.060 Personen einschl. Kindern bezog. Dies entspricht immerhin einem Drittel der Größe des SOEP, so daß dessen Hochrechnungsrahmen

nahezu identisch übernommen werden konnte²⁴. Es ergab sich der erwartete Effekt, daß allein die demographische Struktur einen starken Einfluß auf den Anteil der Ein-Personenhaushalte ausübte. Mit 23,9 vH erreichte er fast den Wert des SOEP-Ost. Diese Ähnlichkeit erlaubte es zur besseren Vergleichbarkeit der beiden Stichproben, dem WfS die für das SOEP geschätzte Haushaltsgrößenstruktur vorzugeben. Diesem Plan stand entgegen, daß der Wohlfahrtssurvey als Personenstichprobe angelegt wurde; die erwähnte Haushalts- und Personengewichtung war nicht verwendbar, da nur ein gutes Drittel der in diese Hochrechnung berücksichtigten Personen (735 von 2.060) effektiv befragt wurden. Um die gewünschte Haushaltsstruktur zu erzielen, wurde ein Trick angewendet, indem die Haushaltsgröße als Personenmerkmal definiert wurde. Dies führte zwar zu dem Ergebnis, daß die Hochrechnung der Haushalte statt der zu erwartenden Zahl von knapp 6,8 Mill. insgesamt 12,7 Mill. Haushalte ausweist, was gerade der Gesamtzahl aller Personen entspricht, die 18 Jahre und älter sind. Die Struktur der erzielten Haushaltsgrößenverteilung entspricht aber exakt der des SOEP, so daß die korrekte Absolutzahl durch nachträgliche Skalierung mit dem Faktor 0,5349 ermittelt werden muß.

Der Hochrechnungsrahmen des SOEP besteht - vereinfacht formuliert - aus einem regionalen und einem demographischen Block, die auf den Wohlfahrtssurvey zu übertragen sind:

Die regionale Gewichtung wird wieder durch Anpassung der Bevölkerung nach Bezirk und Geschlecht vorgenommen. Hierbei trat das Problem auf, daß eine entsprechende Statistik für volljährige Personen nicht zur Verfügung stand. Sie ist altersmäßig nur

24 Es mußten lediglich bei den Verwitweten und Geschiedenen Altersklassen zusammengefaßt werden.

gegliedert nach

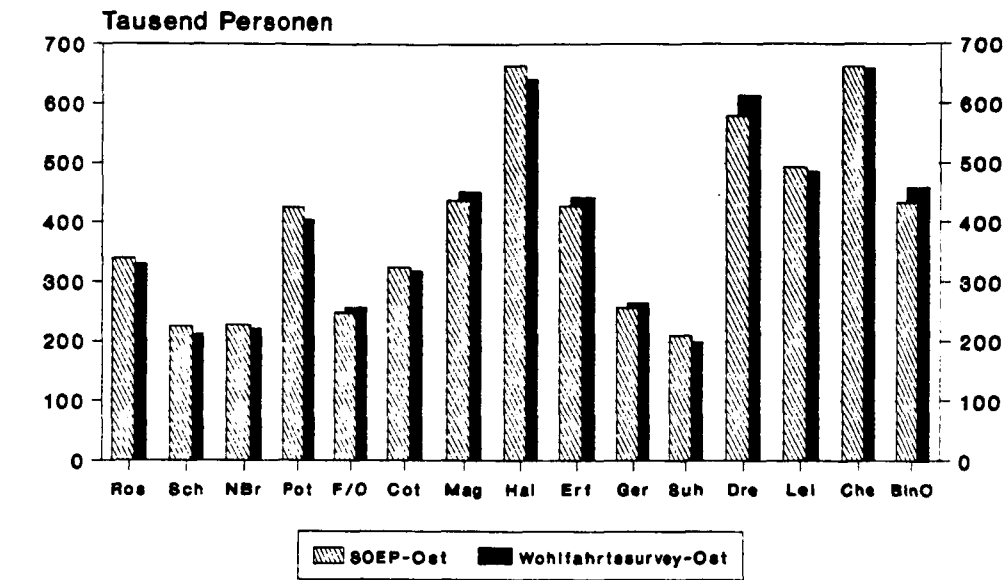
Kindern (nicht arbeitsfähiges Alter),
 Personen im arbeitsfähigem Alter und
 Personen im Rentenalter (nicht arbeitsfähiges Alter).

Ersichtlich fehlen Angaben zu Personen im arbeitsfähigen Alter, die jünger als 18 Jahre alt sind. Da sich der Anteil dieses Personenkreises insgesamt auf nur ca. 3 vH beläuft, ließen sich die erforderlichen Eckdaten aus der genannten Statistik recht gut schätzen. Sie bildeten die Restriktionen für die Regionalverteilung.

In den Abbildungen 10a und 10b sind die Schätzungen den entsprechenden Daten aus dem Hochrechnungsergebnis des SOEP gegenübergestellt.

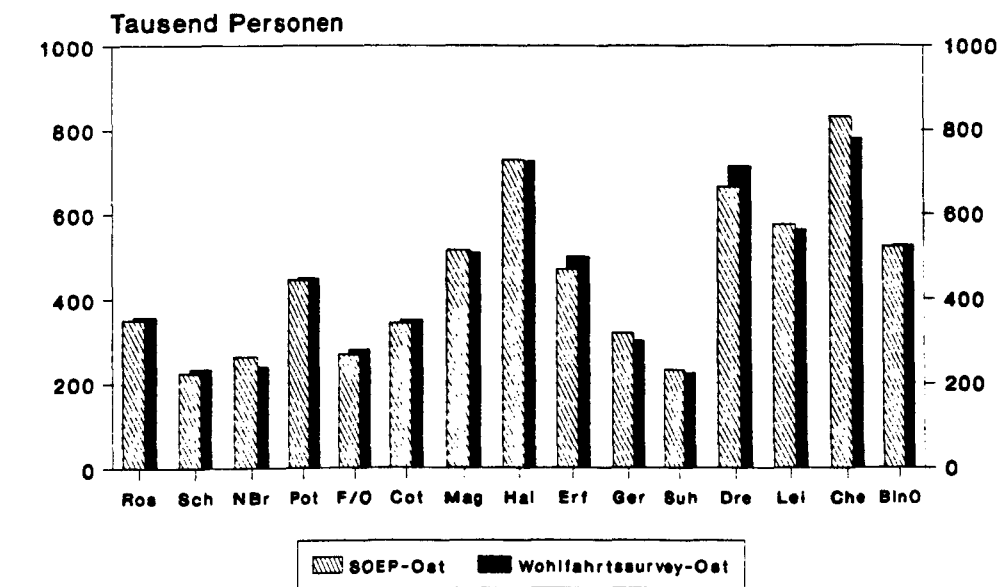
Fallzahlprobleme ergaben sich bei der Anpassung an Alter, Familienstand und Geschlecht. Zwar konnte für die Frauen und Männer insgesamt die feingliedrige Abstufung in fünfer Klassen im Altersbereich von 25 bis 69 Jahren beibehalten werden, die jüngeren und älteren Jahrgänge mußten zu einer Klasse zusammengefaßt werden. Die Verteilung nach dem Familienstand bereitete bei den Männern einige Probleme: In der Stichprobe wurden nur 14 ledige Männer befragt, die älter als 29 Jahre waren. Weiterhin gab es insgesamt nur 7 Witwer und wiederum 14 Geschiedene. Die Konsequenz war, daß für verheiratete Männer nur drei, für verwitwete eine und schließlich für geschiedene Männer lediglich zwei Altersklassen vorgesehen werden konnten. Günstiger - aus statistischer Sicht - erwies sich die Verteilung nach Familienstand für die Frauen. Verheiratete Frauen konnten in sechs, verwitwete und geschiedene Frauen in jeweils drei Altersklassen gegliedert werden.

Abbildung 10a
Vollj. Männer in der DDR nach Bezirken
SOEP-Ost vs. Wohlfahrtssurvey-Ost '90



Quellen: SOEP-Ost, Wohlfahrtssurvey-Ost,
Stat. Amt d. DDR, eig. Berechn.

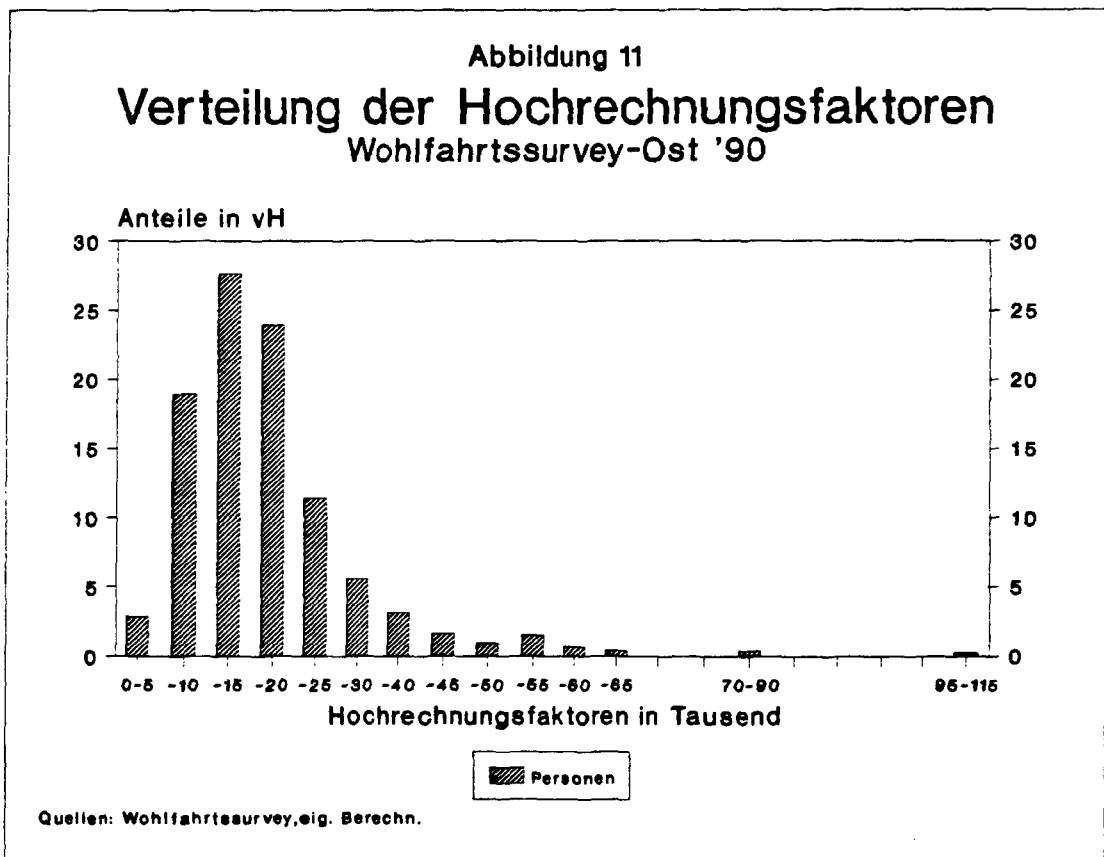
Abbildung 10b
Vollj. Frauen in der DDR nach Bezirken
SOEP-Ost vs. Wohlfahrtssurvey-Ost '90



Quellen: SOEP-Ost, Wohlfahrtssurvey-Ost,
Stat. Amt d. DDR, eig. Berechn.

5.2 Ergebnisse der Gewichtung

Wie für das SOEP-Ost werden zunächst einige statistische Eigenschaften der Gewichtung betrachtet. Auf eine ausführliche Darstellung der demographischen Strukturen wird dagegen verzichtet, da dies vielfach eine Wiederholung bereits gezeigter Daten bedeuten würde. Vielmehr werden an zwei ausgewählten Verteilungen die Unterschiede zwischen den beiden Stichproben demonstriert.



5.2.1 statistische Eigenschaften der Hochrechnungsfaktoren

Die Verteilung der Hochrechnungsfaktoren für den WfS sind in der Abbildung 11 dargestellt. Das Durchschnittsgewicht beträgt

rund 17.280; sie ist also mehr als sechs mal so hoch wie im SOEP, die Konsequenz niedrigerer Fallzahlen. Der niedrigste Wert wurde mit 1.550 ermittelt. Die betreffende Person geht also nur mit einem Elftel (ein Achtel beim SOEP) des Durchschnittsgewicht ein. Mit einem Wert von 111.173 ist der größte Hochrechnungsfaktor 6,4 mal so hoch wie das arithmetische Mittel (SOEP:6,2). Insgesamt ist die Verteilung der Hochrechnungsfaktoren des Wohlfahrtssurvey ungünstiger als die des SOEP. Dies spiegelt sich auch in der Höhe des Variationskoeffizienten wider, der sich als Quotient aus Standardabweichung und arithmetischem Mittel berechnet: Mit 0,68 ist er um fast 50 vH größer als der des SOEP (0,47).

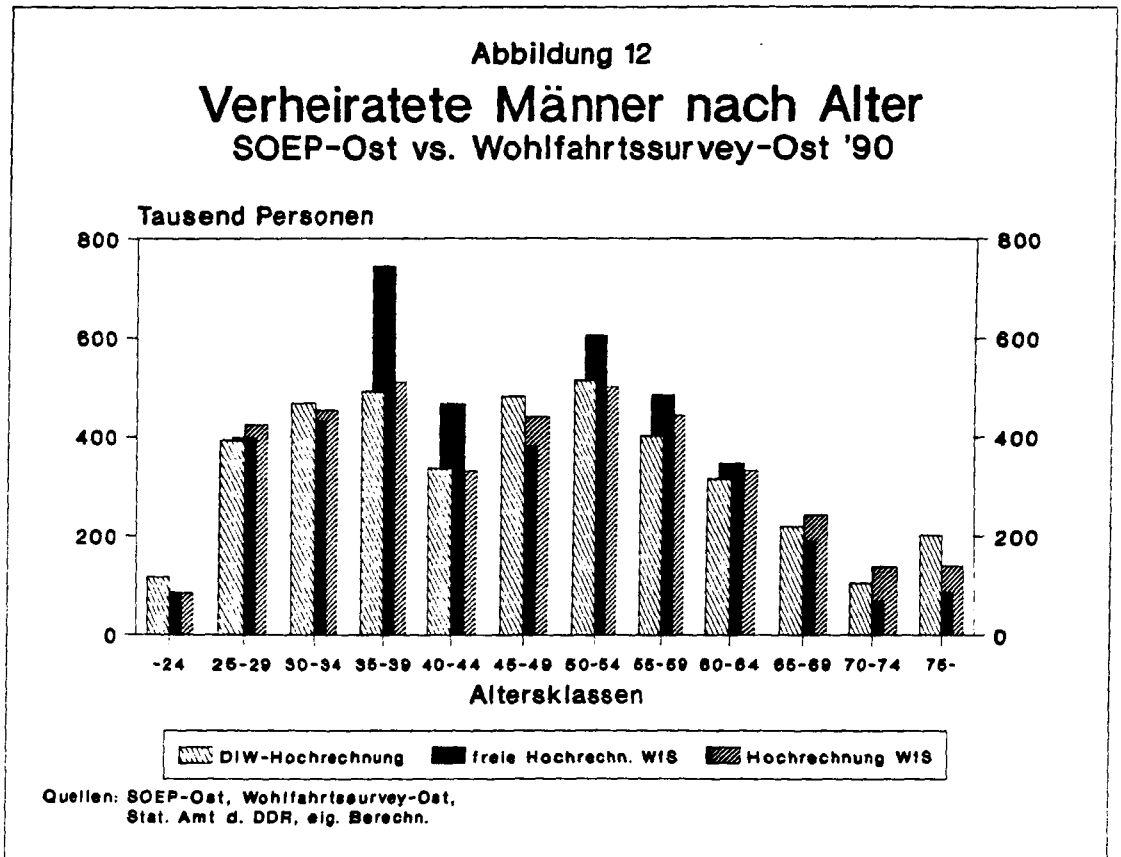
Die wichtigsten statistischen Kennziffern der Hochrechnungsfaktoren auf einen Blick:

Kennziffer	Personen
Anzahl	735
Summe	12.700.622
Arithmetisches Mittel	17.279,76
Standardabweichung	11.754,57
Variationskoeffizient	0,68
Minimum	1.550
Maximum	11.1173

5.2.2 Zwei Beispiele zur demographischen Verteilung

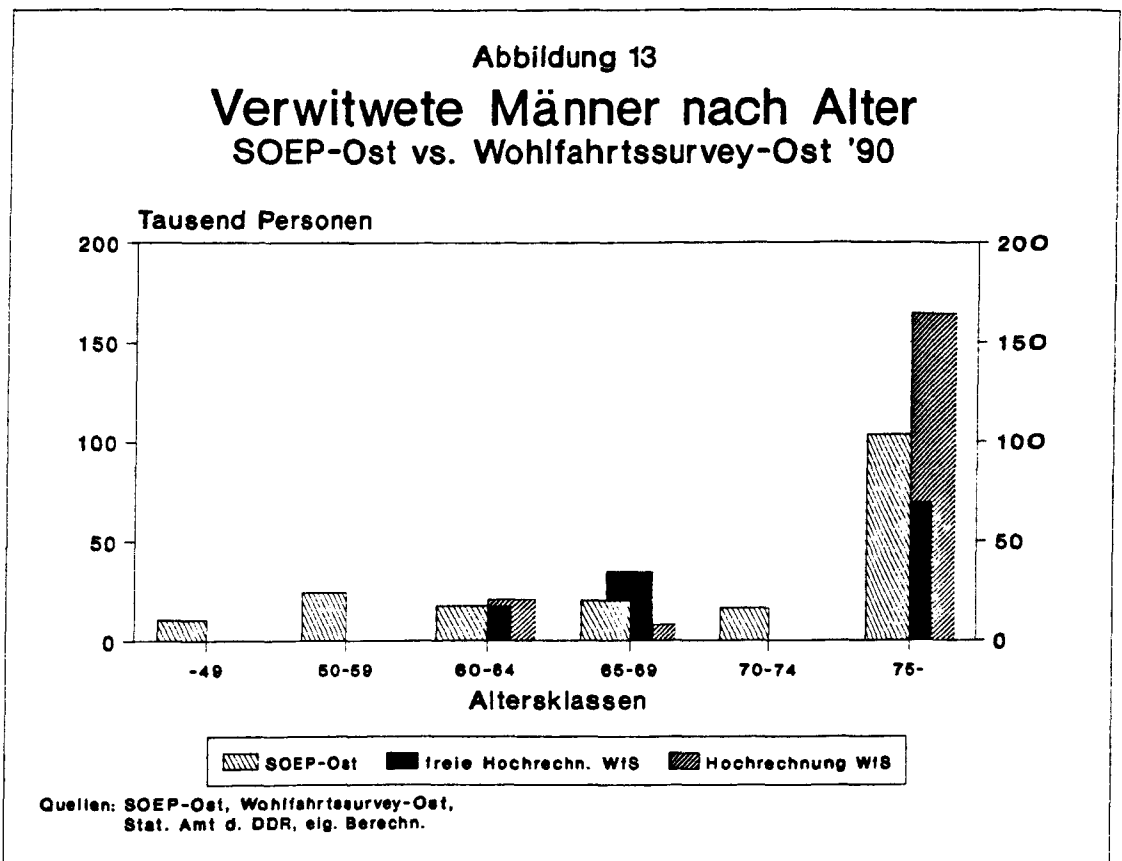
Die bedeutend geringeren Fallzahlen des Wohlfahrtssurvey führten bei der Gliederung nach dem Familienstand zu wesentlich weniger Altersklassenklassen als beim SOEP. Dies galt in höherem Maß für die Männer. So gab es für verheiratete Männer nur drei Altersklassen, Witwer wurden altersmäßig nicht unterschieden. Es ist reizvoll zu sehen, welche Alterstrukturen sich für diese Subpopulationen nach der Hochrechnung im Vergleich zu den Vorgaben durch das Statistische Amt der DDR bzw. durch das SOEP ergeben.

Abbildung 12 zeigt die Verteilung der verheirateten Männer in der DDR. Die Werte für das SOEP sind dieselben wie in Abbildung 6a. Obwohl nur für drei Altersklassen Ecksummen vorgegeben



wurden, zeigt sich insgesamt eine zufriedenstellende Anpassung an die "wahre" Struktur, die in diesem Fall durch die Basiserhebung des DDR-Panel repräsentiert wird. Dieser Effekt ist nicht durch eine besonders gute Ausgangsverteilung bewirkt worden, was sich besonders an der schlechten Anpassung der freien Hochrechnung für die verheirateten Männer im Alter von 35 bis 44 Jahre ablesen läßt, die um 46 vH übererfaßt wurden. Vielmehr gaben hierfür die - dem SOEP ähnlich stark - disaggregierten Restriktionen für die Männer insgesamt, den Ausschlag.

Als Gegenstück zu dem recht erfreulichen Beispiel, muß die Altersverteilung der Witwer herangezogen werden: Abbildung 13 zeigt, daß eine vernünftige Anpassung bei nur sieben beobachteten Fällen selbstverständlich nicht möglich ist. Drei der sechs gezeigten Altersklassen sind beim Wohlfahrtssurvey nicht besetzt, können also weder in der freien noch in der geschätzten Hochrechnung Werte aufweisen. Dennoch mußte die Gesamtzahl der Witwer im Hochrechnungsrahmen unbedingt berücksichtigt werden, da



diese um gut ein Drittel untererfaßt wurden. Im Ergebnis ist dies aber für Auswertungen nicht hinderlich, da die Gruppe der Witwer aufgrund der kleinen Fallzahlen ohnehin nicht auswertbar ist.

6 Soziale Stellung in SOEP und Wohlfahrtssurvey - Ein empirisches Beispiel -

Am Beispiel der sozialen Lage der Frauen in der (ehemaligen) DDR, soll die Vergleichbarkeit der behandelten Stichproben geprüft werden. In der Abbildung 14 ist die Sozialstruktur vor und nach Gewichtung dargestellt²⁵. Ein Vergleich der beiden Erhebungen erfordert zunächst den Hinweis auf einige inhaltliche Unterschiede:

- Die Altersuntergrenze beträgt liegt beim SOEP bei 16 Jahren, bei dem WfS bei 18 Jahren. Dies schlägt sich vornehmlich im Anteil der Auszubildenden nieder.
- Die SOEP-Stichprobe datiert von Mitte 1990, die des Wohlfahrtssurveys vom November 1990. Das führt zu erheblichen Abweichungen in den Arbeitslosenanteilen.
- Im Wohlfahrtssurvey treten in der Klasse der über 60jährigen vier Frauen auf, die sich als "ehemals selbständig" bezeichnen, eine Kategorie die im SOEP nicht enthalten ist. Da der Anteil mit ca. 0,5 sehr niedrig ist, wirkt sich dies kaum auf die gezeigten Verteilungen aus.

Der Vergleich beider Abbildungen zeigt gewichtet wie ungewichtet erstaunliche Übereinstimmungen. Relativ hohe Abweichungen sind nur bei den bereits erwähnten Anteilen der Arbeitslosen und der Auszubildenden zu bemerken. Diese sind aber allein durch die Erhebungsunterschiede zu erklären.

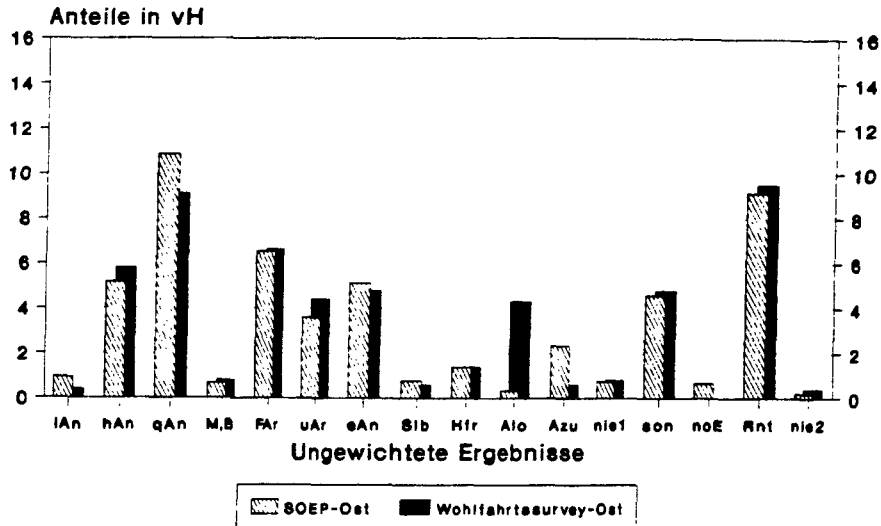
Die Gewichtung wirkt auf beide Erhebungen gleichartig. Der wesentlichste Effekt zeigt in einer Erhöhung des Rentneranteils um ca 5 vH-Punkte, was eindeutig auf die mehrfach erwähnte Untererfassung der älteren Menschen zurückzuführen ist.

Insgesamt erweisen sich die beiden Stichproben - legt man dieses Beispiel als Meßplatte zu Grunde - gewichtet wie ungewichtet als gut vergleich- und auswertbar.

25 Die Daten für diesen Vergleich stellte freundlicherweise Detlef Landua vom WZB zur Verfügung.

Abbildung 14a

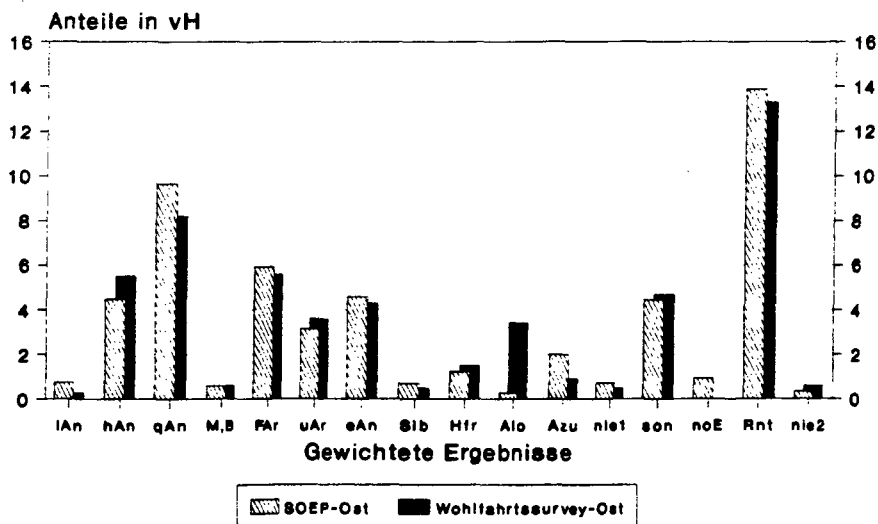
Soziale Lagen der Frauen in der DDR SOEP-Ost vs. Wohlfahrtssurvey-Ost '90



Quellen: SOEP-Ost, Wohlfahrtssurvey-Ost,
Stat. Amt d. DDR, eig. Berechn.

Abbildung 14b

Soziale Lagen der Frauen in der DDR SOEP-Ost vs. Wohlfahrtssurvey-Ost '90



Quellen: SOEP-Ost, Wohlfahrtssurvey-Ost,
Stat. Amt d. DDR, eig. Berechn.

Legende

Bis 60 Jahre alt:

lAn	Leitende Angestellte
hAn	Hochqual. Angestellte / Höhere Beamte
qAn	Qualifizierte Ang. / Gehob. Beamte
M.B	Meister oder Brigadier
fAr	Facharbeiter
uAr	Un- angelernter Arbeiter
eAn	Einfache Angestellte / Beamte
Sib	Selbständige
Hfr	Hausfrauen

Alo	Arbeitslose
Azu	Auszubildende
nie1	Noch nie Erwerbstätige
son	sonstige nicht Erwerbstätige
61 Jahre und älter:	
noE	noch Erwerbstätige
Rnt	Rentner
nie2	Noch nie Erwerbstätige

7 Zusammenfassung und abschließende Bemerkungen

In diesem Beitrag wurden konsistente Hochrechnungen und Gewichtungen für das sozio-ökonomische Panel-Ost und der Pilotstudie des Wohlfahrtssurvey-Ost vorgestellt. Grundgedanke für die Anlage der Gewichtungsverfahren war, die repräsentativ angelegte Stichproben weitgehend zu erhalten, d.h. von einer Gleichverteilung der Gewichte ausgehend, diese nach dem Prinzip des minimalen Informationsverlustes nur so weit zu verändern, daß eine vorgegebene Anzahl von Randbedingungen erfüllt sind. Diese Bedingungen waren beim SOEP 114 Restriktionen zur Personenverteilung nach Region, Geschlecht, Familienstand und Alter. Nur eine Restriktion - die Zahl der DDR-Haushalte insgesamt - wurde aus empirischen Gründen für die Haushaltsebene vorgesehen, so daß 115 Restriktionen in die Gewichtung des SOEP eingingen.

Wegen seiner geringen Fallzahl mußte die Hochrechnung des WfS mit bedeutend weniger Restriktionen auskommen. Diese reine Personenwichtung setzte sich aus 68 Restriktionen zur regionalen und demographischen Beschreibung und vier vorgegebenen Ecksummen zur Haushaltsgröße zusammen.

Gewiß bestand die Möglichkeit, weitere Restriktionen z.B. durch Einbeziehung von Gemeindegrößenklassen aufzunehmen. Indes ist es nicht unbedingt von Vorteil, die Zahl der Restriktionen nach Belieben auszudehnen. Mit jeder zusätzlichen Restriktion nimmt die Varianz der Hochrechnungsfaktoren zu. Diese überträgt sich grundsätzlich auch auf die Stichprobenvarianz. Dies wiederum führt zu größeren Konfidenzintervallen der zu schätzenden Parameter und somit zu Signifikanzverlusten.

Insgesamt werden durch die vorgestellten Hochrechnungen das SOEP-Ost und der WfS-Ost zu vergleichbaren Stichproben, wie an einem ausgewählten Beispiel zur sozialen Lage der Frauen in Ostdeutschland demonstriert werden konnte. Eine analytische Verknüpfung beider Erhebungen ist möglich.

Literatur

- Bartholmai, Bernd, Manfred Melzer und Erika Schulz (1990):** Privathaushalte und Wohnungsbedarf in Deutschland bis zum Jahr 2000, in: Wochenbericht des DIW, Nr. 42/90, S.591-598, Berlin.
- Bedau, Klaus - Dietrich und Heinz Vortmann (1990):** Die Einkommensverteilung nach Haushaltsgruppen in der ehemaligen DDR. In: Wochenbericht des DIW, Nr. 47/90, S.655-660, Berlin.
- Galler, Heinz Peter (1977):** Die Ermittlung eines konsistenten Hochrechnungsrahmens für das IMDAF-69, SPES-Arbeitspapier Nr. 77, Frankfurt.
- Horvitz, D. G., und D. J. Thompson (1952):** A Generalization of Sampling without Replacement from a Finite Universe, in: Journal of the American Statistical Association, Volume 47, S.663-685.
- Habich, Roland und Detlef Landua (1991):** Der Wohlfahrtssurvey im Herbst 1990. In: Projektgruppe Panel (Hg.) Lebenslagen im Wandel - Basisdaten - und -analysen zur Entwicklung in den neuen Bundesländern. Frankfurt und New York.
- Infratest Sozialforschung (1990):** Das Sozio-ökonomische Panel - Basiserhebung in der DDR - . Methodenbericht zur Durchführung der Befragung, München.
- Merz, Joachim (1983):** Die konsistente Hochrechnung von Mikrodaten nach dem Prinzip des minimalen Informationsverlustes, in: Allgemeines Statistisches Archiv, Heft 4, S. 342-366.
- Pischner, Rainer (1991):** Eine konsistente Haushalts- und Personengewichtung für die DDR-Basisbefragung des SOEP. Berlin 1991, DIW-Diskussionspapier Nr. 18, Berlin.
- Rendtel, Ulrich (1987):** Methodische Konzepte für die Hochrechnung von Panel-Daten, in: Vierteljahrshefte zur Wirtschaftsforschung, Heft 4, S. 278-290, Berlin.
- Rendtel, Ulrich (1988):** Repräsentativität und Hochrechnung der Datenbasis, in: Krupp, Hans-Jürgen und Jürgen Schupp (Hrsg.) 1988: Lebenslagen im Wandel: Daten 1987, Band 3 der Reihe: Sozio-ökonomische Daten und Analysen für die Bundesrepublik Deutschland, Frankfurt/New York.
- Rosenblatt, Bernhard von und Jürgen Schupp (1990):** Die DDR-Stichprobe des SOEP. In: Beiträge zur Arbeitsmarkt- und Berufsforschung Nr. 143.
- Schupp, Jürgen und Gert Wagner (1990):** Die DDR-Basisbefragung des Sozio-ökonomischen Panels. In: Vierteljahrsheft zur Wirtschaftsforschung, 1990, Heft 2-3, S. 152-159.
- Schupp, Jürgen und Gert Wagner (1991):** Konzeption und Feldarbeit der Basiserhebung des SOEP in der DDR. In: Projektgruppe Panel (Hg.) Lebenslagen im Wandel - Basisdaten - und -analysen zur Entwicklung in den neuen Bundesländern. Frankfurt und New York.