ECONSTOR
Make Your Publications Visible.

A Service of
ZBW
Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Pischner, Rainer

Article — Digitized Version

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

Vierteljahrshefte zur Wirtschaftsforschung

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, Vierteljahrshefte zur Wirtschaftsforschung, ISSN 0340-1707, Duncker & Humblot, Berlin, Vol. 60, Iss. 1/2, pp. 50-64

This Version is available at:
https://hdl.handle.net/10419/141010

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.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

Eine konsistente Haushalts- und Personengewichtung für die DDR-Basisbefragung des SOEP und für die Ost-Pilotstudie des Wohlfahrtssurveys[1]

von Rainer Pischner

Aktuelle statistische Informationen über die sozio-demographische Struktur der privaten Haushalte liegen für die ehemalige DDR nur spärlich vor[2]. 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 Konzepts und unter Einbeziehung der Ergebnisse wird dann die Gewichtung des Wohlfahrtssurveys vorgenommen.

1. Die Stichprobenanlage des SOEP-Ost und des Wohlfahrtssurveys

Konzeption und Feldarbeit der DDR-Basisbefragung werden an anderer Stelle ausführlich dargestellt (vgl. Schupp und Wagner, 1991). Eine ausführliche Beschreibung des Stichprobenplans ist im Methodenbericht des Erhebungsinstituts zu finden (Infratest Sozialforschung, 1990). Über den Wohlfahrtssurvey berichten Habich und Landua (1991). 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 mit deutscher Bezugsperson bzw. deutschem Haushaltsvorstand und alle 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 bestimmt. Die Startadressen wurden proportional zur Zahl der benötigten Sample-Points je Gemeinde ermittelt. 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 ausgewählt und schließlich acht Interviews vorgenommen worden (vgl. hierzu auch weiter unten den Abschnitt 2).

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. Von insgesamt 10 000 Adressen wurden für den WfS 2 000 Adressen zur Verfügung gestellt. 212 Start- und wei-

---

[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. DIW (1990a).

tere 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 Stichproben 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-Ost. Diese Stichprobe wurde trotz der sehr kurzen Vorbereitungszeit dem bundesdeutschen Standard entsprechend als bevölkerungsrepräsentative Stichprobe angelegt.

So sorgfältig Erhebungen auch geplant und durchgeführt werden, bei der Feldarbeit kommt es zu Fehlern und meist auch zu systematischen Verzerrungen, die die Repräsentativität der Stichprobe beeinträchtigen: 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 wurden[3]. 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 heraus, weil sie — bedingt durch die kurze Feldzeit — nicht rechtzeitig bearbeitet werden konnten. Die insgesamt 502 Ausfälle gelten zwar als qualitätsneutral, d.h. sie sind 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 bezogenen Ausschöpfungsquote von 70 vH. Von den ursprünglich 3 616 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 die Parameter bedingter Verteilungen unter bestimmten Voraussetzungen auch anhand nichtrepräsentativer Stichproben erwartungstreu geschätzt werden können, ist 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 multplizieren sind. Deshalb werden in diesem Aufsatz die Begriffe Gewichte und Hochrechnungsfaktoren synonym verwendet. Sollte eine Unterscheidung notwendig sein, wird 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[4 5]:

Den Hochrechnungen sämtlicher Stichproben liegt der Grundgedanke von Horvitz/Thompson[6] zugrunde, daß die inverse Auswahlwahrscheinlichkeit eines Stichprobenelements einen unverzerrten Schätzwert für den Hochrechnungsfaktor der Elemente darstellt.

Für das SOEP-Ost soll hier unter Auswahlwahrscheinlichkeit die Wahrscheinlichkeit verstanden werden, daß von einem beliebigen privaten Haushalt mit deutscher Bezugsperson ein ausgefüllter Haushaltsfragebogen vorliegt sowie von allen Personen, die 16 Jahre und älter sind, ein Personenfragebogen 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,

---

[3] Vgl. von Rosenbladt und Schupp (1990).

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

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

[6] Siehe hierzu die grundlegenden Arbeiten von Horvitz und Thompson (1952).

— 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 Auswahlsatz, also als Grundgesamtheit, dividiert durch die Zahl der Stichprobenelemente. Wie bereits angedeutet, sieht die Realität freilich anders aus. Bis die erforderliche Zahl von Interviews tatsächlich 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.

— Die Sample-Points sind gleichmäßig über die DDR verteilt gewesen, aber die Ausschöpfung fiel ungleichmäßig aus, so daß die Bezirke unterschiedlich repräsentiert werden. *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 solche Anpassung die Stichprobe vornehmlich nur regional entzerrt 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, berüchsichtigt werden als beim WfS.

Für die Basisbefragung des SOEP-Ost werden 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 übereinstimmen[7].
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 wie folgt zu erfüllen: 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. Anschließend wird der Vektor nach dem Prinzip der Minimierung des Informationsverlusts, also möglichst geringfügig, korrigiert, so daß vorgegebene Randbedingungen erfüllt werden. Diese Randbedingungen werden im nächsten Abschnitt zum Hochrechnungsrahmen näher beschrieben.

4. Die Gewichtung der DDR-Basisbefragung

4.1 Der Hochrechnungsrahmen

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

Seit Mitte 1989 verließen hunderttausende Bürger der DDR das 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 einen 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[8]. Es erhebt sich die Frage, warum unter diesen Umständen überhaupt an Eckwerte angepaßt wird, die oh-

[7] 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.

[8] Das Statistische Amt der DDR geht bis Ende Juni 1990 von etwa 200 Tausend Personen aus, berücksichtigt aber nur ordnungsgemäß abgemeldete Personen. Dagegen spricht das Statistische Bundesamt Wiesbaden von etwa 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.

nehin 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 Wanderungsverluste von 1 bis 2 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 die offiziellen Zahlen von Ende 1989 unverzichtbar, auch wenn diese Daten überholt sind. Es bleibt dem Anwender immer noch überlassen, auf Basis der Hochrechnung Aktualisierungen der Rahmendaten vorzunehmen.

2. Wie weiter unten 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,* nach Bezirken vorgenommen, weist einige Besonderheiten auf, die erklärungsbedürftig sind:

Im Unterschied zur Bundesrepublik war die DDR politisch in 15 Bezirke und nicht nach Ländern gegliedert. Die Bezirke wurden 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)

Die Basisbefragung wurde allerdings am 30.6.90 zu einem Zeitpunkt abgeschlossen, als einige Kreise den neuen Ländern wegen anstehender Volksentscheide noch nicht endgültig zugeordnet waren. Dieser Wechsel der Zuordnungen schlug sich auch in der Stichprobe nieder: Bei den 2 179 befragten Haushalten gibt es in 71 Fällen (3,3 vH) Inkonsistenzen zwischen Bezirks- und Länderabgrenzungen. 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 dem Geschlecht aus. Hieraus ergeben sich 2*(15-1) = 28 Restriktionen[9].

Tabelle 1 ist die vorgegebene Regionalverteilung zu entnehmen:

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

Einschließlich der Kinder umfaßt das SOEP-Ost 6 044 Personen. Gemessen an der Wohnbevölkerung von 16 433 796 Personen ergibt sich damit 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 Geschlechtern ergeben sich die in der Tabelle ausgewiesenen Werte. Sie weichen nur geringfügig von dem Durchschnittswert ab. Die Zahl der Männer ist mit einem Faktor von 2 670 um 1,8 vH überrepräsentiert[10], die der Frauen mit einem Faktor von 2 766 um etwa 1,7 vH untererfaßt[11] [12].

Regional gibt es weitaus größere Abweichungen. Mit 20 vH am stärksten unterrepräsentiert sind die Frauen im Bezirk Erfurt (HF=3 403). Mit fast 50 vH wurden sie im Bezirk Suhl übererfaßt (HF=1 825).

---

[9] 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.

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

[11] Generell gilt: Liegen die Werte in den Spalten „DDR/SOEP" über dem Durchschnittswert von 2719, liegt Untererfassung vor, da ein Stichprobenelement mehr Personen zu repräsentieren hat. Werte kleiner als 2719 zeugen von einer überproportionalen Ausschöpfung der jeweiligen Merkmalskombination.

[12] Diese und die weiteren prozentualen Abweichungen ergeben sich durch Vergleich der Ist- und Sollzahlen in der Stichprobe: Insgesamt wurden 3095 Frauen erhoben. Aufgrund der oben erwähnten Auswahlwahrscheinlichkeit von 0,03678 vH wäre im SOEP-Ost die Zahl von 3149 Frauen zu erwarten gewesen. Somit sind 54 von 3149, also ca. 1,7 vH Frauen, zu wenig erfaßt worden.

Tabelle 1

Wohnbevölkerung in der DDR und im SOEP-Ost nach Bezirken und Geschlecht

Bezirke	DDR	Stichprobe SOEP	DDR/SOEP	DDR	Stichprobe SOEP	DDR/SOEP
	Männer			Frauen		
Rostock	444 731	162	2 745	465 109	169	2 752
Schwerin	285 808	123	2 324	304 363	102	2 984
Neubrandenburg	301 168	110	2 738	314 599	109	2 886
Potsdam	537 263	186	2 889	573 947	188	3 053
Frankfurt/Oder	344 944	118	2 923	361 172	126	2 866
Cottbus	426 093	181	2 354	449 488	199	2 259
Magdeburg	594 518	284	2 093	643 389	303	2 123
Halle	836 539	251	3 333	911 491	281	3 244
Erfurt	586 458	187	3 136	636 439	187	3 403
Gera	347 943	135	2 577	380 136	144	2 640
Suhl	262 320	138	1 901	282 951	155	1 825
Dresden	812 012	281	2 890	901 074	313	2 879
Leipzig	631 097	259	2 437	702 045	270	2 600
Chemnitz	853 468	326	2 618	964 019	338	2 852
Berlin(Ost)	611 244	208	2 928	670 274	211	3 177
Insgesamt	7 873 300	2 949	2 670	8 560 496	3 095	2 766

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

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: Deutliche Abweichungen von den vorgegebenen Randverteilungen zeigen sich, wenn zur Bewertung der Repräsentativität die Verteilung nach dem Familienstand herangezogen wird. So sind Verheiratete um etwa 4 vH überrepräsentiert, für diese Gruppe vermutlich kein schlechtes Ergebnis. Auch für die Ledigen errechnet sich insgesamt nur eine relativ geringe Abweichung vom Erwartungswert (etwa +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 ist 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 ist nicht — wie vielleicht zu erwarten wäre — allein auf das höhere Durchschnittsalter jenes Bevölkerungskreises zurückzuführen, betrifft die Minderausschöpfung doch 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 vom 60. Lebensjahr an, wenn viele Menschen in Rente gehen, gesundheitlich nicht mehr auf voller Höhe sind und allmählich weniger aktiv am Leben teilnehmen. Deutlich erkennbar ist der steile Anstieg des empirischen Hochrechnungsfaktors. Zu niedrige Ausschöpfungsquoten finden sich auch bei den jungen Menschen um Zwanzig, die erfahrungsgemäß überdurchnittlich oft verweigern. Dagegen sind Personen im Alter von 30 bis 44 Jahren 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 im Jahr 1990 ist nur wenig abgesichert. Vornehmlich der Anteil der Ein-Personenhaushalte ist ungewiß.

Die amtliche Statistik bietet lediglich Ergebnisse der letzten Volkszählung aus dem Jahre 1981 sowie eine „amtliche Fortschreibung" an[13]. Allein Wohnungsstatistiken jüngeren Datums ergänzen dieses Material[14]. Danach schwanken die Anteile der Ein-Personenhaushalte an allen Privathaushalten zwischen 25,1 vH (Volkszählung 1981) und 29,1 vH (amtliche Fortschreibung 1988). Letztlich sind die amtlichen Zahlen aus der DDR schon vor Beginn der

[13] DIW (1990a), S.656.

[14] Vgl. DIW (1990).

Tabelle 2

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

Männer und Frauen insgesamt	DDR	Stichprobe SOEP	DDR/SOEP	DDR	Stichprobe SOEP	DDR/SOEP
Altersklasse	Männer			Frauen		
0 - 4 J.	545 191	229	2 381	517 254	188	2 751
5 - 9 J.	577 826	272	2 124	551 517	246	2 242
10 - 14 J.	518 660	264	1 965	491 694	240	2 049
15 - 19 J.	502 511	195	2 577	478 014	199	2 402
20 - 24 J.	633 170	194	3 264	596 900	190	3 142
25 - 29 J.	719 458	247	2 913	672 170	274	2 453
30 - 34 J.	649 482	270	2 405	611 288	285	2 145
35 - 39 J.	631 305	247	2 556	601 112	254	2 367
40 - 44 J.	414 361	197	2 103	404 597	209	1 936
45 - 49 J.	579 616	203	2 855	582 704	186	3 133
50 - 54 J.	598 589	216	2 771	609 967	214	2 850
55 - 59 J.	456 572	158	2 890	482 116	155	3 110
60 - 64 J.	353 307	100	3 533	474 354	134	3 540
65 - 69 J.	250 538	68	3 684	468 452	120	3 904
70 - 74 J.	126 770	34	3 729	261 042	70	3 729
75 - 79 J.	162 990	33	4 939	369 108	69	5 349
80 - 99 J.	152 954	22	6 952	388 207	62	6 261
Insgesamt	7 873 300	2 949	2 670	8 560 496	3 095	2 766
Ledige Männer und Frauen *)						
0 - 4 J.	545 191	229	2 381	517 254	188	2 751
5 - 9 J.	577 826	272	2 124	551 517	246	2 242
10 - 14 J.	518 660	264	1 965	491 694	240	2 049
15 - 19 J.	500 938	194	2 582	467 691	199	2 350
20 - 24 J.	510 339	161	3 170	344 060	109	3 157
25 - 29 J.	283 954	79	3 594	137 160	58	2 365
30 - 34 J.	117 518	33	3 561	58 982	19	3 104
35 - 44 J.	100 277	27	3 713	56 829	13	4 371
45 - 54 J.	59 509	11	5 410	53 287	7	7 612
55 - 99 J.	29 264	6	4 877	171 088	39	4 387
Insgesamt	3 243 476	1 276	2 542	2 849 562	1 118	2 549
nachrichtlich:						
0 - 19 J.	2 142 615	959	2 234	2 028 156	873	2 323
20 - 99 J.	1 100 861	317	3 473	821 406	245	3 353
Verheiratete Männer und Frauen						
0 - 24 J.	116 913	34	3 439	244 999	75	3 267
25 - 29 J.	392 229	156	2 514	475 388	202	2 353
30 - 34 J.	467 143	218	2 143	478 827	239	2 003
35 - 39 J.	490 313	215	2 281	480 438	218	2 204
40 - 44 J.	336 099	173	1 943	325 393	182	1 788
45 - 49 J.	481 230	184	2 615	467 694	162	2 887
50 - 54 J.	513 394	198	2 593	478 856	173	2 768
55 - 59 J.	402 822	145	2 778	352 096	110	3 201
60 - 64 J.	315 007	87	3 621	299 208	97	3 085
65 - 69 J.	219 497	61	3 598	222 942	57	3 911
70 - 74 J.	105 647	28	3 773	87 641	24	3 652
75 - 99 J.	201 270	36	5 591	127 947	16	7 997
Insgesamt	4 041 564	1 535	2 633	4 041 429	1 555	2 599
Verwitwete Männer und Frauen						
0 - 49 J.	10 678	5	2 136	39 826	13	3 064
50 - 59 J.	24 339	6	4 057	98 730	25	3 949
60 - 64 J.	17 739	9	1 971	99 229	17	5 837
65 - 69 J.	20 406	6	3 401	166 902	43	3 881
70 - 74 J.	16 410	5	3 282	136 663	38	3 596
75 - 79 J.	37 412	3	5 776	238 233	53	4 495
80 - 99 J.	66 192	13	6 637	297 646	50	5 953
Insgesamt	193 176	47	4 110	1 077 229	239	4 507
Geschiedene Männer und Frauen						
0 - 29 J.	50 307	12	4 192	75 586	19	3 978
30 - 34 J.	63 895	18	3 550	69 871	27	2 588
35 - 39 J.	69 267	14	4 948	76 088	26	2 926
40 - 44 J.	46 025	13	3 540	52 792	16	3 299
45 - 49 J.	59 551	10	5 955	71 213	19	3 748
50 - 54 J.	48 678	10	4 868	62 413	26	2 401
55 - 99 J.	57 361	14	4 097	184 313	50	3 686
Insgesamt	395 084	91	4 342	592 276	183	3 236
nachrichtlich nur für Frauen **)						
55 - 64 J.	.	.	.	80 822	31	2 607
65 - 99 J.	.	.	.	103 491	19	5 447

* Ledige Männer und Frauen wurden nicht direkt in den Hochrechnungsrahmen einbezogen. Sie sind hier nur der Vollständigkeit halber aufgeführt. — ** Diese Unterteilung wurde nur für die Frauen vorgenommen, da die Zellenbesetzung bei den Männern in diesen Altersklassen zu gering war.

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

Übersiedlerwelle im Herbst 1989 nicht sehr verläßlich. Die Auswirkungen der starken Wanderungsbewegungen auf die Haushaltsstrukturen sind schon gar nicht quantifizierbar. Gewiß ist der Anteil von 15,6 vH der Ein-Personenhaushalte im ungewichteten SOEP zu gering.

Wegen der schlechten Datenlage soll die Stichprobe nur an die vom DIW geschätzte Ecksumme von 6 794 Tausend Haushalten angepaßt werden[15]. Tabelle 2 ist 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 bei einer Hochrechnung, die Geschlecht, Familienstand und Alter berücksichtigt, auch der Anteil der Ein-Personenhaushalte steigt. Unter Umständen erübrigt sich wegen dieser 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. Zuverlässigere Informationen über die Haushaltsstruktur würden bei signifikanten Abweichungen eine neue Hochrechnung erfordern[16].

Die Haushaltsgewichte werden auf sämtliche Personen in den Haushalten einschließlich der Kinder unter 16 Jahren übertragen. Auf diese Weise ist die Identität von Haushalts- und Personengewichten immer gewährleistet.

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 1 dargestellt.

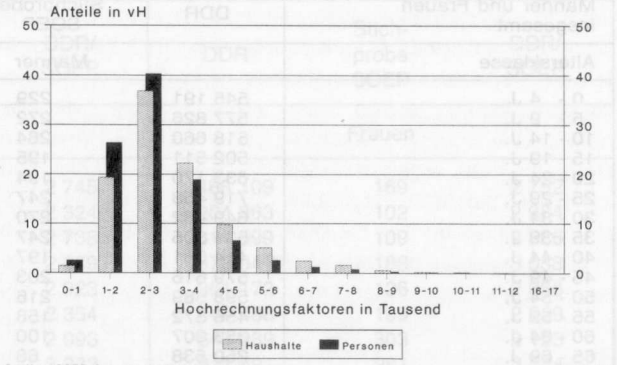
Der niedrigste Wert für einen Hochrechnungsfaktor wurde mit 364 ermittelt. Dies heißt, daß die betreffenden Personen und Haushalte nur mit etwa einem Achtel des Durchschnittsgewichts bei Analysen und Auswertungen berücksichtigt werden. Nur 31 Haushalte (1,4 vH) und 157 Personen (2,6 vH) repräsentieren weniger als 1 000 Mitbürger und 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

Abbildung 1

Verteilung der Hochrechnungsfaktoren

SOEP-Ost Basiserhebung 1990



Quellen: SOEP-Ost, eigene Berechnungen.

DIW 91

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 krass überproportional bewertet. Sechs Haushalte (0,3 vH) mit insgesamt acht Personen (0,1 vH) repräsentieren mehr als 10 000 Haushalte bzw. Personen.

Der weitaus größte Teil der Gewichte streut nicht sehr stark. 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 der Standardabweichung 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	2179	6044
Summe	6793998	16433794
Arithmetisches Mittel	3117,9	2719,0
Standardabweichung	1510,9	1277,1
Variationskoeffizient	0,48	0,47
Minimum	363,9	363,9
Maximum	16798,4	16798,4

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 Stichprobenele-

[15] DIW (1990a).

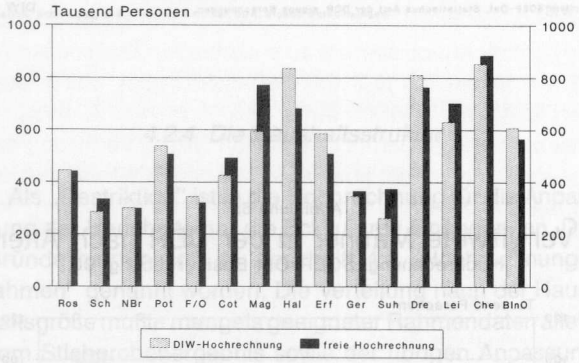
[16] Unter Umständen wird man anhand einer Hochrechnung für die 2. oder gar 3. Welle eine Rückrechnung auf die Basiserhebung vornehmen können. Dies setzt allerdings parallele Mikrozensusergebnisse für die neuen Bundesländer und den Ostteil Berlins voraus.

mente mit dem arithmetischem Mittel der Personengewichte, also mit 2719, multipliziert wird. Im wesentlichen werden die im Kapitel „Hochrechnungsrahmen" bereits diskutierten Eigenschaften graphisch verdeutlicht (Abbildungen 2a und 2b). (Legende siehe S. 63). Offenichtlich ist die demographische Verteilung durch die freie Hochrechnung (schwarze Flächen) gut wiedergegeben. Größere Abweichungen von der „wahren" Verteilung (schraffierte Flächen) zeigen sich nur für die Bezirke Magdeburg, Suhl, Halle und Cottbus.

Abbildung 2a

Männer in der DDR nach Bezirken

Hochrechnung: SOEP-Ost Basiserhebung '90

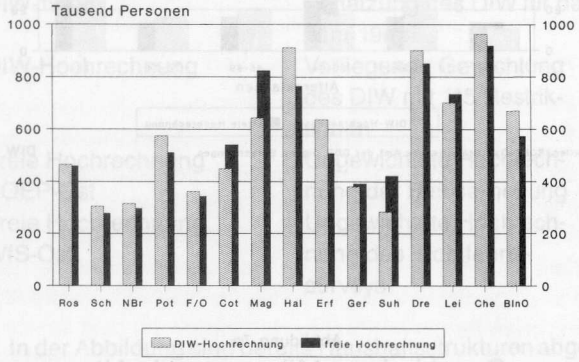


Quellen: SOEP-Ost, Statistisches Amt der DDR, eigene Berechnungen. DIW 91

Abbildung 2b

Frauen in der DDR nach Bezirken

Hochrechnung: SOEP-Ost Basiserhebung '90



Quellen: SOEP-Ost, Statistisches Amt der DDR, eigene Berechnungen. DIW 91

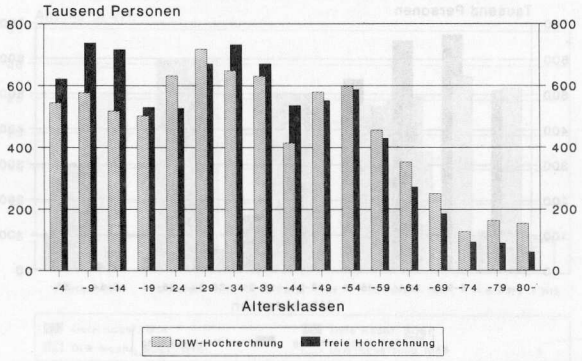
4.2.3 Die demographische Verteilung

Die Abbildungen 3a und 3b geben die Verzerrungen in der Altersstruktur wieder, wenn frei hochgerechnet wird. Auffällig ist, daß der Geburtenausfall nach dem Zweiten Weltkrieg, der sich 1989/90 in der Klasse der 40 bis 44jährigen zeigte, in der Männerstichprobe zu schwach, in der Frauenstichprobe kaum nachgezeichnet wird. Die unterschiedlichen Ausschöpfungsraten der älteren Bürger gegenüber denen der 30 bis 44jährigen werden in der Graphik für die Frauen besonders deutlich.

Abbildung 3a

Männer in der DDR nach Alter

Hochrechnung: SOEP-Ost Basiserhebung '90

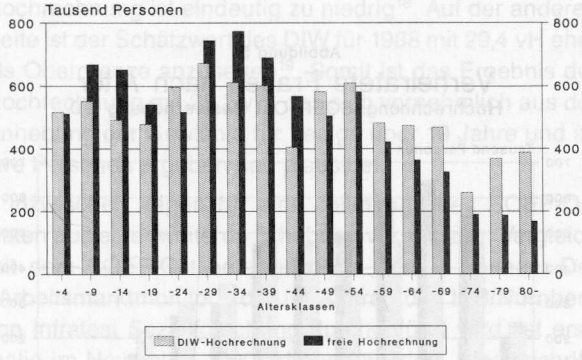


Quellen: SOEP-Ost, Statistisches Amt der DDR, eigene Berechnungen. DIW 91

Abbildung 3b

Frauen in der DDR nach Alter

Hochrechnung: SOEP-Ost Basiserhebung '90



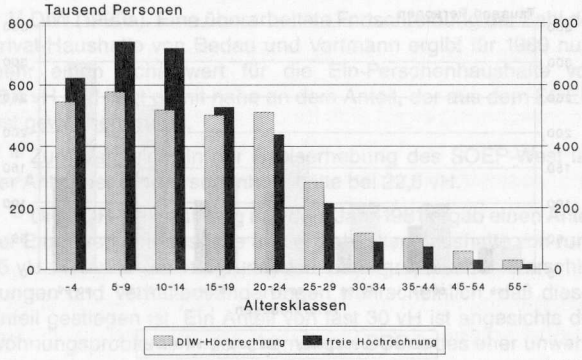
Quellen: SOEP-Ost, Statistisches Amt der DDR, eigene Berechnungen. DIW 91

In den folgenden Abbildungen 4a bis 7b sind Aufgliederungen nach dem Familienstand vorgenommen worden. Auch diese Graphiken erklären sich von selbst. Erwähnt werden soll lediglich noch Abbildung 6b, die die verwitweten Frauen betrifft. Hier sieht man die generelle und die mit dem Alter zunehmende Untererfassung dieser Subpopulation.

Abbildung 4a

Ledige Männer in der DDR nach Alter

Hochrechnung: SOEP-Ost Basiserhebung '90

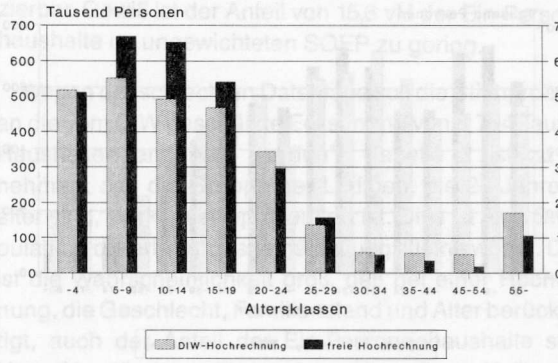


Quellen: SOEP-Ost, Statistisches Amt der DDR, eigene Berechnungen. DIW 91

Abbildung 4b

Ledige Frauen in der DDR nach Alter

Hochrechnung: SOEP-Ost Basiserhebung '90

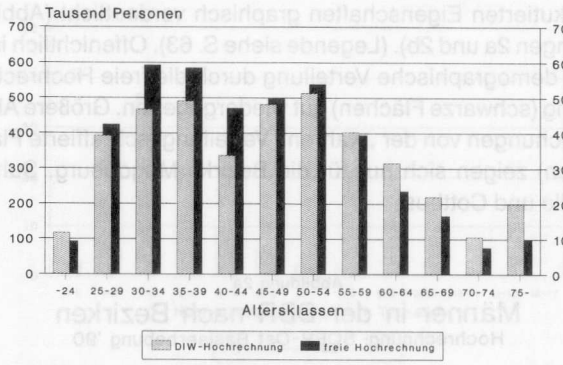


Quellen: SOEP-Ost, Statistisches Amt der DDR, eigene Berechnungen. DIW 91

Abbildung 5a

Verheiratete Männer nach Alter

Hochrechnung: SOEP-Ost Basiserhebung '90

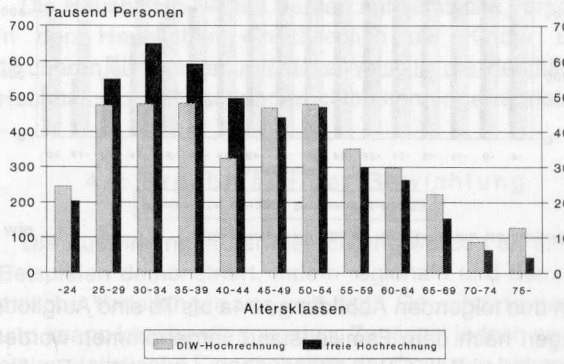


Quellen: SOEP-Ost, Statistisches Amt der DDR, eigene Berechnungen. DIW 91

Abbildung 5b

Verheiratete Frauen nach Alter

Hochrechnung: SOEP-Ost Basiserhebung '90

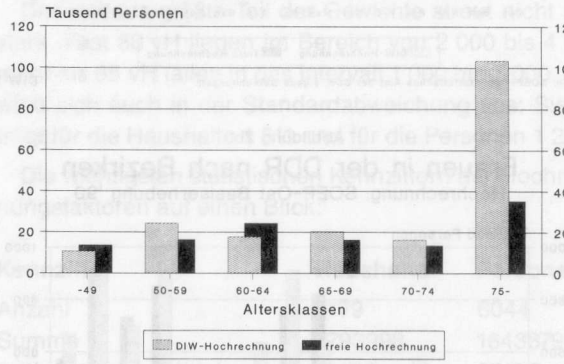


Quellen: SOEP-Ost, Statistisches Amt der DDR, eigene Berechnungen. DIW 91

Abbildung 6a

Verwitwete Männer in der DDR nach Alter

Hochrechnung: SOEP-Ost Basiserhebung '90

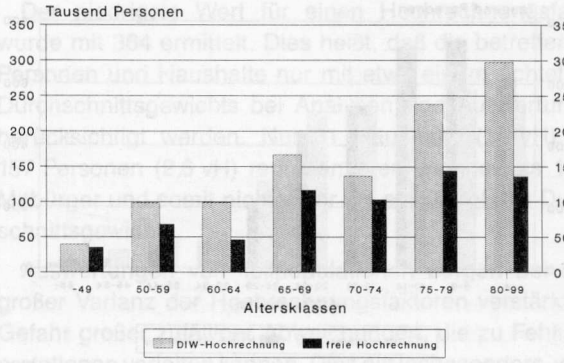


Quellen: SOEP-Ost, Statistisches Amt der DDR, eigene Berechnungen. DIW 91

Abbildung 6b

Verwitwete Frauen in der DDR nach Alter

Hochrechnung: SOEP-Ost Basiserhebung '90

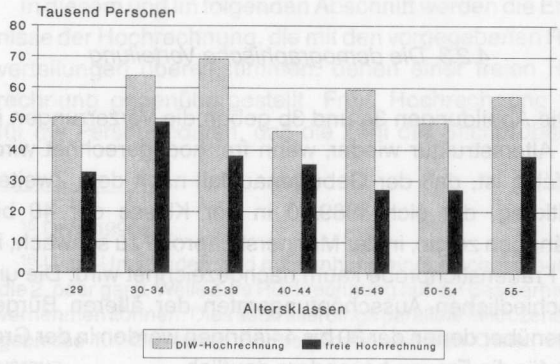


Quellen: SOEP-Ost, Statistisches Amt der DDR, eigene Berechnungen. DIW 91

Abbildung 7a

Geschiedene Männer nach Alter

Hochrechnung: SOEP-Ost Basiserhebung '90

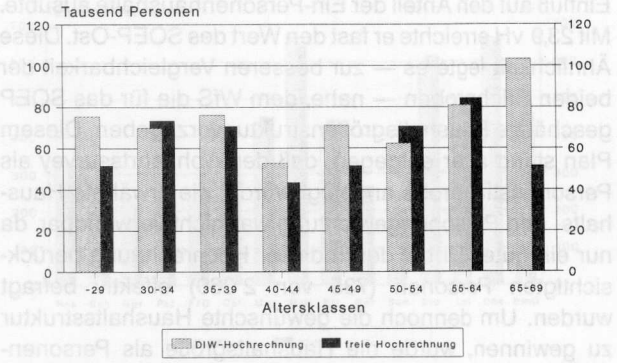


Quellen: SOEP-Ost, Statistisches Amt der DDR, eigene Berechnungen. DIW 91

Abbildung 7b

Geschiedene Frauen nach Alter

Hochrechnung: SOEP-Ost Basiserhebung '90

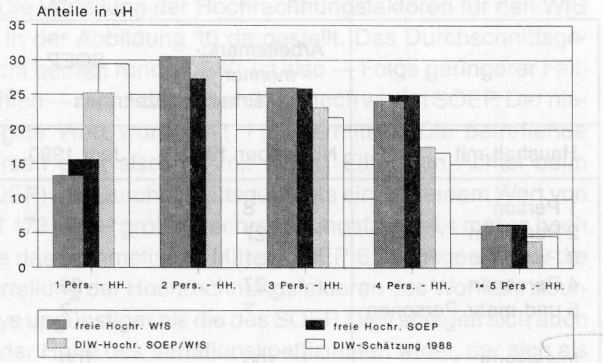


Quellen: SOEP-Ost, Statistisches Amt der DDR, eigene Berechnungen. DIW 91

Abbildung 8

Haushalte nach Größe in der DDR

Ergebnisse verschiedener Schätzungen



Quellen: SOEP-Ost, Wohlfahrtssurvey-Ost, eigene Berechnungen. DIW 91

4.2.4 Die Haushaltsstruktur

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

In Abbildung 8 sind Haushaltsstrukturen dargestellt worden, die sich aus vier verschiedenen Schätzungen ergaben:

DIW-1988	Schätzung des DIW für das Jahr 1988[17]
DIW-Hochrechnung	Vorliegende Gewichtung des DIW mit 115 Restriktionen
Freie Hochrechnung SOEP-Ost	Ungewichtete Hochrechnung der Basiserhebung
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 sowie für die 5 und Mehr-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: Einmal ist die Antreffwahrscheinlichkeit und somit die Antwortwahrscheinlichkeit bei diesen Personen geringer, zum anderen verweigern die häufig sehr jungen oder relativ alten Personen — wenn auch aus unterschiedlichen Gründen — überdurchschnittlich oft die Auskunft.

Der schon genannte Anteil von 15,6 vH aus der freien Hochrechnung ist eindeutig zu niedrig[18]. Auf der anderen Seite ist der Schätzwert des DIW für 1988 mit 29,4 vH eher als Obergrenze anzusehen[19]. Somit ist das Ergebnis der Hochrechnung mit 25,2 vH, das sich vornehmlich aus der Anhebung der Gewichte für Ledige über 19 Jahre und ältere Personen ergeben hat, plausibel.

Inzwischen liegen für eine Teilmenge des SOEP-Ost Daten aus einer weiteren Erhebung vor, die zum Vergleich mit dem SOEP-Ost herangezogen werden können: 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 1926 bis 1974. Mit dieser Befragungsform werden Ein-Personenhaushalte grundsätzlich 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:

---

[17] DIW (1990a). Eine überarbeitete Fortschreibung der Zahl der Privat-Haushalte von Bedau und Vortmann ergibt für 1989 nunmehr einen Schätzwert für die Ein-Personenhaushalte von 25,4 vH und liegt damit nahe an dem Anteil, der aus dem SOEP-Ost gewonnen wurde.

[18] Zum Vergleich: In der Basiserhebung des SOEP-West lag der Anteil der Ein-Personenhaushalte bei 22,6 vH.

[19] Die DDR-Volkszählung aus dem Jahr 1981 ergab einen Anteil der Ein-Personenhaushalte an der Zahl aller Haushalte 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.

Tabelle 3
Personenbezogene Auswertung (in vH aller Personen) der Bevölkerung im erwerbsfähigen Alter in Ostdeutschland nach Haushaltsgröße (gewichtete Auswertung)

	Arbeitsmarkt-monitor	SOEP
	Erhebungszeitraum	
Haushalt mit ...	November 1990	Juni 1990
1 Person	8	7
2 Personen	27	27
3 Personen	31	33
4 Personen	27	26
5 und mehr Personen	7	7
Insgesamt	100	100

Quellen: Infratest Sozialforschung, eigene Berechnungen.

Ersichtlich stimmen beide Verteilungen gut überein; der etwas höhere Anteil der Ein-Personenhaushalte im Monitor 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. Einmal beträgt die Fallzahl des Wohlfahrtssurveys 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 Folge, 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 8). Dieser liegt noch unter dem Wert im SOEP-Ost (15,6 vH). Da im Wohlfahrtssurvey die demographische Struktur des Gesamthaushalts erfaßt wird, war es möglich, analog zum Panel eine simultane Haushalts- und Personengewichtung vorzunehmen, die sich auf Angaben von 2 060 Personen einschließlich Kindern bezog. Dies entspricht immerhin einem Drittel der Größe des SOEP, so daß dessen Hochrechnungsrahmen nahezu identisch übernommen werden konnte[20]. 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 legte es — zur besseren Vergleichbarkeit der beiden Stichproben — nahe, dem WfS die für das SOEP geschätzte Haushaltsgrößenstruktur vorzugeben. Diesem Plan stand aber entgegen, daß der Wohlfahrtssurvey als Personenstichprobe angelegt wurde; die erwähnte Haushalts- und Personengewichtung war nicht verwendbar, da nur ein gutes Drittel der in dieser Hochrechnung berücksichtigten Personen (735 von 2 060) effektiv befragt wurden. Um dennoch die gewünschte Haushaltsstruktur zu gewinnen, wurde die Haushaltsgröße als Personenmerkmal definiert. Dies führte zwar zu dem Ergebnis, daß die Hochrechnung statt der zu erwartenden Zahl von knapp 6,8 Mill. insgesamt 12,7 Mill. Haushalte ergibt, was gerade der Gesamtzahl aller Personen entspricht, die 18 Jahre und älter sind. Die Struktur der erzielten Größenverteilung der Haushalte 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 gegliedert nach

Kindern (nicht arbeitsfähiges Alter),

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

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

In den Abbildungen 9a und 9b sind die Schätzungen den entsprechenden Daten aus dem Hochrechnungsergebnis des SOEP gegenübergestellt. (Legende siehe S. 63).

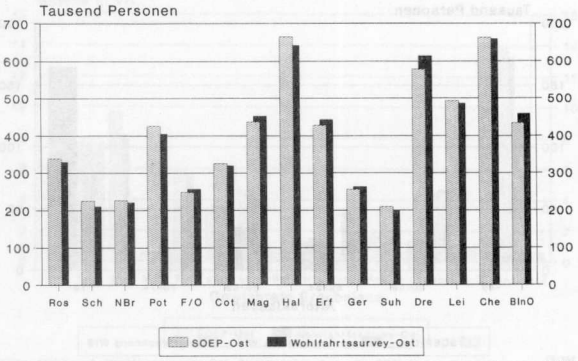
Fallzahlprobleme ergaben sich bei der Anpassung an *Alter, Familienstand und Geschlecht.* Zwar konnte für die Frauen und Männer insgesamt die feingliedrige Abstufung in Fünfer-Klassen im Altersbereich von 25 bis 69 Jahren beibehalten werden, die jüngeren und älteren Jahrgänge mußten aber zu einer Klasse zusammengefaßt werden. Die Verteilung nach dem Familienstand bereitete bei den Män-

[20] Es mußten lediglich bei den Verwitweten und Geschiedenen Altersklassen zusammengefaßt werden.

Abbildung 9a

Vollj. Männer in der DDR nach Bezirken

SOEP-Ost vs. Wohlfahrtssurvey-Ost '90

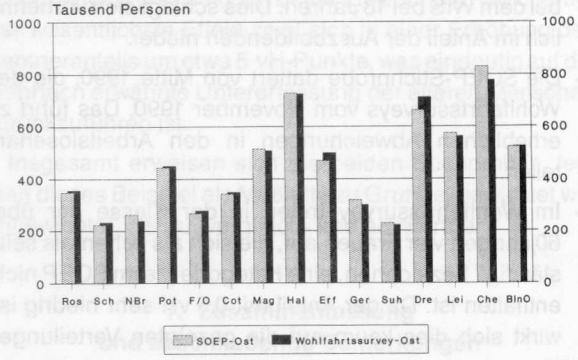


Quellen: SOEP-Ost, Wohlfahrtssurvey-Ost, Statistisches Amt der DDR, eigene Berechnungen. DIW 91

Abbildung 9b

Vollj. Frauen in der DDR nach Bezirken

SOEP-Ost vs. Wohlfahrtssurvey-Ost '90



Quellen: SOEP-Ost, Wohlfahrtssurvey-Ost, Statistisches Amt der DDR, eigene Berechnungen. DIW 91

nern 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 dem Familienstand für die Frauen. Verheiratete Frauen konnten in sechs , verwitwete und geschiedene Frauen in jeweils drei Altersklassen gegliedert werden.

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.

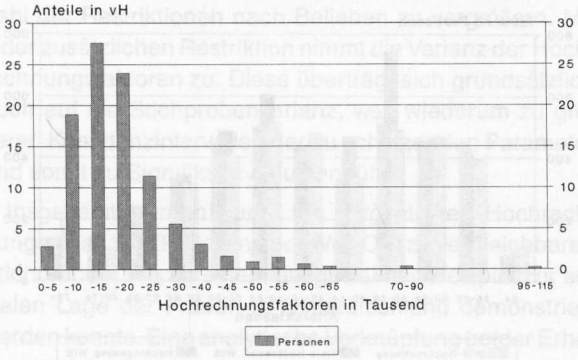
5.2.1 Statistische Eigenschaften der Hochrechnungsfaktoren

Die Verteilung der Hochrechnungsfaktoren für den WfS ist in der Abbildung 10 dargestellt. Das Durchschnittsgewicht beträgt rund 17 280, ist also — Folge geringerer Fallzahlen — mehr als sechsmal so hoch wie im SOEP. Der niedrigste Wert wurde mit 1 550 ermittelt. Die betreffende Person geht also nur mit einem Elftel (ein Achtel beim SOEP) des Durchschnittsgewichts 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 Wohlfahrtssurveys 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).

Abbildung 10

Verteilung der Hochrechnungsfaktoren

Wohlfahrtssurvey-Ost '90



Quellen: Wohlfahrtssurvey-Ost, Statistisches Amt der DDR, eigene Berechnungen. DIW 91

Die wichtigsten statistischen Kennziffern der Hochrechnungsfaktoren auf einen Blick:

Kennziffer	*Personen*
Anzahl	735
Summe	12700622
Arithmetisches Mittel	17279,76
Standardabweichung	11754,57
Variationskoeffizient	0,68
Minimum	1550
Maximum	111173

5.2.2 Zwei Beispiele zur demographischen Verteilung

Die bedeutend geringeren Fallzahlen des Wohlfahrtssurveys führten bei der Gliederung nach dem Familienstand zu wesentlich weniger Altersklassen als beim SOEP. Dies galt besonders für die Männer. So gab es für verheiratete Männer nur drei Altersklassen, Witwer wurden alters-

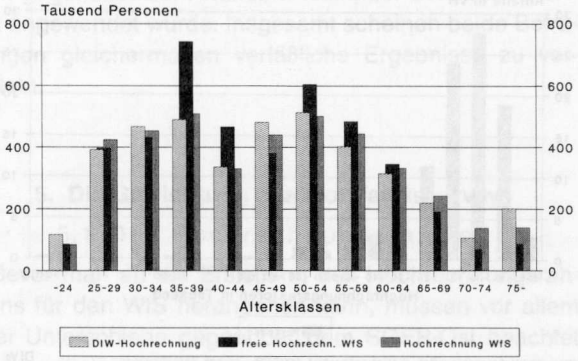
mäßig nicht unterschieden. Es ist reizvoll zu sehen, welche Alterstrukturen sich für diese Subpopulationen nach der Hochrechnung im Vergleich zu den Vorgaben durch das Statistische Amt der DDR bzw. durch das SOEP ergeben.

Abbildung 11 zeigt die Altersverteilung der verheirateten Männer in Ostdeutschland. Die Werte im SOEP sind dieselben wie in Abbildung 5a. 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-Panels 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 Jahren ablesen läßt, deren Zahl um 46 vH zu hoch erfaßt wurde. Vielmehr gaben hierfür die — dem SOEP ähnlich — stark disaggregierten Restriktionen für die Männer insgesamt den Ausschlag.

Abbildung 11

Verheiratete Männer nach Alter

SOEP-Ost vs. Wohlfahrtssurvey-Ost '90



Quellen: SOEP-Ost, Wohlfahrtssurvey-Ost, Statistisches Amt der DDR, eigene Berechnungen. DIW 91

Als Gegenstück zu dem recht erfreulichen Beispiel ist die Altersverteilung der Witwer zu nennen: Abbildung 12 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. Im Ergebnis ist dies aber für Auswertungen nicht hinderlich, da die Gruppe der Witwer wegen der kleinen Fallzahlen ohnehin nicht auswertbar ist.

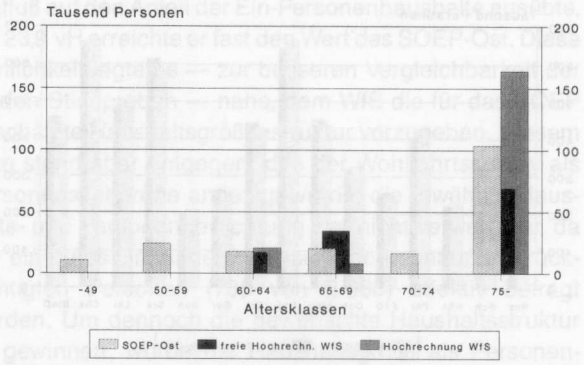
6. Soziale Stellung in SOEP und Wohlfahrtssurvey Ein empirisches Beispiel

Am Beispiel der sozialen Lage der Frauen in der (ehemaligen) DDR soll die Vergleichbarkeit der behandelten Stichproben geprüft werden. In den Abbildungen 13a und 13b sind die Sozialstrukturen vor und nach Gewichtung dargestellt[21]. (Legende siehe S. 63). Ein Vergleich der

Abbildung 12

Verwitwete Männer nach Alter

SOEP-Ost vs. Wohlfahrtssurvey-Ost '90



Quellen: SOEP-Ost, Wohlfahrtssurvey-Ost, Statistisches Amt der DDR, eigene Berechnungen. DIW 91

beiden Erhebungen erfordert zunächst den Hinweis auf einige inhaltliche Unterschiede:

— Die Altersuntergrenze liegt beim SOEP bei 16 Jahren, bei dem WfS bei 18 Jahren. Dies schlägt sich vornehmlich im Anteil der Auszubildenden nieder.

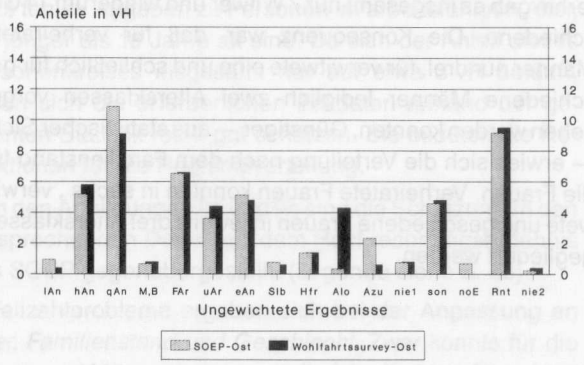
— Die SOEP-Stichprobe datiert von Mitte 1990, die des Wohlfahrtssurveys vom November 1990. Das führt zu erheblichen Abweichungen in den Arbeitslosenanteilen.

— 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 0,5 vH sehr niedrig ist, wirkt sich dies kaum auf die gezeigten Verteilungen aus.

Abbildung 13a

Soziale Lagen der Frauen in der DDR

SOEP-Ost vs. Wohlfahrtssurvey-Ost '90



Quellen: SOEP-Ost, Wohlfahrtssurvey-Ost, Statistisches Amt der DDR, eigene Berechnungen. DIW 91

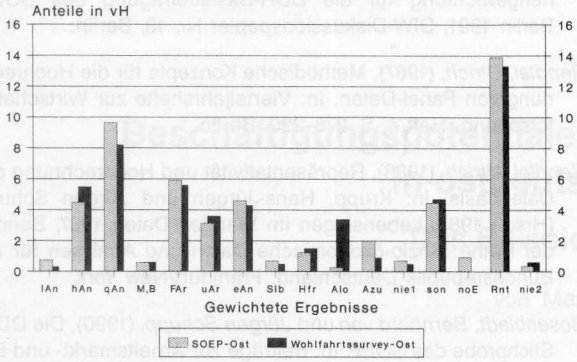
Der Vergleich beider Abbildungen zeigt, gewichtet wie ungewichtet, erstaunliche Übereinstimmungen. Relativ hohe Abweichungen sind nur bei den bereits erwähnten

[21] Die Daten für diesen Vergleich stellte freundlicherweise Detlef Landua vom Wissenschaftszentrum für Sozialforschung, Berlin, zur Verfügung.

Abbildung 13b

Soziale Lagen der Frauen in der DDR

SOEP-Ost vs. Wohlfahrtssurvey-Ost '90



Quellen: SOEP-Ost, Wohlfahrtssurvey-Ost, Statistisches Amt der DDR, eigene Berechnungen. DIW 91

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 sich in einer Erhöhung des Rentneranteils um etwa 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 Meßlatte zu Grunde, gewichtet wie ungewichtet, als gut vergleich- und auswertbar.

7. Zusammenfassung und abschließende Bemerkungen

In diesem Beitrag wurden konsistente Hochrechnungen und Gewichtungen für das Sozio-Ökonomische Panel-Ost und die Pilotstudie des Wohlfahrtssurvey-Ost vorgestellt. Grundgedanke für die Anlage der Gewichtungsverfahren war, die repräsentativ angelegten 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 ist. Diese Bedingungen waren beim SOEP 114 Restriktionen zur Personenverteilung nach Region, Geschlecht, Familienstand und Alter. Nur eine Restriktion — die Zahl der 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 Personengewichtung 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 zu vergrößern. Mit jeder zusätzlichen Restriktion nimmt die Varianz der Hochrechnungsfaktoren zu. Diese überträgt sich grundsätzlich auch auf die Stichprobenvarianz, was wiederum zu größeren Konfidenzintervallen der zu schätzenden Parameter und somit zu Signifikanzverlusten führt.

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.

Legende für die Abbildungen 2a, 2b, 9a und 9b:

Ros	Rostock
Sch	Schwerin
NBr	Neu-Brandenburg
Pot	Potsdam
F/O	Frankfurt/Oder
Cot	Cottbus
Mag	Magdeburg
Hal	Halle
Erf	Erfurt
Suh	Suhl
Ger	Gera
Dre	Dresden
Lei	Leipzig
Che	Chemnitz
BlnO	Berlin(Ost)

Legende für die Abbildungen 13a und 13b:

Frauen bis 60 Jahre:

lAn	Leitende Angestellte
hAn	Hochqual. Angestellte / Höhere Beamte
qAn	Qualifizierte Ang. / Gehob. Beamte
M,B	Meister oder Brigadier
FAr	Facharbeiter
uAr	Un- und angelernter Arbeiter
eAn	Einfache Angestellte / Beamte
Slb	Selbständige
Hfr	Hausfrauen
Alo	Arbeitslose
Azu	Auszubildende
nie1	Noch nie Erwerbstätige
son	sonstige nicht Erwerbstätige

Frauen 61 Jahre und älter:

noE	noch Erwerbstätige
Rnt	Rentner
nie2	Noch nie Erwerbstätige

Literaturverzeichnis

DIW (1990), Privathaushalte und Wohnungsbedarf in Deutschland bis zum Jahr 2000. Bearb.: Bartholmai, Bernd, Manfred Melzer und Erika Schulz. In: Wochenbericht des DIW, Nr. 42/1990, S. 591 ff.

DIW (1990a), Die Einkommensverteilung nach Haushaltsgruppen in der ehemaligen DDR. Bearb.: Bedau, Klaus-Dietrich und Heinz Vortmann. In: Wochenbericht des DIW, Nr. 47/1990, S. 655 ff.

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.

Rosenbladt, 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: Vierteljahrshefte 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.