



DIW Berlin

Deutsches Institut
für Wirtschaftsforschung

Data Documentation 22



Rainer Pischner

**Die Querschnittsgewichtung und die
Hochrechnungsfaktoren des
Sozio-oekonomischen Panels (SOEP)
ab Release 2007 (Welle W)**

Modifikationen und Aktualisierungen

IMPRESSUM

© DIW Berlin, 2007

DIW Berlin
Deutsches Institut für Wirtschaftsforschung
Königin-Luise-Str. 5
14195 Berlin
Tel. +49 (30) 897 89-0
Fax +49 (30) 897 89-200
www.diw.de

ISSN 1861-1532

All rights reserved.
Reproduction and distribution
in any form, also in parts,
requires the express written
permission of DIW Berlin.



Data Documentation 22

Rainer Pischner*

Die Querschnittsgewichtung und die Hochrechnungsfaktoren des Sozio-oekonomischen Panels (SOEP) ab Release 2007 (Welle W)

- Modifikationen und Aktualisierungen -

Berlin, 25. Juni, 2007

* Längsschnittstudie SOEP am DIW Berlin. rpischner@diw.de

Inhaltsverzeichnis

1 Hintergrund und Zielsetzung	2
2 Die Stichproben im SOEP.....	3
3 Zum Konzept der Hochrechnung des SOEP	4
4 Die Startgewichtungen der Stichproben A – H	6
4.1 Stichprobe A und B (1984).....	8
4.2 Stichprobe C (1990).....	9
4.3 Stichprobe D (1994/95).....	9
4.4 Stichprobe E (1998)	10
4.5 Stichprobe F (2000)	10
4.6 Stichprobe G (2000).....	11
4.7 Stichprobe H (2006).....	11
5 Änderungen der Hochrechnung ab Erhebungsjahr 2005	12
5.1 Angaben für das Land Berlin nicht mehr nach West und Ost getrennt.	12
5.2 Änderungen bei Zahl und Art der Hochrechnungsfaktoren.....	13
5.3 Leichte Modifizierung des Hochrechnungsrahmens rückwirkend für die Erhebungsjahre 2005 und später.....	15
6 Querschnittshochrechnung für Personen.....	18

Verzeichnis der Tabellen

Tabelle 1-1 Designgewichte im Sozio-oekonomischen Panel	7
Tabelle 2 -1 Hochrechnungsfaktoren für Privathaushalte im Sozio-oekonomischen Panel für die Wellen A – W (1984 – 2006) – In den Gewichten enthaltene Stichproben und Eckdaten –	15
Tabelle 3 –1 Stichprobenspezifischer Eckwerte zur Querschnittshochrechnung der Privathaushalte für die Jahre 2005 und 2006 (vorläufig)	17
Tabelle 4-1 Stichprobenunabhängige Eckwerte zur Querschnittshochrechnung der Privathaushalte für die Jahre 2005 und 2006 (vorläufig)	18
Tabelle 5-1 Hochrechnungsfaktoren für Personen in Privathaushalten im Sozio-oekonomischen Panel für die Wellen A – W (1984 – 2006) – In den Gewichten enthaltene Stichproben und Eckdaten –	19
Tabelle 5-2 Eckwerte zur Querschnittshochrechnung der Personen in Privathaushalten für die Jahre 2005 und 2006 (vorläufig)	20

Zusammenfassung

Mit der Datenweitergabe für Release 2007 ergeben sich für die Querschnittsgewichtung des SOEP mehrere Änderungen und Neuerungen. Diese betreffen folgende Punkte:

- Einbeziehung der Ergänzungsstichprobe H.
- Zuordnung von Berlin (West) zu den neuen Ländern.
- Leichte Modifizierung des Stichprobenrahmens.
- Konstruktion von wellenspezifischen Gewichtungs- und Hochrechnungsvariablen, bei denen die Gewichte der Haushalte neuer Stichproben auf Null gesetzt sind, um Probleme, die bei Erstbefragung komplexer Sachverhalte entstehen können, zu vermeiden. Damit werden insbesondere bei Betrachtung von Zeitreihen „Sprünge“ abgemildert.

Einen Überblick über die Haushalts- und Personengewichte des SOEP für deskriptive Querschnittsanalysen wird tabellarisch gegeben.

Abstract

With SOEP's 2007 data release, a number of changes and additions have been made in the cross-sectional weighting scheme. They apply to the following:

- Inclusion of refreshment sample H.
- Classification of West Berlin as belonging to the "new" federal states (of former East Germany).
- Slight modification of the sample framework.
- Construction of wave-specific weighting variables, where the weights for households from new samples are set to zero in order to avoid problems that may arise when asking respondents about complex issues in their first survey year. This helps in particular in avoiding discontinuities between time series.

An overview of the household and individual weights used in SOEP for representative cross-sectional analyses are provided in tabular form.

JEL Classification: C81, C19, Z13

Keywords: Gewichtung, Weighting, Hochrechnung, Querschnitt, cross-sectional, Stichproben, sampling, SOEP

1 Hintergrund und Zielsetzung

Mit der Datenweitergabe für das Jahr 2006 (Welle W, Release 2007) ergeben sich für die Querschnittshochrechnungsfaktoren bzw. Querschnittsgewichte¹ zwei wesentliche Neuerungen und Änderungen: Zum einen ist das SOEP im Erhebungsjahr 2006 um eine weitere Stichprobe – die Ergänzungsstichprobe H – erweitert worden („Refreshment Sample“). Die 1506 neuen befragten Haushalte haben an der Gesamtstichprobe (12499 Haushalte) einen Anteil von ca. 12 %. Diese neuen Haushalte, die repräsentativ für die gesamte Bundesrepublik Deutschland sind, waren in die Hochrechnung zu integrieren.

Weiterhin weist seit dem Jahr 2005 das Statistische Bundesamt Angaben für das Land Berlin nicht mehr nach West und Ost getrennt aus. Dies führt zu leichten Modifizierungen des Hochrechnungsrahmens für Haushalte; rückwirkend seit 2005.

Die Neuerungen wurden zusätzlich zum Anlass genommen, Zahl und Art der standardmäßig weitergegebenen Hochrechnungsfaktoren zu modifizieren. Dadurch soll die Nutzung der Standard-Hochrechnungsfaktoren deutlich erleichtert werden. Eine sehr wichtige Neuerung ist die Konstruktion der von Hochrechnungsvariablen, bei denen die Gewichte von neuen Teilstichproben auf Null gesetzt sind. Hintergrund: In den ersten Wellen weisen die Befragten bei komplexen Erhebungskonstrukten ein „schlechteres“ Antwortverhalten auf als in späteren Wellen (z. B. bezüglich Lebenszufriedenheit und Jahreseinkommen). Eine Ausnahme bildet hier die Stichprobe C. Befragte in der DDR wiesen in ihrem Antwortverhalten im Jahr 1990 nicht die üblichen Probleme bei einer Erstbefragung auf (vgl. Abschnitt 4.2).

Zum Aufbau dieser Dokumentation: Abschnitt 1 gibt einen kurzen Überblick über die bisher verfügbaren Stichproben des sozio-oekonomischen Panels (SOEP). Allein wegen der komplizierten Struktur des SOEP und des unterschiedlichen Ziehungsdesign der diversen Teilstichproben ist für deskriptive Analysen eine Design-Hochrechnung unerlässlich. Im Abschnitt 2 wird das verwendete Hochrechnungskonzept zusammenfassend vorgestellt. Damit, wie die jeweils ersten Wellen der acht Teilstichprobenproben grundsätzlich in die Gewichtung einbezogen wurden, befasst sich der Abschnitt 3. Änderungen und Neuerungen bei der Querschnittsgewichtung, die seit dem Erhebungsjahr 2005 erforderlich waren, sich aber teilweise auf den gesamten Erhebungszeitraum bezogen, finden sich im Abschnitt 4. Die Logik und

¹ Die Begriffe Hochrechnungsfaktoren und Gewichte werden synonym verwendet. Wenn nicht explizit erwähnt, sind immer Querschnittsgewichte gemeint.

Struktur der Querschnittsgewichte für Haushalte wird kompakt in Tabellen zusammengefasst. Diese finden sich auch für Personen im Abschnitt 5.

2 Die Stichproben im SOEP

Es ist zweckmäßig, zunächst eine kurze Übersicht über die Stichproben des Sozio-oekonomischen Panels und ihre Einbeziehung in die Hochrechnung zu geben.

Mit der Ergänzungsstichprobe H umfasst das Sozio-oekonomische Panel (SOEP), das 1984 startete, nunmehr acht Stichproben und – für die Stichproben A und B - 23 Wellen (1984 – 2006).² Repräsentiert werden:

Stichprobe A: Deutsche Haushalte³ der Bundesrepublik Deutschland (Start 1984)

Stichprobe B: Ausländische Haushalte⁴ der Bundesrepublik Deutschland (Start 1984)

Stichprobe C: Haushalte der DDR (Start 1990).

Stichprobe D: Zuwandererhaushalte (Start 1994/95)

Stichprobe E: Haushalte in Deutschland, Ergänzungsstichprobe (Start 1998)

Stichprobe F: Haushalte in Deutschland, Ergänzungsstichprobe (Start 2000)

Stichprobe G: Hocheinkommenshaushalte in Deutschland, Hocheinkommensstichprobe (Start 2002)

Stichprobe H: Haushalte in Deutschland, Ergänzungsstichprobe (Start 2006)

Insbesondere für eher deskriptiv angelegte Analysen ist eine Hochrechnung für den SOEP-Datensatz unbedingt erforderlich, da die Ziehungswahrscheinlichkeiten der einzelnen Teilstichproben sich vom Design her unterscheiden (d. h. die jeweiligen Beobachtungen in den einzelnen Teilstichproben unterschiedlich viele Personen / Haushalte in der jeweiligen Grundgesamtheit repräsentieren). Hinzu kommt ein unterschiedliches Befragtenverhalten nach Welle 1 (selektive Attrition), wobei ebenfalls Unterschiede zwischen den Teilstichpro-

² Eine aktuelle Beschreibung des SOEP findet sich bei Wagner, Frick, Schupp (2007).

³ Genauer: Haushalte, deren Haushaltsvorstand nicht türkischer, italienischer, jugoslawischer, griechischer oder spanischer Nationalität ist

⁴ Genauer: Haushalte, deren Haushaltsvorstand türkischer, italienischer, jugoslawischer, griechischer oder spanischer Nationalität ist

ben bestehen (z.B. aufgrund höherer Emigrations Wahrscheinlichkeiten in den Stichproben B und D).

3 Zum Konzept der Hochrechnung des SOEP

Zunächst seien zum besseren Verständnis der Hochrechnung einige allgemeine Anmerkungen zum Hochrechnungskonzept des SOEP gemacht.

Die Gesamtstichprobe des SOEP ist – wie Abschnitt 1 kurz gezeigt hat – ungewöhnlich komplex. Ihre Gewichtung und Hochrechnung kann nur auf Grundlage eines konsistenten, theoretisch abgesicherten Konzeptes erfolgen. Es ist nicht möglich, an dieser Stelle detailliert hierauf einzugehen. Es erfolgt lediglich eine Beschreibung der wesentlichen Komponenten dieses Konzeptes mit entsprechenden Literaturhinweisen. Außerdem werden beispielhaft die Gewichtungen der bisherigen Teilstichproben beschrieben.⁵

Die Schätzung von Gewichten bzw. Hochrechnungsfaktoren erfolgt – wie international üblich – nach dem Ansatz von Horvitz und Thompson (1952) prinzipiell über den Kehrwert der Auswahlwahrscheinlichkeit der jeweiligen Stichprobeneinheit. Hochrechnungsfaktoren unterscheiden sich von Gewichten lediglich durch einen Skalarmultiplikator. Während die nach Horvitz und Thompson berechnete Summe der Hochrechnungsfaktoren der Zahl der Einheiten der Grundgesamtheit entspricht, ist die Summe der Gewichte gleich der Fallzahl der Stichprobe.

Horvitz und Thompson stellten ihr Konzept indes nur für reine Querschnittsdaten vor. Stichprobeneinheiten sind beim SOEP primär die Haushalte, sekundär die Personen, die in diesen Haushalten leben. Der Ansatz von Horvitz und Thompson wurde von Galler (1987) für Paneldaten weiterentwickelt, indem die Schätzung von Auswahlwahrscheinlichkeiten für die Startwelle um die Schätzung der Verbleib- bzw. Antwortwahrscheinlichkeit für die folgenden Wellen erweitert wurde. Hierzu muss zunächst die Wahrscheinlichkeit einer erneuten Kontaktaufnahme, darauf folgend die Wahrscheinlichkeit einer erneuten Antwortgewährung bestimmt werden.⁶ Eng verknüpft mit dieser Aufgabe ist die Analyse des Ausfallverhaltens der

⁵ Die Frage, ob überhaupt eine Stichprobe hochgerechnet werden soll, wird hier nicht diskutiert werden. Siehe zu diesem Thema Rendtel und Pötter (1992)

⁶ Siehe hierzu Galler (1987) und Rendtel (1995).

Stichprobeneinheiten von Welle zu Welle.⁷ Dieses Konzept ermöglicht es, sowohl Querschnittsgewichte ab Welle 2 jeder Stichprobe sowie Längsschnittgewichte theoretisch konsistent und praktisch zufriedenstellend zu ermitteln.

Der Gewichtungprozess im SOEP stellt sich somit – pro Teilstichprobe - vereinfacht wie folgt dar⁸:

Nur Welle 1:

Ermittlung der Startgewichte neuer Teilstichstichproben.

Ab Welle 2:

Schätzung der Bleibe- und Antwortwahrscheinlichkeiten der verbliebenen Haushalte, über die vorläufige Gewichte der aktuellen Welle bestimmt werden.

Schätzung von Startgewichten für seit der Vorwelle neu entstandene Haushalte in den Alt-Stichproben.

Randanpassung an Eckgrößen des aktuellen Mikrozensus

1985 bis 1989: Zahl der Privathaushalte in der Bundesrepublik Deutschland insgesamt

1990 bis 1999: Getrennt nach alten und neuen Ländern:

Haushaltsgröße (5 Kategorien)

Alter der Personen in Privathaushalten (3 Kategorien)

Geschlecht der Personen in Privathaushalten (2 Kategorien)

Nationalität der Personen in Privathaushalten (2 Kategorien)

Seit 2000: Getrennt nach alten und neuen Ländern:

Haushaltsgröße (5 Kategorien)

Alter der Personen in Privathaushalten (17 Kategorien)

Geschlecht der Personen in Privathaushalten (2 Kategorien)

Nationalität der Personen in Privathaushalten (2 Kategorien)

⁷ Siehe hierzu z.B. Pannenberg (2000) sowie Kroh und Spiess (2006) .

⁸ Vgl. hierzu auch Rendtel (1995).

Man beachte, dass die aufgeführten Personenmerkmale (Alter, Geschlecht und Nationalität) als Eigenschaften der Haushalte anzusehen sind. So wird vermieden, dass sich temporäre Ausfälle von einzelnen Haushaltsmitgliedern auf die Haushaltshochrechnung auswirken.

Eine Übersicht zur Hochrechnung des SOEP, die nicht allzu sehr in Details geht, findet sich in Haisken-De New und Frick (2006) .

4 Die Startgewichtungen der Stichproben A – H

Querschnittsgewichtungen hängen vom Erhebungsdesign einer Stichprobe und dem Teilnehmerverhalten der zufällig ausgewählten Erhebungseinheiten ab. Diese Charakteristika variieren jedoch recht stark über die diversen SOEP-Stichproben und deshalb weisen auch die Startgewichtungen unterschiedlich Konzepte im Detail auf. Dennoch ist das Gerüst jeder Startgewichtung im Prinzip gleich:

Zunächst werden sogenannte Designgewichte (Kehrwerte der designbedingten Auswahlwahrscheinlichkeiten) für die befragten Haushalte erstellt, die weitgehend den durchgeführten Stichprobenplan widerspiegeln sollen. Diese Designgewichte berücksichtigen unterschiedliche Auswahlwahrscheinlichkeiten; im Falle des SOEP z.B. die Ziehung nach Nationalität und unterschiedliche Teilstichproben. Nicht berücksichtigen können sie Nichtteilnahme von Haushalten, z.B. wegen Nicht-Antreffens oder Verweigerung.

Tabelle 3-1 gibt die Designgewichtung für die einzelnen Stichproben und darin wiederum enthaltenen Teilstichproben wider. Man erkennt zum Beispiel, dass Subsample B aus fünf nach Nationalität des Haushaltsvorstands abgegrenzten Teilstichproben besteht. Da Stichprobe B in Abweichung zu allen anderen Subsamples personen- und nicht haushaltsbezogen gezogen wurde, musste ein haushaltsspezifisches „Redressment“ durchgeführt werden. Dadurch entstand in den einzelnen Schichten Varianz. Da andere Teilstichproben z. T. auch in Schichten zerlegt wurden (D, F und G) gibt es auch dort Varianz in den Designgewichten.⁹

⁹ Vgl. w.u. Abschnitt 3.1. sowie die Kurzdokumentation von Spiess und Kroh (2007).

Tabelle 1–1

Designgewichte im Sozio-oekonomischen Panel

Stichprobe/ Teilstichprobe	Kurz-Beschreibung der Stichprobe (Startjahr)	Kennziffern für die Designgewichte von Haushalten mit <i>positiven</i> Querschnittsgewichten		
		Mittelwert	Standardabweichung	Fallzahl im Startjahr
A	Deutsch (1984)	3344	0	4528
B	Ausländisch (1984)	546	305	1393
B1	Türkisch	772	332	397
B2	Italienisch	624	250	194
B3	Griechisch	355	158	196
B4	Jugoslawisch	487	211	306
B5	Spanisch	261	117	200
C	DDR (1990)	1900	0	2179
D	Zuwanderer (1994/1995)	3279	331	252 *
D1	Übersiedler	2946	0	127 *
D2	Aussiedler	3405	0	44 *
D3	Sonstige	3693	0	81 *
E	Ergänzung (1998)	19081	0	1056
F	Ergänzung (2000)	3406	402	6052
F1	Deutsch	3519	0	5607
F2	Ausländisch	1980	0	445
G	Hocheinkommens- haushalte mit Haus- haltsnettoeinkommen > 4500 € (2000)	924	148	998 *
G1	West: < 5113 €	754	0	483 *
G2	Ost: < 5113 €	638	0	55 *
G3	West: >= 5113 €	525	0	419 *
G4	Ost: >= 5113 €	490	0	41 *
H	Ergänzung (2006)	10432	0	1506

* Folgende Stichproben enthalten weitere Haushalte ohne positive Gewichte für die Querschnittsgewichtung:

122 Haushalte in D, davon 49 in D1, 37 in D2, 36 in D3. Diesen Haushalten konnten kein Designgewicht zugeordnet werden, da ihre Auswahlwahrscheinlichkeit nicht zu bestimmen ist. (vgl. Abschnitt 3.3)

226 Haushalte in G, davon 152 in G1, 23 in G2, 43 in G3 und 8 in G4. Diese Haushalte überschritten in der zweiten Welle nicht die vorgeschriebene Einkommensgrenze von 4500 € (vgl. Abschnitt 3.6)

Quelle: Das Sozio-oekonomische Panel. Release 2007; eigene Berechnungen.

Um differentiellen Non-Response in Welle 1 auszugleichen (d. h. unterschiedliche hohe Responderaten nach sozio-oekonomischen Merkmalen), wird auf Basis der Designgewichte eine Randanpassung an wesentliche Eckdaten der amtlichen Statistik oder an andere Merkmale auf Haushaltsbasis vorgenommen. Diese Randanpassung erfolgt nach dem Prinzip des minimalen Informationsverlusts.¹⁰

Ausgehend von den so ermittelten Haushaltsgewichten werden Personengewichte bestimmt. Da grundsätzlich sämtliche Personen, ab dem Kalenderjahr, in dem sie 17 Jahre alt werden, an der Befragung teilnehmen sollen, gilt für repräsentative Stichproben prinzipiell die Gleichheit von Haushalts- und Personengewicht. D.h. die geschätzten Haushaltsgewichte können 1:1 auf sämtliche Personen des Haushalts (einschl. der Kinder) übertragen werden.

4.1 Stichprobe A und B (1984)

Die Hochrechnungen der Stichproben A und B unterscheiden sich grundsätzlich in der Bestimmung der Designgewichte, da die Ausländerhaushalte (Stichprobe B) über Personenregister gezogen worden sind, was zwangsläufig zu einer Überrepräsentation großer Haushalte führte. Diese Haushalte hatten mehrfache Chancen, via eines der Haushaltsmitglieder in die Stichprobe zu gelangen. Stichprobe B musste deshalb einem sog. Redressment unterzogen werden. Diese Designgewichte wurden in einem weiteren vorbereitenden Schritt an die regionale Verteilung der gezogenen Adressen angepasst. Auch die Ausschöpfungsquoten der Sample-Points wurden berücksichtigt.¹¹

Die Randanpassung beider Stichproben erfolgte an 316 Merkmale. Diese beruhen auf interpolierten Daten der Mikrozensus 1982 und 1985 sowie der EG-Arbeitskräfte-Stichprobe des Jahres 1984. Einerseits wurden die Stichproben A und B sehr detailliert an soziodemographische Eckzahlen angepasst, andererseits ging diese Genauigkeit zu Lasten einer großen Varianz der Hochrechnungsfaktoren.

Diesem – zu Beginn des SOEP nicht klar erkannten – Effizienzproblem wurde bei den folgenden Startgewichtungen Rechnung getragen, die sparsamer an externe Ränder angepasst wurden.

¹⁰ Siehe hierzu Merz (1983). Die Randanpassungen wurden durchgeführt mit der MS-DOS-Version des Programms ADJUST von Joachim Merz, Universität Lüneburg.

¹¹ Zur Hochrechnung der Stichproben A und B siehe Pischner (1994)

4.2 Stichprobe C (1990)

Im Juni 1990 – also nur wenige Monate vor der Wiedervereinigung - wurde die DDR-Basisbefragung durchgeführt. Nach Abschluss der Feldarbeiten lagen Daten für 2179 Haushalte vor. Anstatt eine Designgewichtung auf Basis der Schichten der Stichprobe vorzunehmen, wurden regionale Informationen direkt in die Randanpassung einbezogen. Eine weitere Besonderheit war, dass keine amtlichen Daten für die Verteilung der Haushalte in der DDR nach ihrer Größe verfügbar waren. Insgesamt wurden nur 115 Restriktionen berücksichtigt.

Als Startgewichte für die Randanpassung wurden die freien Hochrechnungsfaktoren - d.h. das Verhältnis der befragten Haushalte zur Gesamtzahl der Privathaushalte - verwendet.¹²

4.3 Stichprobe D (1994/95)

Der überwiegende Teil der Immigranten, die nach 1984 in die Bundesrepublik einwanderten, konnte von einer Stichprobe wie dem SOEP auch im Laufe der Zeit nicht adäquat erfasst werden, da diese Immigranten i.d.R. neue Haushalte gründen, die entsprechend eine Ziehungswahrscheinlichkeit von Null im SOEP haben. Berücksichtigt werden von einer laufenden Panelstichprobe also nur selektive Netzwerk-Migranten, die in bestehende Haushalte ziehen. Deshalb wurde eine spezielle Zuwandererstichprobe gezogen. Sie umfasst ca. 500 Haushalte. Die Designgewichtung ist bei Rendtel et. al. (1997) beschrieben. Unterschieden wurde nach validen Haushalten in „Schneeball – Haushalten“, die auf Adressen beruhen, die ihrerseits von zufällig gewählten Haushalten genannt wurden. So konnte preiswert die Fallzahl erhöht werden; jedoch wurde am Ende davon abgesehen, diese Haushalte in die Hochrechnung einzubeziehen; d. h. für Schneeball-Haushalte wird ein Gewichtungsfaktor von Null bereitgestellt.

Die Randanpassung konnte mangels empirischer Grundlagen und wegen der sehr geringen Fallzahl nur an sehr wenige Werte erfolgen.¹³

¹² Siehe hierzu Pischner (1991)

¹³ Siehe hierzu Schupp, Wagner (1995).

4.4 Stichprobe E (1998)

Stichprobe E ist die erste von bisher drei Stichproben, um das SOEP bevölkerungsrepräsentativ zu ergänzen. Sie wurde 1998 erhoben und umfasste damals 1067 befragte Haushalte. Sie ist nicht nach Befragungsmerkmalen geschichtet; deshalb gab es vom Design her keine systematischen Über- oder Untererfassungen von Haushalten oder Personen. Eine Randanpassung war wie üblich erforderlich.¹⁴

Die Randanpassung wurde getrennt nach alten und neuen Ländern durchgeführt. Angesichts der relativ geringen Fallzahl wurden lediglich neun Restriktionen - abgesehen von der Unterteilung nach Ost und West - berücksichtigt:

Haushaltsgröße (1- 2 -, 3-, 4 – sowie 5 und mehr Personenhaushalte)

Altersklasse (15-69 Jahre, 70 Jahre und älter)

Geschlecht (Männlich)

Nationalität (nicht deutsch)

4.5 Stichprobe F (2000)

Stichprobe F verdoppelte nahezu den Umfang der „Altstichproben“ A bis E. Sie ist wie Stichprobe E nicht geschichtet - von einer Besonderheit für Ausländer abgesehen - ; aber es gab – wie üblich – selektiven Non-Response.

Angesichts der hohen Fallzahl (6052 neue befragte Haushalte) und dem entsprechenden Gewicht dieser Stichprobe für den gesamten SOEP-Stichprobenbestand kommt einerseits eine sparsame Gewichtung, wie sie für die Ergänzungsstichprobe E durchgeführt wurde, nicht in Frage, andererseits wurde Wert darauf gelegt, die Varianz der Hochrechnungsfaktoren möglichst klein zu halten.

Der bekannten Untererfassung von Ausländerhaushalten bei nach dem ADM-Mastersample durchgeführten Random-Route-Befragungen¹⁵ konnte teilweise entgegengewirkt werden bei Anlage des Stichprobenplans der Stichprobe F, in dem doppelt so viele Adressen wie notwendig erfasst wurden, jedoch von den „überzähligen“ Haushalten nur solche befragt wurden, in

¹⁴ Siehe hierzu Pischner (1999).

¹⁵ Diese Untererfassung beruhte auf den Stimmbezirken der Bundestagswahl als Auswahlgrundlage, die naturgemäß nur Wahlberechtigte, d.h. keine Ausländer, enthalten.

denen Ausländer leben. Dieser Tatsache wurde bei der Generierung der Designgewichte Rechnung getragen.¹⁶

Weiterhin hätten die verfügbaren Informationen es erlaubt - wie bei den Stichproben A und B praktiziert - Ausschöpfungsquoten der Sample-Points zusätzlich zu berücksichtigen. Im Hinblick auf möglichst geringe Varianz der Hochrechnungsfaktoren wurde ein eleganterer Weg eingeschlagen, auftretende Probleme, die durch eine unterschiedlicher Ausschöpfung von Sample-Points auftreten, zu lösen. Dies erfolgte durch Berücksichtigung sog. Quartiersvariablen, die von den Interviewern bei der Auflistung der Brutto-Adressen erhoben wurden. Detaillierte Angaben sind zu finden bei Pischner (2001):

- Entfernung des Haushalts zur nächsten Großstadt (100 Tsd. Einwohner und mehr)
- Gebäudetyp.
- Unmittelbare Umgebung des Wohnhaushalts.
- Soziale Schicht der Mehrzahl der Bewohner im Wohngebiet des zu befragenden Haushalts.

4.6 Stichprobe G (2000)

Die Hocheinkommensstichprobe weist nur Haushalten, die im Jahr 2003 – also in der zweiten Welle - über ein Netto-Haushaltseinkommen von wenigsten 4500 € verfügen, einen positiven Hochrechnungsfaktor zu. Haushalte, die der Stichprobe G, deren Einkommen niedriger lag, wurde im Rahmen der Weiterverfolgung nicht berücksichtigt und auch für Welle 1 mit einem Gewicht von Null versehen. Die Stichprobe G ist nicht als Ergänzung wie die Stichproben E ,F sowie ab 2006 H zu sehen, zählt deshalb somit nicht zum „Standardpanel“¹⁷ und wird gesondert behandelt.¹⁸

4.7 Stichprobe H (2006)

Mit der Datenweitergabe Welle W (2006) ist die Datenbasis des SOEP um die Ergänzungstichprobe H erweitert worden. Dadurch erhöht sich die Zahl der Haushalte um mehr als 1500

¹⁶ Siehe hierzu Spiess (2001).

¹⁷ Siehe hierzu unten Abschnitt 4.2.

¹⁸ Siehe hierzu Wagner und Schupp (2003) und Wagner et al. (2007).

Haushalte. Eine Startgewichtung der neuen Stichprobe H wurde *bislang* nur indirekt und vorläufig vorgenommen; sie ist vorläufig in die Standardhochrechnung integriert. Auch für die Stichprobe H würde eine Berücksichtigung allzu vieler Restriktionen zu einer deutlichen Erhöhung der Varianz in den Gewichten führen.

Stichprobe H wurde – wie üblich getrennt nach alten und neuen Ländern – an die Haushaltsgrößenstruktur des Mikrozensus angepasst. Die demographischen Merkmale Alter, Geschlecht und Nationalität der Haushaltsmitglieder werden nur für die Gesamtstichprobe an die Eckgrößen des Mikrozensus angepasst (s.w.u.).

Bis zur Datenweitergabe 2008 wird auf Basis sehr detaillierter Ausfallanalysen eine Hochrechnung analog der Hochrechnung von Stichprobe F anhand von „Quartier-Beschreibungen“ vorgenommen werden.¹⁹

5 Änderungen der Hochrechnung ab Erhebungsjahr 2005

5.1 Angaben für das Land Berlin nicht mehr nach West und Ost getrennt.

Die wellenspezifischen Hochrechnungs- und Gewichtungsfaktoren des SOEP werden jährlich an die jeweiligen Daten des Mikrozensus (MZ) angepasst, so dass die SOEP-spezifische Verteilung nach Region, Alter, Geschlecht, Haushaltsgröße und Nationalität in den Eckdaten derjenigen des MZ entspricht. Ab dem Jahr 2005 werden Angaben für das Land Berlin nicht mehr nach West und Ost getrennt ausgewiesen, sondern insgesamt den neuen Ländern zugeordnet. Deshalb kann seit dem Jahr 2005 (Welle 22, V) die Variablen \$SAMPREG, die die Stichprobenregion nach alten und neuen Ländern zum Zeitpunkt der Wiedervereinigung unterscheidet, nicht mehr im Hochrechnungsrahmen zur Trennung zwischen alten und neuen Ländern Verwendung finden. Die Hochrechnung, die sparsam an nur wenige Eckgrößen des Mikrozensus anpasst, um eine hohe Varianz der Gewichte zu vermeiden, kann einen solchen Bruch in der amtlichen Statistik nur bedingt ausgleichen. Denn durch die Zuordnung der Großstadt Berlin (West) mit ca. 1,3 Mio. Haushalten zu den stärker ländlich geprägten neuen Ländern vergrößerte sich die Population der "Ostdeutschen" Haushalte um 18%. Die Interpre-

¹⁹ Wobei freilich die „Quartierbeschreibungen“ nicht mehr von den Infratest-Interviewern erhoben wurden, sondern aus der kommerziellen Datenbank MICROM stammen (vgl. Schraepler et al. 2007).

tation von Ergebnissen, die sich auf Veränderungen des Jahres 2005 gegenüber 2004 beziehen, sollte diesem Umstand unbedingt Rechnung tragen.

5.2 Änderungen bei Zahl und Art der Hochrechnungsfaktoren.

Zum besseren Verständnis werden die Bezeichnungen für die Gewichte vorab erläutert: Jedes Gewicht wird mit $\$xHRFy$ bezeichnet:

Es bedeuten: $\$$ = Wellenkennzeichen **A,B,...,W** für die Jahre 1984,1985,...,2006.

x = Unterscheidung nach Haushalten ($x = H$) und Personen ($x = P$)

HRF kennzeichnet die Variable als Hochrechnungsfaktor

y = Eine Zusatzkennung, die die Art des Gewichts beschreibt

$y = \langle \text{leer} \rangle$, also nicht besetzt, bezeichnet Standardhochrechnungsfaktoren. Standardgewichte umfassen sämtliche Samples mit Ausnahme der Hocheinkommensstichprobe G. Diese Gewichte sind für sämtliche Wellen verfügbar.

$y = 1$ bezeichnet modifizierte Standardhochrechnungsfaktoren

Diese sind normalerweise identisch mit den Standardgewichten; allerdings werden die Gewichte von Stichproben in ihrer ersten Welle auf Null gesetzt. Zur erstmals mit der Datenauslieferung 2007 durchgeführten Null-Setzung der Querschnittsgewichte der samplespezifischen ersten Wellen sei angemerkt: Es ist inzwischen bekannt, dass Befragte bei der Erhebung ungewohnter Merkmale (wie z.B. Lebenszufriedenheit und Haushaltseinkommen) in den ersten Wellen eines Panels signifikante Lerneffekte zeigen.²⁰ Deswegen empfiehlt es sich, erste Wellen nicht gewichtet in deskriptive Analysen einzubeziehen.²¹ Durch das Nullsetzen der Hochrechnungsfaktoren geschieht dies gewissermaßen automatisch²². Nur nicht für Sample C, da in der damaligen DDR das Einkommen sehr einfach strukturiert war und deshalb eine Analyse auch der ersten Welle sinnvoll erscheint.

$y = \text{ALL}$, umfasst sämtliche erhobene Stichproben des SOEP

$y = D$ kennzeichnet die isolierte Zuwanderer-Stichprobe D

²⁰ Vgl. Frick et al. (2006)

²¹ Erste Welle-Effekte können indes mit multivariaten Verfahren kontrolliert werden.

²² Dies hat zur Konsequenz, dass mit Hilfe von $\$HRF1$ keine Analysen für das Jahr 1984 möglich sind.

$y = G$ kennzeichnet die isolierte Reichenstichprobe G

In der folgenden Tabelle 4.2 - 1 sind diese Zusammenhänge noch einmal tabellarisch zusammenfassend dargestellt. Die ausgewiesenen Fallzahlen und Ecksummen der Gewichte beziehen sich nur auf die Privathaushalte; Anstaltshaushalte, die bekanntlich nicht repräsentativ im SOEP erfasst sind, bleiben unberücksichtigt.

Tabelle 2 –1

Hochrechnungsfaktoren für Privathaushalte im Sozio-oekonomischen Panel für die Wellen A – W (1984 – 2006) – In den Gewichten enthaltene Stichproben und Eckdaten –

Welle \$	Jahr	\$HHRF	\$HHRF1	\$HHRFALL	\$HHRFy *	Zahl der Privathaushalte in der Standard- Stichprobe**	Zahl sämtlicher Haushalte**	Hochgerechnete Privathaushalte in der Grundgesamtheit in Tsd.
A	1984	AB	= 0	.	.	5853	5921	26076
B	1985	AB	AB	.	.	5238	5322	26367
C	1986	AB	AB	.	.	4991	5090	26739
D	1987	AB	AB	.	.	4920	5026	27006
E	1988	AB	AB	.	.	4719	4814	27402
F	1989	AB	AB	.	.	4602	4690	27793
G	1990	ABC	ABC	.	.	6722	6819	34848
H	1991	ABC	ABC	.	.	6581	6699	35256
I	1992	ABC	ABC	.	.	6564	6665	35700
J	1993	ABC	ABC	.	.	6537	6637	36230
K	1994	ABC	ABC	.	.	6459	6559	36695
L	1995	ABCD	ABC	.	.	6656	6768	36938
M	1996	ABCD	ABCD	.	D	6591	6698	37281
N	1997	ABCD	ABCD	.	D	6508	6617	37456
O	1998	ABCDE	ABCD	.	D	7359	7486	37532
P	1999	ABCDE	ABCDE	.	D	7905	7215	37794
Q	2000	ABCDEF	ABCDE	.	D und F	12905	13.078	38123
R	2001	ABCDEF	ABCDEF	.	D	11667	11783	38455
S	2002	ABCDEF	ABCDEF	ABCDEFG	D und G	11202	12308	38720
T	2003	ABCDEF	ABCDEF	ABCDEFG	D und G	10987	11910	38945
U	2004	ABCDEF	ABCDEF	ABCDEFG	D und G	10641	11642	39121
V	2005	ABCDEF	ABCDEF	ABCDEFG	D und G	10321	11294	39178
W	2006	ABCDEFH	ABCDEF	ABCDEFGH	D und G	11409	12361	39178***

* y=D: Zuwanderer. y=F Ergänzung 2000 y=G: Hocheinkommensstichprobe.

** Nur Haushalte mit positivem Gewicht.

*** vorläufig. Die Daten basieren auf dem Mikrozensus 2005.

5.3 Leichte Modifizierung des Hochrechnungsrahmens rückwirkend für die Erhebungsjahre 2005 und später

Eine Umstellung innerhalb der amtlichen Statistik und die Einbeziehung der Stichprobe H wurde zum Anlass genommen, den Stichprobenrahmen leicht zu modifizieren. Außerdem wurde die Abgrenzung der sog. Standard-Hochrechnungsfaktoren rückwirkend ab 1984 geändert.

Seit dem Erhebungsjahr 2000, also mit Einführung der Stichprobe F, wurden die Standardhochrechnungsfaktoren im Prinzip auf der Basis von vier Hochrechnungsrahmen ermittelt; nämlich getrennt nach alten und neuen Ländern und noch einmal unterteilt nach den Altstichproben A-E und der Ergänzungsstichprobe F. Eingeflossen in die Hochrechnung waren jeweils 22 Restriktionen, wobei deren fünf auf die Haushaltsgröße und 17 auf Ausländeranteil, Geschlecht und Altersklassen entfielen.

Beginnend mit Welle V wurde zwar auf eine gesonderte Hochrechnung für die Altstichproben A-E vs. Ergänzungsstichprobe F verzichtet, deren Stichprobenanteile werden in der Klassifizierung nach Haushaltsgröße indes weiterhin berücksichtigt.

Somit gibt es einen stichprobenspezifischen Teil (Tabelle 4.3-1) und einen von den Stichproben unabhängigen Teil des Hochrechnungsrahmens (Tabelle 4.3-2).

Dieses Vorgehen ermöglicht es, bei Bedarf weiterhin getrennte Auswertungen z.B. für Stichprobe F für das Jahr 2005 vorzunehmen:

Zunächst setzt man sämtliche Gewichte für Haushalte, die nicht der Stichprobe F entstammen auf Null. Die verbleibenden Gewichte müssen nur noch durch den Stichprobenanteil – also 0.45 – dividiert werden und man erhält einen Gewichtungsrahmen, der die Struktur der Haushalte nach Größe korrekt widerspiegelt. Die Stichprobenanteile sind in der Tabelle 4.3-1 explizit ausgewiesen. Die übrigen Restriktionen werden nur annähernd erfüllt; an dieser Stelle müssen leichte Abweichungen von den Ecksummen hingenommen werden.

Data Documentation 22

5 Änderungen der Hochrechnung ab Erhebungsjahr 2005

Tabelle 3 –1

Stichprobenspezifischer Eckwerte zur Querschnittshochrechnung der Privathaushalte für die Jahre 2005 und 2006 (vorläufig)

		Welle V (2005)		Welle W (2006)			Welle W (2006)	
		Standard: ABCDE	VHHRF F	Standard: ABCDE	WHHRF F	Alle Stichproben H	WHHRFALL ABCDEFG	H
	Insgesamt	Anteil in %	Anteil in %	Anteil in %	Anteil in %	Anteil in %	Anteil in %	Anteil in %
	Insgesamt	0,55	0,45	0,4785	0,3915	0,13	0,88	0,12
Früheres Bundesgebiet ohne Berlin (West)	Hochgerechnete Privathaushalte in Tsd.							
1-Personen Haushalte	11323	6228	5095	5418	4433	1472	9964	1359
2-Personen Haushalte	10332	5683	4649	4944	4045	1343	9092	1240
3-Personen Haushalte	4207	2314	1893	2013	1647	547	3702	505
4-Personen Haushalte	3517	1934	1583	1683	1377	457	3095	422
5+Personen Haushalte	1353	744	609	647	530	176	1191	162
Insgesamt	30732	16903	13829	14705	12032	3995	27044	3688
Neue Länder und Berlin								
1-Personen Haushalte	3373	1855	1518	1614	1321	438	2968	405
2-Personen Haushalte	2934	1614	1320	1404	1149	381	2582	352
3-Personen Haushalte	1270	699	572	608	497	165	1118	152
4-Personen Haushalte	695	382	313	333	272	90	612	83
5+Personen Haushalte	174	96	78	83	68	23	153	21
Insgesamt	8446	4645	3801	4041	3307	1098	7432	1014
Deutschland								
Insgesamt	39178	21548	17630	18747	15338	5093	34477	4701

Tabelle 4–1

Stichprobenunabhängige Eckwerte zur Querschnittshochrechnung der Privathaushalte für die Jahre 2005 und 2006 (vorläufig)

Bevölkerung in Privathaushalten

Alter von.. bis unter.. Jahren	Deutschland	Früheres Bundes- gebiet	Neue Länder und Berlin
	In Tsd.		
Insgesamt	82676	65917	16759
Unter 5	3487	2869	618
5 - 10	3939	3365	574
10 - 15	4076	3513	563
15 - 20	4950	3832	1118
20 - 25	5137	3975	1162
25 - 30	4928	3889	1038
30 - 35	5142	4174	968
35 - 40	6700	5472	1228
40 - 45	7243	5721	1522
45 - 50	6170	4846	1325
50 - 55	5621	4345	1277
55 - 60	4830	3874	956
60 - 65	5018	3887	1131
65 - 70	5471	4248	1223
70 - 75	3764	2937	827
75 - 80	3022	2406	616
80 und mehr	3174	2564	613
Darunter:			
Im deutschen Haus- halt lebend	75364	59376	15998
Männer	40602	32364	8238

Quelle: Sonderauswertungen des Mikrozensus 2005

6 Querschnittshochrechnung für Personen

Die Methode zur Schätzung der Personengewichte blieben abgesehen von der veränderten regionalen Zuordnung von Berlin (West) zu den neuen Ländern unverändert.

Die nachfolgenden Tabellen geben einen allgemeinen Überblick über die verfügbaren Gewichte für Personen seit 1984 (Tabelle 5-1) sowie eine detailliertere Ansicht für die Jahre 2005 und 2006 (Tabelle 5 – 2).

Tabelle 5–1

Hochrechnungsfaktoren für Personen in Privathaushalten im Sozio-oekonomischen Panel für die Wellen A – W (1984 – 2006) – In den Gewichten enthaltene Stichproben und Eckdaten –

Welle \$	Jahr	\$PHRF	\$PHRF1	\$PHRFALL	\$PHRFy *	Zahl Personen in Privathaushalten in der Standard-Stichprobe**	Zahl sämtlicher Personen **	Hochgerechnete Zahl Personen in Privathaushalten am Hauptwohnsitz in der Grundgesamtheit in Tsd.
A	1984	AB	= 0	.	.	16099	16173	60501
B	1985	AB	AB	.	.	14443	14508	60160
C	1986	AB	AB	.	.	13742	13804	60247
D	1987	AB	AB	.	.	13496	13563	60404
E	1988	AB	AB	.	.	12817	12872	60587
F	1989	AB	AB	.	.	12393	12443	61094
G	1990	ABC	ABC	.	.	18165	18254	78678
H	1991	ABC	ABC	.	.	17756	17844	79015
I	1992	ABC	ABC	.	.	17350	17429	79624
J	1993	ABC	ABC	.	.	16993	17072	80320
K	1994	ABC	ABC	.	.	16642	16715	80588
L	1995	ABCD	ABC	.	.	17246	17345	80787
M	1996	ABCD	ABCD	.	D	16860	16942	81010
N	1997	ABCD	ABCD	.	D	16474	16570	81216
O	1998	ABCDE	ABCD	.	D	18132	18234	81110
P	1999	ABCDE	ABCDE	.	D	17382	17487	81208
Q	2000	ABCDEF	ABCDE	.	D	30587	30764	81368
R	2001	ABCDEF	ABCDEF	.	D	27762	27920	81466
S	2002	ABCDEF	ABCDEF	ABCDEFG	D und G	26119	29072	81690
T	2003	ABCDEF	ABCDEF	ABCDEFG	D und G	25202	27868	81734
U	2004	ABCDEF	ABCDEF	ABCDEFG	D und G	24372	26916	81704
V	2005	ABCDEF	ABCDEF	ABCDEFG	D und G	23292	25638	81639
W	2006	ABC- DEFH	ABCDEF	ABC- DEFGH	D und G	25213	27442	81639***

* y=D: Zuwanderer. y=F Ergänzung 2000 y=G: Hocheinkommensstichprobe.

** nur Personen mit positiven Gewicht.

*** vorläufig. Die Daten basieren auf dem Mikrozensus 2005.

Tabelle 5–2

Eckwerte zur Querschnittshochrechnung der Personen in Privathaushalten für die Jahre 2005 und 2006 (vorläufig)

Bevölkerung in Privathaushalten am Ort der Hauptwohnung

Alter von.. bis unter..	Deutschland	Früheres Bundesgebiet	Neue Länder und Berlin
Jahren	In Tsd.		
Insgesamt	81639	65044	16595
unter 5	3478	2861	617
5 - 10	3923	3351	572
10 - 15	4058	3497	561
15 - 20	4898	3794	1103
20 - 25	4885	3774	1111
25 - 30	4779	3766	1013
30 - 35	5067	4110	958
35 - 40	6639	5419	1220
40 - 45	7182	5668	1514
45 - 50	6113	4796	1317
50 - 55	5569	4300	1269
55 - 60	4783	3833	949
60 - 65	4960	3836	1124
65 - 70	5420	4201	1218
70 - 75	3731	2907	825
75 - 80	2998	2385	613
80 und mehr	3157	2546	611

Quelle: Sonderauswertungen des Mikrozensus 2005

Literatur

- Däubler*, Thomas (2002): Nonresponseanalysen der Stichprobe F des SOEP. DIW Materialien Nr. 15.
- Galler*, Heinz P. (1987): Zur Längsschnittgewichtung des Sozio-oekonomischen Panels. In: Hans-Jürgen *Krupp* und Ute *Hanefeld*: Lebenslagen im Wandel: Analysen 1987, Band 2 der Reihe: Sozio-oekonomische Daten und Analysen für die Bundesrepublik Deutschland, Frankfurt, S. 295-317.
- Frick*, Joachim R., Jan *Goebel*, Edna *Schechtman*, Gert G. *Wagner*, and Shlomo *Yitzhaki*, (2006): Using Analysis of Gini (ANoGi) for detecting whether two sub-samples represent the same universe: The German Socio-Economic Panel Study (SOEP) Experience. *Sociological Methods & Research*, 34 (4): 427-468
- Haisken-De New*, John P. and Joachim R. *Frick* (eds.) (2006): DTC - Desktop Companion to the German Socio-Economic Panel Study, DIW Berlin.
- Horvitz*, D.G. and D.J. *Thompson* (1952): A Generalisation of Sampling without Replacement From a Finite Universe, *Journal of the American Statistical Association* 47, S. 663-685.
- Kroh*, Martin und Martin *Spiess* (2006): Documentation of Sample Size and Panel Attrition in the German Socioeconomic Panel . DIW- Data Documentation No. 15
- Merz*, Joachim (1983): Die konsistente Hochrechnung von Mikrodaten nach dem Prinzip des minimalen Informationsverlusts, *Allgemeines Statistisches Archiv* 67, S. 342-366.
- Pannenberg*, Markus (2000): Documentation of the Sample Sizes and Panel Attrition in the German Socio-Economic Panel (GSOEP), DIW Diskussionspapier No. 196.
- Pischner*, Rainer (1991): Eine konsistente Haushalts- und Personengewichtung für die DDR-Basisbefragung des SOEP und für die Ost-Pilotstudie des Wohlfahrtssurveys. *Vierteljahrshefte zur Wirtschaftsforschung*, Heft 1 – 2, Berlin, S. 50-64.
- Pischner*, Rainer (1994): Quer- und Längsschnittgewichtung des Sozio-oekonomischen Panels. In: Siegfried *Gabler*, Jürgen H. P. *Hoffmeyer-Zlotnik*, Dagmar *Krebs* (Hrsg.): Gewichtung in der Umfragepraxis. Opladen, S. 166-187.
- Pischner*, Rainer (1999): Überarbeitete Querschnittshochrechnung der Wellen G-N (1990 bis 1997) des Sozio-oekonomischen Panels (SOEP) unter Einbeziehung der Ergänzungsstichprobe E. Siehe: <http://www.diw.de/deutsch/sop/service/doku/index.html>
- Pischner*, Rainer (2001): Hochrechnung der ersten Welle der Stichprobe F. Siehe: http://www.diw.de/deutsch/sop/service/doku/docs/hoch_f_0.pdf
- Rendtel*, Ulrich und Ulrich *Pötter* (1992): Über Sinn und Unsinn von Repräsentationsstudien. DIW-Diskussionspapier 61, Berlin.
- Rendtel*, Ulrich (1995): Lebenslagen im Wandel: Panelfälle und Panelrepräsentativität. Campus: Frankfurt/New York.
- Rendtel*, Ulrich, Markus *Pannenberg*, und Sascha *Daschke* (1997): Die Gewichtung der Zuwandererstichprobe des Sozio-oekonomischen Panels (SOEP). *Vierteljahrshefte zur Wirtschaftsforschung* 66, S. 271-285.
- SOEP Group* (2001): The German Socio-Economic Panel (GSOEP) after more than 15 years – Overview. In: Elke *Holst*, Dean R. *Lillard*, und Thomas A. *DiPrete* (eds.): Proceedings of the 2000 Fourth International Conference of German Socio-Economic Panel Study Users (GSOEP2000). *Vierteljahrshefte zur Wirtschaftsforschung* 70 (1), S. 7-14.

- Schraepler, Joerg-Peter, Jürgen Schupp, und Gert G. Wagner (2007): Who Are the Nonrespondents? An Analysis Based on a New Subsample of the German Socio-Economic Panel (SOEP) including Microgeographic Characteristics and Survey-Based Interviewer Characteristics. DIW Berlin Research Note (in Vorbereitung).*
- Schupp, Jürgen und Gert G. Wagner (1995): The German Socio-Economic Panel: a Database for Longitudinal International Comparisons. Innovations 8 (1), S. 95-108.*
- Schupp, Jürgen und Gert G. Wagner (2003): Repräsentative Analyse der Lebenslagen einkommensstarker Haushalte. In: Lebenslagen in Deutschland. Armuts- und Reichtumsberichterstattung der Bundesregierung. S131 ff.*
- Spiess, Martin (2001): Derivation of design weights: The case of the German Socio-Economic Panel (GSOEP), DIW-Materialien / Research notes No.5*
- Spiess, Martin und Ulrich Rendtel (2000): Combining an ongoing panel with a new cross-sectional sample. DIW-Diskussionspapier No. 198*
- Spiess, Martin und Martin Kroh (2007): Documentation of the dataset DESIGN of the Socio economic Panel Study (SOEP). Siehe: <http://www.diw.de/deutsch/sop/service/doku/index.html#1.2>*
- Wagner, Gert G., Joachim R. Frick, und Jürgen Schupp (2007): The German Socio-Economic Panel Study (SOEP) – Scope, Evolution and Enhancements. Schmollers Jahrbuch 127 (1). (Im Druck).*
- Wagner, Gert G. et al. (2007): High Incomes in Household Surveys: The Case of Sample G of the German Socio-Economic Panel Study (SOEP). DIW Data Documentation (in Vorbereitung).*